CESifo WORKING PAPERS

9365
2021

October 2021

Boosting Tax Revenues with Mixed-Frequency Data in the Aftermath of Covid-19: The Case of New York

Kajal Lahiri, Cheng Yang

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9365*

Boosting Tax Revenues with Mixed-Frequency Data in the Aftermath of Covid-19: The Case of New York

Abstract

We forecast New York state tax revenues with a mixed-frequency model using a number of machine learning techniques. We found boosting with two dynamic factors extracted from a select list of New York and U.S. leading indicators did best in terms of correctly updating revenues for the fiscal year in direct multi-step out-of-sample forecasts. These forecasts were found to be informationally efficient over 18 monthly horizons. In addition to boosting with factors, we also studied the advisability of restricting boosting to select the most recent macro variables to capture abrupt structural changes. Since the COVID-19 pandemic upended all government budgets, our boosted forecasts were used to monitor revenues in real time for the fiscal year 2021. Our estimates showed a drastic year-over-year decline in real revenues by over 16% in May 2020, followed by several upward nowcast revisions that led to a recovery to -1% in March 2021, which was close to the actual annual value of -1.6%.

JEL-Codes: C220, C320, C500, C530, E620.

Keywords: revenue forecasting, machine learning, real time forecasting, mixed frequency, fiscal policy.

Kajal Lahiri
University of Albany: SUNY
Department of Economics
USA – NY 12222
klahiri@albany.edu

Cheng Yang
Liaoning University
Advanced Institute of Finance and Economics
Liaoning / China
yangcheng@lnu.edu.cn

July 31, 2021
Preprint submitted to International Journal of Forecasting.

1. Introduction

Revenue forecasts provide a fiscal baseline that serves as the budget constraint for spending plans and tax policies. Because they are the fundamental building blocks of the budget, revenue forecasts are often sources of controversy, since forecast manipulations can be tools for controlling the state spending plans. Thus, one of the most sensitive functions of a state government is the production of its revenue forecast for annual budgets. As Rodgers and Joyce (1996) once noted, "Budgets are public statements about who gets what, when, where, how and why", and forecast is a key component of it.

Since spending at the state governments in the U.S. is around 12% of GDP and 25% of total government spending, economists have studied the mechanics of state budgeting given the special patterns of fiscal federalism in America. Notable amongst these is the balanced budget rule that limits a state's capability to react to fiscal shocks by varying degrees, depending on the stringency of the rule in a particular state. Whereas much of the increases on state expenditures over last 60 years have been induced by partially-funded federal mandates, states have increasingly relied on own-source revenues like income and sales taxes to balance the budget, cf. Baicker et al. (2012). Since these revenue sources are highly volatile and pro-cyclical, forecasting state revenues has become increasingly difficult. The largest forecast errors are found during recessions, and have gotten worse over the years. Interestingly, Boyd and Dadayan (2014) found that this pattern of errors are generated exactly by a random walk benchmark, suggesting government revenue forecasters fail to incorporate the possibility of recessions in their forecasts.

The COVID-19 pandemic, which Ludvigson et al. (2021) characterize as an unprecedented 192σ disaster shock, has dramatically affected the state economies and their budgets. Even with an optimistic scenario, Makridis and McNab (2020) have estimated a 6.7% tax-revenue decline for an average state in the U.S. due to the pandemic. In order to help the budget making process and monitor fiscal developments, we develop a high-frequency model to forecast tax revenues using a number of machine learning techniques in such a way that, in addition to the dynamics of monthly tax revenues, the forecasts for future macroeconomic conditions using a large number of local and national level leading indicators are used as critical inputs to forecasting. Our hope is that a model like ours will be able to track the effects of cataclysmic events like COVID-19 lockdown in real time. We specialize on New York not only because of our familiarity with the state, but also because

forecasting New York tax revenue is highly challenging. Apart from having a high-variance revenue sources due to its massive financial sector, New York has a bad reputation of passing late budgets due to political gridlocks, see Anderson et al. (2012). In New York, the independent forecasts by the governor and the legislature lay the foundation for budget negotiations between the Governor, Senate Majority leader and the Assembly Speaker, who have their own divergent interests, political agendas and spending needs - a fertile landscape for budget impasses in divided governments, cf. Sun (2008).

To coincide with the budget making process, our model starts forecasting annual tax revenue from six months before the beginning of a fiscal year and updates it every month till March using the latest information. This is consistent with a series of recommendations made by Boyd et al. (2011) and National Association of Budget Officers (NASBO, 2013) to improve the budget process by forecasting revenues more often, and more promptly combining many predictors. Even though we implement our methodology on the State of New York, it can be readily amended for the other 50 states in the U.S., and other countries who face budget rules like Euroland's Stability and Growth Pact under which 3% of GDP is the set as the ceiling for a country's budget deficit, see Frankel (2011). Asimakopoulos et al. (2020) have used quarterly data to nowcast and monitor annual budgetary outtutns for a number of European Union countries.[1]

We use a number of machine learning techniques to choose among a large number of high-frequency variables and their lags that will minimize prediction errors at each forecasting horizon. In recent years, these methods have been used successfully to forecast recessions, See, for instance, Ng (2014), Dopke et al. (2017), Vrontos et al. (2020), Davig and Hall (2019), and Piger (2020). The literature on machine-learning forecasting with mixed-frequency data is also growing, see Lehrer et al. (2019) for example. Babii et al. (2021) have used sparse-group LASSO (sg-LASSO) in Mixed-Data-Sampling (MIDAS) models. But these powerful methods have not been widely used in governmental budget forecasting. An exception is Buxton et al. (2019), who used deep learning to forecast the sales tax revenues in Illinois. One of the weaknesses of existing literature is that the importance of business cycles is not fully integrated in budget forecasting. Our study is also close to a comprehensive study by Ghysels et al. (2020), who used MIDAS model to estimate the effect of the COVID-19 pandemic on forecasting the revenues and expenditures of 48 contiguous

[1] See also Pérez (2007), Pedregal and Pérez (2010), Onorante et al. (2010), and Paredes et al. (2014).

States of the U.S. using quarterly Commerce department fiscal data.

We experimented with a number of boosting (Breiman, 1998, Friedman, 2001, Bühlmann and Yu, 2003) and LASSO (Tibshirani, 1996) algorithms to study which one would work best given our novel context and the data. We used the conventional boosting with monthly lags of taxes and macro variables, boosting with macro factors, LASSO, sg-LASSO, boosting/sg-LASSO with Legendre Polynomials and boosting with tree base-learner in our analysis. Since boosting may not always chose the latest macroeconomic data, we also restricted boosting to select among the most recent macro variables to capture abrupt structural changes promptly. After balancing the pros and cons of all the alternatives, we found boosting with factors worked best in our context.

The plan of the paper is as follows: We briefly introduce the machine learning techniques and the MIDAS models in Section 2. Section 3 outlines the revenue forecasting schedule in New York and the data. Empirical results are presented in four subsections in Section 4. Two sub-sections of Section 5 present our estimates of revenue short fall due to the COVID-19 pandemic for the fiscal year 2021, and a comparison with some recent studies. Section 6 summarizes the main empirical findings of the paper.

2. Models

2.1. Component-Wise Gradient Boosting

Before discussing aspects of forecasting with mixed-frequency data, we first introduce the basic component-wise gradient boosting, which is a useful tool for our purpose. The methodology is closer to regular statistical or econometric models and more interpretable compared with many of other machine learning techniques, cf Ng (2014). It can deal with highly non-linear lag functions in high dimensional models nonparametrically, cf.Bühlmann and Hothorn (2007).

Given data (y_i, x_i) with $x_i = (x_{i1}, \ldots, x_{ip})$, the goal is predicting y with x. Defining a smooth and convex loss function $\rho(,)$, the goal is minimizing the expected loss:

$$\min_{f(.)} \frac{1}{n} \sum_{i=1}^{n} \rho(y_i, f(x_i)). \tag{1}$$

A parametric function $f(x_i)$ is referred to as a weak learner. Boosting then starts with an initial prediction $\hat{f}^0$ (e.g. $\hat{f}^0 = \bar{y}$). At each iteration m, define negative gradient as $u_i^m = -\frac{\partial}{\partial f}\rho(y_i, \hat{f}^{m-1}(x_i))$

for $i = 1, \ldots n$, thus $u_i^1 = -\frac{\partial}{\partial f}\rho(y_i, \hat{f}^0(x_i))$. Given a set of base-learners chosen, boosting fits each one of them to u_i^m to get $\hat{u}_i^m$, solving the following problem in Iteration m:

$$\min_{f(.)} \frac{1}{n} \sum_{i=1}^{n} \rho(u_i^{(m)}, f(x_i)). \tag{2}$$

Each base-learner can be a simple estimator with a few or just one of the predictors in x_i, using simple models such as OLS, regression trees (Hothorn et al., 2006) or P-splines (Eilers and Marx, 1996). Boosting procedure then chooses one base-learner that helps the gradient descent the most and adds it to update the prediction $\hat{f}^{(m)} = \hat{f}^{(m-1)} + \nu\hat{u}^{(m)}$. The shrinkage parameter $\nu \in (0, 1)$ is the step size that prevents the model from taking the full contribution of the updates, and helps to avoid over-fitting. Thus, the estimation $\hat{f}^{(m)}$ is continuously improved by the little boosts $\nu\hat{u}^{(m)}$. The iteration stops at a chosen $m = M$. In our application, we use the usual L_2 loss function: $\rho(y, f) = (y - f)^2/2$. This framework can be easily adapted to forecasting problems with time series, even when the data are mixed-frequency.

2.2. Boosting with Mixed-frequency Data

It has already been shown that MIDAS models can improve the forecasts of fiscal variables by incorporating potentially richer information contained in high-frequency variables, cf. Ghysels and Ozkan (2015), and Ghysels (2020). The improvement can be achieved through the lag distribution function which does temporal aggregation in a way that is better than taking arithmetic averages. However, even if the parameters of the function can cover different shapes of a distribution, the flexibility is still limited. The range of shapes is fixed once the type of lag distribution function is chosen. The distributions still may not adapt to various cases where the weight changes in a complicated way across high-frequency lags. Boosting is a procedure that does not require a specific distribution function restriction. It also does not suffer from numerical problems of collinearity when there are multiple high-frequency predictors, cf. Andreou et al. (2013). Further, only simple weak learners are needed for boosting. Another issue is that the leading characteristics of the predictors can be different, especially in times of recessions. Depending on the horizon, some lag(s) may be helpful but others may not (Ng 2014). Boosting can be a helpful tool that also deals with this issue.

Therefore we integrate L_2 boosting with ADL-MIDAS model, using monthly and quarterly data. The model is adapted to mixed-frequency data following Bai and Ng (2009) [2].

$$y_{t+h} = \alpha + \gamma(L^{12})y_t + \beta(L)x_{t+\omega} + \epsilon_{t+h}, \qquad t = 12, 24, 36, \ldots \tag{3}$$

with

$$\beta(L)x_{t+\omega} = \sum_{k=1}^{n_M} \beta_k(L)x_{t+\omega,k} \tag{4}$$

and

$$\beta_k(L)x_{t+\omega,k} = \sum_{j=0}^{J_M-1} w(j;\beta_k)x_{t+\omega-j,k}, \tag{5}$$

where L is monthly lag operator, L^{12} is annual lag operator, y_{t+h} the annual target variable, $x_{t+\omega}$ is a $n_M \times 1$ vector of monthly variables, ω controls the number of months with released data within a fiscal year, $\gamma(L^{12})$ defines the coefficients of the lagged annual dependent variable, $\beta(L)$ is a $1 \times n_M$ vector defining the coefficients of the lags of the monthly variables, J_M is the number of monthly lags, and $w(;)$ is the weight of each lag. $\beta(L)x_{t+\omega}$ can be large. For example, if there are 30 monthly variables with 12 lags in each, $\beta(L)x_{t+\omega}$ will have 360 terms. Thus the large dimension of predictors comes from both the number of predictors and the fact that the predictors can be of high-frequency. Clearly it is a type of U-MIDAS model (Foroni et al., 2015), and can be implemented by machine learning (Babii et al., 2021). Note that $x_{t+\omega}$ can include factor(s) extracted from the macro variables, which will be introduced in the next sub-section. If quarterly variables are also used, as we do, the model becomes

$$y_{t+h} = \alpha + \gamma(L^{12})y_t + \beta(L)x_{t+\omega} + \kappa(L^3)x^Q_{t+\omega^Q} + \epsilon_{t+h}, \qquad t = 12, 24, 36, \ldots \tag{6}$$

where L^3 is quarterly lag operator, $x^Q_{t+\omega^Q}$ is a $n_Q \times 1$ vector of quarterly variables, $\kappa(L^3)$ is a $1 \times n_Q$ vector defining the coefficients of their lags, ω^Q helps to control the most recent released quarterly data within a fiscal year. $\kappa(L^3)x^Q_{t+\omega^Q}$ can be explained in a way similar to Equations (4) and (5). When implementing the boosting procedure, we first define $\tilde{z}_{t+\omega}$ as the set of high-frequency

[2] Yousuf and Ng (2020) have extended the boosting procedure of Bai and Ng (2009) and allow for time-varying parameters, when the target variable is only locally stationary. We do not adopt this extension since our target variable is non-trending with limited number of annual observations.

predictors aligned at low-frequency level

$$\tilde{z}_{t+\omega} = (\tilde{y}_t, \tilde{x}_{t+\omega}, \tilde{x}^Q_{t+\omega}) \qquad t = 12, 24, 36, \ldots \tag{7}$$

where $\tilde{y}_t = (y_t, \ldots, y_{t-12n_y+12})$, $\tilde{x}_{t+\omega} = (x_{t+\omega}, x_{t+\omega-1}, \ldots, x_{t+\omega-J_M+1})$, $\tilde{x}^Q_{t+\omega^Q} = (x^Q_{t+\omega^Q}, x^Q_{t+\omega^Q-3}, \ldots, x^Q_{t+\omega^Q-3J_Q+3})$; n_y is the number of ADL lags and we let it be 2 in our application; J_M and J_Q are number of monthly and quarterly lags which may vary with forecast horizons. Thus all the low-frequency and high-frequency lags relevant to forecast y_{t+h} are turned into a vector of predictors $\tilde{z}_{t+\omega}$. We let the total number of predictors be $K = n_y + n_m J_M + n_Q J_Q$, which is allowed to be greater than the low-frequency time length $T/12$. In our empirical analysis, we also experiment setting $J_M = 1$ and $J_Q = 1$ for macro variables and restrict them to be most recent at the time the forecasts and nowcasts are made.

A boosting procedure for mixed-frequency data with linear base-learner is summarized as the following:

(1) Set the initial forecasts as the sample mean: $\hat{\Psi}^{(0)}(\tilde{z}_{t+\omega}) = \bar{y}_{t+h},\ t = 12, 24, \ldots, T-h$;

(2) For iteration $m = 1, \ldots, M$, do (a)-(c) as:

(a) Compute $\hat{u}^{(m)}_{t+h} = y_{t+h} - \hat{\Psi}^{(m-1)}(\tilde{z}_{t+\omega}),\ t = 12, 24, \ldots, T-h$;

(b) Choose $\hat{i}_m = argmin_{i \le K} \sum_{t=12,24,\ldots}^{T-h} (\hat{u}^{(m)}_{t+h} - \hat{b}^{(m)}_i \tilde{z}_{i,t+\omega} - \hat{a}^{(m)}_i)^2$,

where $\hat{a}^{(m)}_i$ and $\hat{b}^{(m)}_i$ are OLS estimators for the linear equation

$$\hat{u}^{(m)}_{t+h} = a + b\tilde{z}_{i,t+\omega} + \tilde{\epsilon}_{t+h}, \quad t = 12, 24, \ldots, T-h,$$

where $\tilde{\epsilon}_{t+h}$ is an error term.

(c) Denoting $\hat{h}_m(\tilde{z}_{\hat{i}_m,t+\omega}) \equiv \hat{b}^{(m)}_{\hat{i}_m} \tilde{z}_{\hat{i}_m,t+\omega} + \hat{a}^{(m)}_{\hat{i}_m}$, update the forecasts $\hat{\Psi}^{(m)}(\tilde{z}_{t+\omega}) = \hat{\Psi}^{(m-1)}(\tilde{z}_{t+\omega}) + \nu \hat{h}_m(\tilde{z}_{\hat{i}_m,t+\omega}),\ t = 12, 24, \ldots, T-h$. $\nu \in (0, 1)$ is a shrinkage parameter;

(3) Get the prediction $\hat{\Psi}^{(M)}(\tilde{z}_{t+\omega}) = \hat{\Psi}^{(0)}(\tilde{z}_{t+\omega}) + \nu \sum_{m=1}^{M} \hat{h}_m(\tilde{z}_{\hat{i}_m,t+\omega}),\ t = 12, 24, \ldots, T-h$;

(4) The out-of-sample forecast of horizon $h - \omega$ is $\hat{\Psi}^{(M)}(\tilde{z}_{T+\omega})$.

The forecasts are slowly improved as predictors are selected through the iterations. A trivial innovation that differs from the original formulation is that our number of monthly or quarterly lags can be different across predictors and horizons. The procedure also applies to other base-learners.

Following the literature, we set the shrinkage parameter ν to be 0.1, which is a relatively small fraction, making the learning slower. With a fixed ν, the trade off between the number of iterations M and ν gets determined, and the number of iterations M for each horizon can be chosen by a model selection criterion like AIC. The main procedure of boosting can be implemented by R package *mboost*.[3] Some functions from R package *midasr* are used to help align mixed-frequency data.[4]

2.3. Boosting with Factors

As stated in the previous subsection, $x_{t+\omega}$ can be factors extracted from the macro variables instead of the macro variables themselves. We get two monthly dynamic factors following Bańbura and Modugno (2014)

$$\dot{x}_t = \Lambda f_t + u_t \tag{8}$$

$$f_t = A f_{t-1} + \eta_t \tag{9}$$

where $\dot{x}_t$ is the vector of macro variables, Λ is the factor loading matrix and f_t is the vector of factor. A defines the matrix of autoregressive coefficients for the vector of factors. Both u_t and η_t are normal, and the variance covariance matrix for u_t is diagonal. We use all macro variables in Table 2 and 3 to get the factors. When boosting with factors, the model becomes

$$y_{t+h} = \alpha + \gamma(L^{12})y_t + \beta_1(L)f_{t+\omega} + \beta_2(L)MTax_{t+\omega} + \epsilon_{t+h}, \qquad t = 12, 24, 36, \ldots \tag{10}$$

where MTax is the monthly tax variable (Table 1), which directly adds up to the annual target variable, and will be introduced in the next section. Our model can deal with different types of missing data. The first type comes from publication lags or different releasing dates (the ragged-edge problem). The second type comes from the fact that some macro variables are observed every quarter and are treated as monthly variables with two missing observations in every quarter. The third type comes from the fact that some variables might not have been available in the past. One important example used in our study is the Business Leaders Survey from Federal Reserve Bank

[3] See Hofner et al. (2014) for a useful tutorial.
[4] See Ghysels et al. (2016) details of the package.

of New York, which started only from 2004.

We set the number of factors to be 3 and the first 2 monthly factors are used as predictors. We tried different ways to choose the number of factors (cf. Bai and NG (2002)) and experimented with having additional factors in boosting. The results showed that the third and fourth factor were never selected by boosting even though they are included in the model.

2.4. Other Models

As robustness checks, we also considered few other models or specifications to make the forecasts: The first one is LASSO, which also selects from a large number of variables. The second one is sg-LASSO by Simon et al. (2013), which is adopted by Babii et al. (2021) to time series forecasting with large mixed-frequency data. sg-LASSO assigns certain group of variables to belong to specific groups. It selects groups of variables and variables within groups simultaneously. Due to the nature of time series variables, we treat all the lags from the same high-frequency variable as being in the same group. The third one is sg-LASSO/boosting with Legendre Polynomials, introduced by Babii et al. (2021), where the high-frequency data is aggregated by a lag distribution function that is approximated by multiple fixed-shaped distribution functions, which are selected by sg-LASSO or boosting. Our boosting models with mixed-frequency data can be treated as a special case, where 100% weight is assigned to each high-frequency lag as a distribution function. Other alternatives we experimented with include other base-learners in boosting such as tree base-learner and P-splines base-learner.

2.5. Benchmark Models

We use ADL-MIDAS (Ghysels et al., 2007) with only one high-frequency predictor as the first benchmark model. Aggregating monthly taxes by taking arithmetic averages over previous J months, we obtain

$$y_{t+h} = \alpha + \beta_0 y_t + \beta_1 \frac{1}{J} \sum_{j=0}^{J-1} MTax_{t+\omega-j} + \epsilon_{t+h}, \qquad t = 12, 24, 36, \ldots \tag{11}$$

It is a special case of ADL-MIDAS model since the lag distribution is flat and fixed at $1/J$, and is updated monthly. Deseasonalized monthly tax variable is turned into percentage change from the

average of previous fiscal year. The number of lags J is chosen such that all available lags of a fiscal year at particular horizon are picked and the lags do not extend back to another fiscal year. Thus the average reflects information about cumulative tax of a fiscal year. Note that this benchmark model is not naive since the information contained in the monthly tax variable is a rich source of high-frequency information for nowcasting.

Random walk model (with drift) at annual frequency is an alternative benchmark

$$y_{t+12} = \mu + y_t + \epsilon_{t+12}, \qquad t = 12, 24, 36, \ldots. \tag{12}$$

This benchmark model depends on horizons: When comparing with our mixed-frequency machine learning models at horizons longer than 12 months, where y_t is not known, y_{t-12} is used and iterated to make forecast for y_{t+12}; For horizons shorter than or equal to 12 months, y_t is known and used to make nowcast for y_{t+12}.

3. Data

The data we use includes the monthly total net tax receipts from the New York State Department of Tax and Finance. Currently, monthly tax collection reports starting from FY 1996 are posted on its website. We worked with the tax department staff to collect and record data going back to FY 1987. Each report contains tax collections of all categories in a month. Unfortunately, the reports are not directly usable for econometric analysis because of changes in data definitions over years, change of variable names, shifts and rumps in series, transcript errors, outliers, paired opposite outliers, etc. Williams (2008) discusses a number of such issues while assembling revenue data at the state level. For us, the most challenging issue was that of reconciling reporting methodologies in different years in order to create a consistent time series. The majority of the conversions from hard copies to Excel had to be done manually over several weeks as there were over 1,000 pages of spreadsheets. After considerable discussions in many meetings with the Department of Tax and Finance staff, we developed a clean data base of 32 consistent series of individual taxes beginning 1987, as shown in Table 1. Developing a clean data set for econometric analysis was a non-trivial part of our efforts.[5] These 32 monthly series are usually aggregated into personal in-

[5] John Delaney and Onur Bugdayci spent considerable time with undergraduate interns to develop the data base and doing the preliminary statistical work on seasonal adjustments and factor analysis. They found the individual

come tax (PIT), sales and use tax, business taxes and others. The annual sum of these components make up total tax revenue (MTax) that we forecast in this paper, and is very close to the series New York Division of the Budget (NY DOB) forecasts.[6]

In addition, we selected more than thirty cyclical macroeconomic variables, part of which are related to New York state, New York metropolitan area, Port Authority and the Northeast region of the U.S. We chose these from a much larger set of variables based on their leading values. Thus, many of the coincident and lagging economic indicators were not included deliberately. One of the most important variables for our purpose is the Leading Economic Index (LEI) from Conference Board. LEI is a composite index computed using 10 variables including initial unemployment insurance claims, Institute for Supply Management (ISM) new orders diffusion index and 7 other variables. We also use several other ISM indexes such as ISM Manufacturing Employment Index and ISM Non-Manufacturing New Orders Index. Another important variable is Blue Chip Real GDP forecasts. For every target year, the Blue Chip data contains monthly forecasts from 1-month horizon to 24-month horizon. Since the horizons span 2 years, we turn the data into current fiscal year forecasts and next year forecasts, and use both in our analysis. We also include withheld income and employment federal taxes, which have been used by Lewis et al. (2020). We also use Google Mobility data, which has been used by Fernández-Villaverde and Jones (2020) measure macroeconomic outcomes of COVID-19. Following Fernández-Villaverde and Jones (2020), we take the average of "Retail/Recreation" and "Workplace" and use it as a predictor. The daily Google mobility data is aggregated to monthly frequency. Since this variable only started after the pandemic began, it is used only to make forecasts and nowcasts for FY 2021. Other notable variables included are New York State Electricity Use, Travel in Millions of Vehicle Miles from U.S. Department of Transportation, and Truck/Auto traffic data from New York State Port Authority. Many of these variables have not been utilized before for forecasting analysis. These are all listed in Tables 2 and 3.

We use X-13-ARIMA-SEATS by U.S. Census Bureau with R package *seasonal* to do the seasonal adjustment. Since our target variable in terms of U.S. dollars is the 12-month sum of the monthly taxes, seasonal adjustment needs to be handled carefully. The fiscal-annual total values of the

tax series to be quite idiosyncratic compared to standard macroeconomic data. For instance, more than 10 principal components explain only 66% of the variation in the data set. See Delaney and Bugdayci (2016).

[6] Our experiments show that using the sum of the components is better than forecasting the components separately and then aggregating.

seasonally adjusted series in each fiscal year were constrained to be equal to the fiscal-annual total values of the original series. Tax data is further turned into percent change over the average of previous fiscal year. April, being the first month of the fiscal year in New York (most states begin their fiscal year on July 1), it is a special month that reports nearly 12% of annual receipts. The values of this month can change abruptly due to federal and state tax policy changes. Shifting of tax revenues between April and March can be one of the major reasons that lead to occasional abnormal April values. These outliers are controlled during the seasonal adjustment process.[7] Finally, all nominal values including the annual total tax receipts are turned into real values using CPI. All macro variables are seasonally adjusted, turned into percent change or log difference from one year ago whenever needed. The time span of our data is from April 1986 to March 2021. [8] Data from beginning to September 2006 is used as the initial training sample. The remaining data from October 2006 until March 2021 is used for evaluations.

In terms of information, the tax revenue data (MTax) of each month is usually published about 3 weeks after the end of the month. Thus the target variable of the fiscal year is know about 3 weeks after the fiscal year has ended. For the monthly macro variables, the publication date is usually one month after the end of the month. Some variables such as Blue Chip Real GDP forecast have publication lags shorter than one month. Some variables such as Transportation Service Index has publication lags longer than one month. Variables such as Personal Income of New York are generally lagged one-quarter. Due to the publication lags, our data set is ragged-edged. We first shift the quarterly variables that have long publication lags onward by one quarter to make sure the data is available when we generate out-of-sample forecasts. Other than this data availability issue, the model by Bańbura and Modugno (2014) described in Section 4.3 can get the factors with arbitrary types of missing data. In addition, since the newer data series such as Business Leaders Survey did not exist before 2004, we back-filled the missing data by multiplying the estimated factors by estimated factor loadings.

[7]For example, the value in April 2008 is abnormally large. This is due to a strong "settlement". A big amount of tax liabilities in FY 2007 were received in April 2008 as estimated taxes and final returns; see "Special Quarterly Report on the State Financial Plan Revenue Results and Trends-First Quarter State Fiscal Year 2008-09" by New York State Comptroller. When the financial crisis of 2007 came, investors sold assets and tax liabilities, generating capital gains. But these taxes were reflected at the beginning of FY 2009 in April 2008. Thus we use a dummy variable for this April as a holiday effect when doing seasonal adjustment.

[8]The time series of real annual growth in rate in NY Tax over 1971 - 2020 is compared with that of the federal taxes in Figure 1, where we find that their movements can be quite different at times, and the correlation between them is just 0.47. Both are stationary according to the standard Fuller-Dickey test.

4. Empirical Results

We generate direct multi-step (DMS) out-of-sample forecasts of real year-over-year growth in tax revenues for each horizon (h=18, 17,...,1). Thus, we develop an empirical model for each horizon since the specifications and lag structures change over horizons. Our approach does not require prediction of the right hand side variables over the prediction periods. By contrast, Ghysels et al. (2020) followed the iterative multi-period scheme (IMS) based on one estimated model for the two horizons they considered. To evaluate forecast performance, we make out-of-sample forecasts recursively using the rolling cross validation approach. Data from April 1986 to September 2006 are used as the initial training sample. Remaining data until March 2021 are used for out-of-sample forecasting evaluations. Thus forecast targets range from FY 2008 to FY 2021, and the standard deviation of the annual values was 4.05. Starting from the initial training sample, all currently available data are used to make forecasts for the target fiscal years. Then data for one more month is included, and forecasts are generated again for the same target. We go through the boosting or other machine learning procedures repeatedly to make forecast(s) for the same target fiscal year, but with models of shorter horizons sequentially. As we do this, the horizon gets shorter by one month at a time for the same fiscal year, which will also change every 18 rounds of monthly forecasts as we move forward. As a result, the training data set expands when more months or quarters of data are added. The forecasts stop when the sample fully exhausts. Note that since our forecast horizons range from 18 months to 1 month, current-year nowcasts and next-year forecasts are sometimes made simultaneously (i.e. forecasts of horizons 6 and 18, 5 and 17, 4 and 16, 3 and 15, 2 and 14, 1 and 13 are made at the same points of time).

To reiterate, There are a few newly created monthly variables which have missing historical values. More generally, our algorithm handles missing data of various types. The missing data of all such variables are imputed with estimated $\hat{\Lambda}\hat{f}_t$ from Equation (8), provided monthly factors are not used as the predictors.

Unless otherwise stated, monthly or quarterly lags of the macro variables fed to boosting procedures go back till the first month of the previous fiscal year. For example, for a 6-month-ahead nowcast, the lags of each variable go from lag 6 to lag 23. In terms of choosing the number of iterations for boosting, AIC is used as a guide.[9] In our experiments, the iteration number at each horizon did

[9] Although not frequently used in time series, we also tried k-fold cross validation. The numbers of iterations chosen were generally smaller, but the out-of-sample forecasts did not differ significantly from results when iteration

not change much across different options for boosting. The out-of-sample forecast accuracy also did not change much when M is between 20 and 50. Thus we use the same M for all boosting options at each horizon, and M is chosen to be between 20 and 50, cf.Wohlrabe and Buchen (2014). Notwithstanding the fact that budget managers may not treat revenue forecasts as an assertion on what will happen in the future, but rather as a guide for public management, we use the usual root mean square error (RMSE) for evaluation. It depends on h and is computed as

$$RMSE_h = \sqrt{\frac{1}{(T-h-T_0)/12+1} \sum_{t=T_0,T_0+12,\ldots}^{T-h} (y_{t+h} - \hat{y}_{t+h})^2} \tag{13}$$

RMSE ratio is obtained by dividing the out-of-sample RMSE of a particular model under consideration by the out-of-sample RMSE of the benchmark linear model for each horizon.

4.1. Main Results with Boosting

The results of RMSE ratios are reported in Tables 4 and 5. The machine learning models in the two tables are the same, but the benchmark model is random walk (with drift) in Table 4 and ADL-MIDAS in Table 5. In both tables, column 1 shows the boosting results when all variables and their lags are simultaneously fed into boosting procedure. As explained earlier, the dimension here is large since there are more than 30 high-frequency variables, each of which has many lags. Column 2 shows the boosting results when the first two extracted factors from macro leading variables together with monthly and annual taxes are used. Column 3 shows the boosting results when lags from all variables are simultaneously fed into boosting procedure when the lags of high-frequency macro variables are restricted to be the most recent ones (i.e. $J_M = 1$ and $J_Q = 1$ for the macro variables), but the number lags of the MTax is not restricted. Column 4 shows results of LASSO with all variables and their lags. Column 5 shows results of sg-LASSO with all variables and their lags. In all options, MTax and two yearly lags of the dependent variable were always used as predictors which were not used to extract factors. P-values of Diebold-Mariano tests (Diebold and Mariano, 1995) are reported in the parentheses below the ratios.

We find in Table 4 that all mixed-frequency machine learning models beat random walk model

numbers were a little larger.

(with drift) at all horizons. Sometimes, the RMSE ratio is even below 0.1. The improvement is due to both better methodologies and richer information in the data. For more meaningful analysis, we focus on Table 5, whose benchmark model is more informative. We find that boosting improves upon the benchmark at horizons from 3 to 10, no matter which option is used (except sg-LASSO at horizon 9). The ratios can be as low as 0.5 or even lower. The Diebold-Mariano tests with the null that the benchmark model has higher accuracy can be easily rejected at many of the horizons. The monthly tax variable or MTax is the main high-frequency predictor that contributes to the nowcasts inside the fiscal year. To see this, we refer to the short-horizon rows in Table 6, which list the selected lags at specific horizons while boosting with all variables. We can observe that the first selected lags at short horizons are all from MTax. The reason why boosting outperforms linear benchmark model at short horizons is that not all monthly data releases are equally helpful. For example, lags 6, 8, and 10 of MTax are not selected at horizon 4 even when they are available. The results are similar in Table 7, which lists the selected lags at specific horizons while boosting with factors. To see this more clearly, we turn the selected lags and their coefficients into implied lag distributions for MTax in Figure 2 with horizon 4 and 6 as examples. The distributions are not only different from uniform lag distribution used in our benchmark ADL-MIDAS, but also different from regular lag distributions controlled by parameters in MIDAS models. These choices are consistent with the fact that many big receipts like estimated payments, final payments and refunds are received on a quarterly schedule. The fact that selected lags vary over horizons underscores the need to specify the direct multi-step (DMS) model for each horizon separately - a multi-step iterative scheme (MIS) will not accommodate such heterogeneities.[10]

Most remarkably, boosting results with factors extracted from the macro variables reported in column 2 (Table 5) are better than boosting with all the variables (column 1) for horizons 18 - 8; at horizons 7 - 1, they are very similar. At long horizons, the RMSE ratios with factors are around 0.7, but they are around 1.0 when boosted with all variables.

Note that there are few boosting options where there is no significant improvement at horizons 12 and 11 compared to the benchmark. One of the reasons is that the fiscal year changes as we move from horizon 12 (March data) to horizon 11 (April data), where the uncertainty can be high in certain years. For example, part of the tax receipts collected are sometimes moved between March

[10] A study about direct and iterated multi-step conditional forecasts has been made by McCracken and McGillicuddy (2019).

and April due to changes in taxpayer behavior as reactions to federal and state tax law changes. The recent passage of the Tax Cuts and Jobs Act of 2017 (TCJA) created a huge incentive for some tax payers to shift income between tax years and delay estimated quarterly tax payments into the extension and final payment period. These issues make state tax revenue forecasting challenging and different compared to forecasting using conventional macroeconomic variables. Another reason is our benchmark model with fixed-target forecasts takes advantage of information from actual monthly tax collections that directly resolves the uncertainty of the total current-year taxes, which is not the same as models based on low-frequency values.

Surprisingly the RMSE ratio is uniformly large at Horizons 1 and 2, meaning that boosting is much worse than a linear MIDAS model (11) at these horizons. The explanation is straightforward: At very short horizons, the uncertainty associated with the remaining tax revenues for the tax year given the total over last 10 or 11 months is very little. The simple linear model that takes current-year monthly average of the taxes as a predictor is very powerful and hard to beat with little uncertainty. Boosting, by opening up to many lagged taxes that are often multicollinear, only adds unnecessary complication. We should point out, however, that although the ratios are large at Horizons 1 and 2, the RMSE values from the five machine learning models in Table 5 are not large - rather the RMSE of the benchmark model is very close to zero. In addition, the need for forecasting at horizons 1 or 2 is little because the target is practically known at this horizon.

The selected lags of variables chosen by boosting, reported in Tables 6 and 7, reveal some interesting information about the role and timing of high-frequency data releases. For example, the New York State Initial Claims is always selected at longer horizons. It is the first selected variable at Horizons 11-14. The Business Confidence Index, ISM indexes and housing variables are also frequently selected. After entering the target fiscal year when horizon is less than 12, the monthly taxes of that fiscal year start to be selected and are selected more frequently as the horizon gets shorter. This is because these monthly taxes add up to the target value which we are predicting. Also, MTax_7 is always the first selected lag after the horizon reaches 7, which means the monthly tax release in August is centrally important for nowcasting current-year tax revenues. This is also true for October and November tax values. Needless to say, these boosting results are not accidental, but are rooted in the institution. As for the selected lags while boosting with lagged values of factors, the overall situation is similar in the sense that only a few lags are selected. It also shows

lags 14 and 15 of the first factor are more important at long horizons outside the current year. In addition, the second factor (F2) is less frequently selected compared with the first factor, which is consistent with intuition.

In column 3 of Table 5 where the lags of high-frequency macro variables are restricted to be the most recent ones in boosting (i.e. $J_M = 1$ and $J_Q = 1$ for the macro variables) , we find that although the accuracy is not as good as boosting with factors, improvements can be found except for horizons 1 and 2 at most of the short and long horizons over benchmark. This option can be important if we see sharp structural breaks or turning points in the sample such that the most up-to-date information can help to track the sharp changes. For example, when boosting with all variables at horizon 12, none of the first selected lags of macro variables is the most recent, which is lag Number 12, or value in March), see Table 6. Note, one of the selected lags is NYICLAIMS_14, which is two-month-old New York State Initial Claims data from January. This may not be ideal for real-time forecasts facing abrupt structural breaks. Also, state tax revenues are affected greatly by recessions, which occurs every 8-10 years. The information related to recessions contained in the past history of the series may not be rich enough for machine learning if an upcoming recession or a structural break is triggered by factors that did not occur in the past (e.g., pandemic-related economic shut down). Thus if the optimal lags of macro variables are not appropriately selected in a specific context, it may be better to let boosting select among their most recent observations to take advantage of the up-to-date information. Since we are imposing the restriction only on macro indicators, this option is expected to be useful at long horizons longer than 11. The COVID-19 related recession, which occurred early in our multi-horizon forecasting scheme, can be one such case. The use of the latest values to forecast turning points follows the long tradition of the leading indicators approach where only contemporaneous values of the leading indicators are used. Thus, in addition to boosting with factors, we will also study the usefulness of this option when forecasting 2021 fiscal year revenue after the COVID-19 lockdown.

We plot the boosting forecasts and nowcasts for each of the fiscal years 2008 to 2020 with factors (i.e., column 3 of Table 5) in Figure 3, together with the MIDAS benchmark forecasts and nowcasts based on equation (11). The exercises were conducted in real time where the seasonal filters and factors were calculated every month recursively. The overall performance of boosted forecasts is better than the performance of the benchmark forecasts. The boosting forecasts and nowcasts ap-

proach the actual values well most of the time except at the shortest horizons of a few years. One dominant reason is (see Table 7) that at very short horizons the latest monthly tax information is not always chosen by any one of the boosting options. Ng (2014) and Babii et al. (2021) have also noted this problem with machine learning - the procedure does not recognize the natural ordering of successive lags in a time series variable. In order to circumvent this problem in the context of fixed target forecasting like ours, a natural remedy may be to restrict the lag to be the most recent lag value. But the cost of this restriction seems to be high at other horizons. In addition, we clearly see in Figure 3 the tremendous contribution of the macro variables in boosting long-horizon forecasts in many of the years including the recession year of FY 2010. In their Pew Center report, Boyd et al. (2011, p. 44) cite an interesting episode where NY DOB revised its budget shortfall estimate for FY 2010 five times - what started at $2.2 billion in April 2009 became $9 billion in March 2010 with the largest jump of $6.7 billion announced on July 2009. By comparing these series of NY DOB revisions with our forecasts in Figure 3 for FY 2010, it is readily seen that our model anticipated these shortfalls much earlier. We forecast quite correctly the final revenue growth of FY 2010. The boosted forecasts has this timing advantage because of its built in macro leading indicators foreseeing recessions like that of 12/2007-06/2009.

4.2. Results from Other models

In this section, we briefly discuss the forecasts from other models as robustness checks. First, in Table 5, we see that the RMSE ratios for LASSO and sg-LASSO also beat the benchmark most of the time, but they were not the best. Second, we experimented with sg-LASSO/boosting using Legendre polynomials. The results are consistent with our main findings that it beats the benchmark, but our results based on factors is still the best. Further, the tree base-learner or P-splines base-learner in boosting did not yield encouraging results. The explanation could be that there is not much non-linearity in the relationship between the predictors and the target variable, and hence there is no need to make the model unnecessarily complicated. Finally, an autoregressive benchmark was also tried and is found to be worse than the ADL-MIDAS benchmark.

4.3. Relative Importance of Selected Variable Lags

One of the benefits of gradient boosting is that the relative importance of each selected variable can be studied. The idea has been implemented by Taieb and Hyndman (2014), where the

contribution of one selected variable is removed from the model at a time. Adapting their method to our model, $\hat{\Psi}(\tilde{z}_{t+\omega})$ in Section 2.2 can be written as

$$\begin{aligned}\hat{\Psi}(\tilde{z}_{t+\omega}) =& \hat{\Psi}^{(0)}(\tilde{z}_{t+\omega}) + \sum_{n=1}^{N} \nu \hat{h}_n(\tilde{z}_{\hat{S}_n,t+\omega}) \\ =& \hat{\Psi}^{(0)}(\tilde{z}_{t+\omega}) + \sum_{j\in\{1,\ldots M\}} \sum_{\{n:\hat{S}_n=j\}} \nu \hat{h}_n(\tilde{z}_{\hat{S}_n,t+\omega}) \\ \equiv& \hat{\Psi}^{(0)}(\tilde{z}_{t+\omega}) + \sum_{j\in\{1,\ldots M\}} \hat{H}_n(\tilde{z}_{j,t+\omega}),\end{aligned} \tag{14}$$

where $\hat{H}_n(\tilde{z}_{j,t+\omega})$ is the contribution of each selected variable, recalling that M is the number of predictors including all low and high-frequency lags.

Next define forecasts removing the contribution of selected predictor i:

$$\hat{\Psi}_{-i}(\tilde{z}_{t+\omega}) = \hat{\Psi}^{(0)}(\tilde{z}_{t+\omega}) + \sum_{j\in\{1,\ldots M\}\setminus\{i\}} \hat{H}_n(\tilde{z}_{j,t+\omega}). \tag{15}$$

Thus the sum of squared error of such forecasts is

$$E^{-i} = \sum_{t=12,24,\ldots}^{T-h} (y^n_{t+h} - \hat{\Psi}_{-i}(\tilde{z}_{t+\omega}))^2. \tag{16}$$

The relative importance of this selected predictor is

$$I_i = \frac{E^{-i} - E}{E} \tag{17}$$

where E is the sum of squared errors for $\hat{\Psi}(\tilde{z}_{t+\omega})$:

$$E = \sum_{t=12,24,\ldots}^{T-h} (y^n_{t+h} - \hat{\Psi}(\tilde{z}_{t+\omega}))^2. \tag{18}$$

The degree of relative importance of a selected variable is large if removing its contribution makes the forecasts significantly less accurate. We compute relative importance for the model where monthly-tax lags and most-recent macro variables are fed into boosting procedure together. Since selected variables may change when doing out-of-sample forecasts and our focus here is on the

selected variables, we compute the relative importance based on in-sample forecasts over the whole sample as a summary measure. The results for all horizons are plotted in Figure 4. Since only a few variables and lags are selected at some horizons, we only report top-3 important ones. We find that the seventh lag of the monthly tax variable ranks first even at horizons shorter than 7, when taxes of more recent months are available and known. This partly explains why, at horizons 1 and 2, boosting does not outperform benchmark model that uses averaged monthly taxes as the predictor. On the other hand, at long horizons, New York Initial Claims, Continued Claims, Business Confidence Index, and some of the ISM indexes are relatively more important.

4.4. Information Efficiency of Boosted Forecasts

An important issue is whether the boosting forecasts generated by the 18 horizon-specific models are informationally efficient, in the sense that as the new revised forecasts are made every month for the same target year, they reflect all relevant information at the time each forecast is made. Nordhaus (1987) suggested a test for such a hypothesis when a number of multi-period forecasts are available for the same target. Since the exact specifications of our forecasting models get different as horizons change, this test will also confirm if each boosted specification is efficient. That is, at each horizon on the average over the sample, if boosting utilizes new information efficiently while revising the previous forecast. Nordhaus (1987) showed that under the null of efficiency, successive forecast revisions for the same target will be uncorrelated. Note that unpredictable forecast revisions would imply that fixed-target forecasts should behave like a random walk (or a martingale) – revisions should be spiky and not smooth. Since we have a series of 18 fixed-target forecasts with 13 separate out-of-sample targets, this test is particularly instructive and powerful in our context.

Ghysels et al. (2020) tested for unbiasedness of the forecasts generated by their MIDAS model pooling 48 states with 2 horizons. There are a number of studies that have examined if the government revenue forecasts are unbiased and efficient, and reached conflicting results depending on the sample. Due to the balanced budget rule, one may speculate that budget offices may have asymmetric loss function such that forecasts will tend to underestimate the actual values, see Krol (2013) and references therein. However, Pew Center-Rockefeller Institute Report (2011) found a clear pattern of positive (actual - forecast) errors during expansions and negative errors during

downturns. These government forecasts, however, embody political and bureaucratic biases. What we report below and those in Ghysels et al. (2020) are based on transparent data-based models without reflecting subjective factors.

Define ${}_{t+\omega}\hat{y}_{t+h}$ as forecast for y_{t+h} made at $t+\omega$, where $t = 12, 24, 36\ldots$, $h = 12$ or 24, $\omega = 0, 1, 2, \ldots, 11$. h and ω are such that the forecast horizons $h - \omega = 1, \ldots, 18$. Define forecast revision ${}_{t+\omega}\hat{v}_{t+h}$ as:

$$ {}_{t+\omega}\hat{v}_{t+h} \equiv {}_{t+\omega}\,\hat{y}_{t+h} - {}_{t-1+\omega}\,\hat{y}_{t+h} \tag{19} $$

Nordhaus (1987) suggests testing $\beta = 0$ in the regression:

$$ {}_{t+\omega}\hat{v}_{t+h} = \alpha + \beta \times {}_{t-1+\omega}\,\hat{v}_{t+h} + u_{t+\omega} \tag{20} $$

Note that due to taking a lag and the difference, $h - \omega = 1, \ldots, 16$ when running the regression, and hence the total number of observations in the regression is 208. A simple OLS regression with the stacked data yielded the following estimates: $\hat{\alpha} = 0.019$ (p-value=0.75) and $\hat{\beta} = -0.079$ (p-value=0.25). Given the reported p-values, we conclude that the coefficients are statistically not different from zero at any reasonable level of significance. Thus, the evidence is overwhelming that the boosted forecasts are informationally efficient at each horizon.

5. Forecasts in the aftermath of COVID-19

5.1. Tax Revenue Forecasts for Fiscal Year 2021

A cataclysmic event like the COVID-19 pandemic will make any forecasting model to fail miserably. However, our model with high-frequency data is based on machine learning, which is known to accommodate abrupt and non-linear variations in the target variable. In addition, the use of high-frequency data can potentially identify the breaks promptly, particularly because our macro variables were specifically selected with leading indicator properties. We have been tracking the forecasting power of the boosting model from 2008 recursively in real time and found that it performed quite well in forecasting abrupt annual variations in revenues like the 2009 economic recession. But the challenge of predicting the short term economic effects of the COVID-19 is

unprecedented and ought to be formidable.

The economic shutdown of New York in March 2020 threw a monkey wrench on the New York Budget making process. Annual budget negotiations on revenue forecasting began as usual from October 2019 with the so-called Albany triad - the Governor, the Assembly and the Senate - that culminated in the Consensus Forecasting Conference on February 27, 2020 for final passage of the budget in March 2020.[11]. New York officially enacted its FY 2021 budget in April 2020. However, the complete lockdown of the economy that converted the vibrant New York city into a ghost town and the resultant uncertainty with revenues rendered the enacted budget useless.

New York Governor Andrew Cuomo issued a stay-at-home order on March 22, 2020. It continued for 54 days, making it the seventh-longest shutdown in the nation. Within a month, the budget analysts projected a 13.3 billion budget shortfall as a result of the free fall of tax revenues due to COVID-19. As immediate stop gap measures, aids to localities were postponed to the tune of $8 billion, and a number of across-the-board cuts, hiring freeze, deferred salary increases, etc. were instituted. The State also got $4.5 billion by PIT note sales for helping with the cash flow during the year. As of November 20, 2020, New York being the epicenter of the pandemic had reported nearly 300,000 total COVID-19 cases, and nearly 530,000 continued unemployment insurance claims, ranking third-highest in the nation. Since January 2020, New York's private employment declined by 12.8%. Most striking, the leisure and hospitality sector experienced the greatest percentage decline in employment across the state, reporting 37.9% fewer jobs since January. Not surprisingly, many of the cyclical indicators reacted immediately to the shock, and declined precipitously. The LEI declined in a month by 7.4% in March 2020, and the initial unemployment claims increased by more than 2500% in New York state. Psychologically, it was widely expected that a recession would start with the COVID-19 outbreak, particularly in New York and the U.S. Northeast.

The forecasting model managed by NY DOB is based on a large number of fragmented equations that are primarily estimated using quarterly and annual observations. Even though its structural model is well spelled out with impressive details, it does not have the capability of monitoring the incoming revenues in conjunction with other macroeconomic developments at high frequencies. As Dadayan (2020) pointed out, state revenue forecasting across the country faced extraordinary

[11] Chris Varvares of Macroeconomic Advisors, Hugh Johnson of Johnson Advisors LLC, Jason Bram of New York Fed and Kajal Lahiri of SUNY Albany testified at the Conference in the presence of the Budget Director, Deputy Controller, Ranking members of the Senate, Ways & Means Finance committees, and other stakeholders on the reasonableness of the executive budget and its underlying assumptions

challenges because of the pandemic related uncertainty and partial lock downs. The frequent updates of yearly revenue growth with the latest information was the only hope for the policy makers to monitor the revenue receipts in real time as the fiscal year unfolded towards the mid-year adjustment. Currently, no such monthly model exists either for New York or any other sub-national entities in the U.S..

Forecasts and nowcasts from our two preferred models, viz., (1) boosting with most recent macro variables and lags of monthly tax variable, and (2) boosting with lags of factors and lags of the monthly tax variable for FY 2021 are plotted in two panels of Figure 5, showing how successive forecasts were updated as new data in each month became available in real time. Our forecasting cycle began in 9/2019 - 6 months before the start of FY 2021 (or 18 months before the end of it), and by 5/2021 all forecasts and nowcasts at all horizons for the FY 2021 were completed. Since the macro indicators were already showing signs of economic slow down in late 2019, our long-horizon forecasts for Fiscal Year 2021 were around 0% growth. These forecasts, though not significantly negative, were much lower than the forecasts made by different governmental agencies (e.g, NY DOB, Senate Finance Committee, Senate Minority Ways and Means, and Ways and Means Minority), and were uniformly in excess of 5% in real terms, see Table 9. Thus, whereas none of these agency forecasters saw slowing down of the revenues, our model due to its reliance of business cycle variables could anticipate it correctly. As the National Bureau of Economic Research (NBER) Business Cycle Dating Committee explained in its announcement of June 8, 2020, the latest recession started in February 2020 before the COVID-19 lockdown due to the softening of the labor market, but was precipitated by the pandemic-related lockdowns.

As seen in Figure 5, our boosting model (i,e., model 1 above) with latest macro variables promptly revised its forecast from around -0.5% in the beginning months of the calendar year 2020 to -3.1% with March data, the first full month affected by New York's restrictions on normal economic activities when the private payroll employment in the state collapsed to 1.8 million below the prior year's level. The major selected macro variables that pulled the forecast down include Withheld Income and Employment Federal Taxes, Continued Claims, Personal Consumption Expenditures, and lagged dependent variable, see Table 10. The nowcasts continued to drop even more sharply to end up at -16.6% with the April data. This dramatic decline is supported by the fact that between April 2019 and April 2020, New York tax revenue fell by 70%. This is reflected in the nowcast since

MTax_11, which represents tax receipts in April 2020, is one of the first selected variables, see again Table 10. The New York employment fell by 18.8% between 2019Q3 and 2020Q3 - the third worst decline amongst all states in the US. The nowcasts then started to improve to -10% with the arrival of May data, as many of the macro indicators and the monthly tax variable improved with signs of recovery. The trough of April 2020 in the tax data coincided exactly with the same as announced by NBER on July 19, 2021 regarding the end of the latest recession that started in February 2020. With July data, there was another large upward revision that made the nowcast become positive and stayed above -5% afterwards. The surge of July 2020 was due to the postponement of filing deadline of 2019 federal income tax return till July 15, 2020. The updated nowcast made in March 2021 stood at -1%, which is slightly above the actual value -1.6%.

The boosting nowcasts using factors (i.e., model 2) followed the previous model 1 with a lag of one month and went down to -2% with the April 2020 data update, and was revised downward to -8.9% with May data. Thus, our boosting model (2) with factors that may not necessarily use the latest lags of macro variables performed similar to the restricted model (1) in generating year-over-year revenue growth, but with less abrupt initial revision and a little lagged responses by a month. As the horizon got shorter, both models with increasing certainty issued very similar nowcasts, which were above -5% and close to actual value in the end. Subsequently, the receipt of $5.1 billion under the Coronavirus Aid, Relief and Economic Security (CARES) Act, a steady growth of income and sales taxes (including on-line sales), and the surge of tax returns in July due to the deferment of the tax filing deadline partially mitigated the initial projection of revenue drop. It is noteworthy that our boosting nowcasts are significantly higher than those of NY DOB, which has not revised its tax revenue projections since its first quarter update, and stayed between -11% and -12% in its FY 2021 Mid-year Update in November. This dire projection was balanced in the budget with an increasing expected federal grants that increased by $20.9 billion from the enacted Executive Budget of April 2020 to the Mid-year Update in November 2020.

At the end of FY 2020, New York state had about $2.5 billion rainy day funds, which is only 3.2% of its total state expenditures. The latter figure for California and Texas are, for instance, close to 13%, see Clemens and Veuger (2020). However, the 2021 Executive Budget expected State's total primary cash reserves including rainy day funds, which can be tapped in an emergency, will be nearly $6 billion at the close of FY 2021. But under the balanced budget guideline, these funds

have to be restored back to the previous year's level within the next fiscal year. Since the budget short falls due to COVID-19 is expected to be last well beyond one year, borrowing from rainy day funds was of little help to the states. Thus, without additional grants from the federal government, it was impossible for New York state to mitigate the monumental projected shortfall for FY 2021 without draconian spending cuts, tax increases or borrowing. It can not be argued that the FY 2021 estimated short fall in New York was due to mismanagement of the budget from previous years. Even though New York state has been notorious in having late budgets over a prolonged period during 1988-2007 (cf. Andersen et al., 2012), every year since 2010, it has successfully enacted and executed timely balanced budgets.

5.2. Comparison with recent studies

Given the importance of forecasting the tax revenues for the current and the next fiscal years during this uncertain period, a number of studies have produced informed estimates taking different approaches. Ghysels et al. (2020) is closest to our approach in terms of forecasting state revenues with high-frequency data in real time. They produce revenue and spending forecasts for 48 contagious states using Census Bureau's quarterly data supplemented by similar information from NASBO during 2014-2019. Their forecasts for real revenue growth during 2020 and 2021 with data up to April 2020 were -2.46% and 10.62% - these are much different from our real-time forecasts for FY 2021 made in April 2020. However, given the model specification, variables used and the assumptions about the future path of the exogenous variables, their estimates are not comparable to our forecasts. In particular, their scheme does not utilize the monthly taxes inside the target year in recursive forecasting. In addition to not requiring any assumption on the future path of the exogenous variables, our DMS forecasts utilize the accumulated monthly tax revenues inside the target year as major determinants of the yearly tax revenue growth on a monthly basis in real time.

Using the same Census Bureau quarterly tax revenue data base, Zhao (2020) uses a simple tax revenue model as a function of state unemployment rate and trend to forecast percentage change in real state tax revenue for six New England states under three different assumed scenarios (low, mid and high) for the future growth in unemployment rates over the forecasting period. For Massachusetts, which is possibly closest to New York in terms of diversity and size, the predicted

declines were -10.50%, -19.30% and -35.21% respectively, depending on the three unemployment scenarios assumed. Thus, Zhao's (2020) projection with the middle unemployment scenario comes close to our 2021 projection made around April 2020.

Taking a slightly different approach, Clemens and Veuger (2020) also assessed monetary shocks to the state government sales and income tax revenues during the initial months of COVID-19 pandemic. They used forecast revisions to personal income and consumption from the January and May Congressional Budget Office (CBO) reports as measures of changes in tax base, and revenue elasticities from the existing literature to find that the pandemic would reduce total U.S. states' tax collections by about $42 billion in 2020Q2 and an additional $106 billion over 2020Q3-2021Q2. We applied their methodology to New York State and obtained -4.63% as the forecast of total tax shortfall over FY 2021. We differed from Clemens and Veuger (2020) in three important ways. First, we used estimates of the relevant elasticities from the Budget Methodology Book of NY DOB (2019) that were very similar to those reported by Anderson and Shimul (2018). Second, rather than using CBO forecasts from January and May 2020 to measure the incidence of COVID-19, we used Blue Chip forecasts from March and May 2020 newsletters to estimate more precisely the net effect of COVID-19 on forecasts for personal income and consumption. Compared to CBO, the Blue Chip gave a much higher forecast for personal income in its May newsletter, possibly because CBO produces current law forecasts, and hence might not have incorporated the impact of the CARES Act and many other income enhancing future legislations. Third, we do not compute the counterfactual tax base values or tax shortfall level for FY 2021 since we only look at the percentage change. Tax shortfall level may vary when different assumptions are used to get the counterfactual tax base values.

We should point out that these timely studies, like most in this area, use the revenue figures from the Annual Survey of State and Local Government Finances produced by the U.S. Census Bureau. This harmonized all-state data set is very useful in analyzing the finances of all states together. But one important drawback, apart from being quarterly, is that these are released three months behind the state agencies publications, and hence less suitable for real time nowcasting/forecasting. In the absence of monthly tax receipts, in Ghysels et al. (2020), revenue short falls are driven by only one state specific variable (quarterly personal income with a lag) and 11 monthly macro and financial variables. In Zhao (2020) state unemployment rate is the only primitive variable, and in

Clemens and Veuger (2020) personal income and consumption are the drivers for PIT and sales tax. By contrast, our model incorporates not only monthly tax receipts, but also a host of New York and U.S. macroeconomic variables that boosting ensembles optimally.

Note that our monthly forecasting model for FY 2021 as the target is uniquely positioned to calculate effect of COVID-19 shutdown on state revenue counterfactually by subtracting the real-time nowcast of model (1) made with data till February 2020 from the nowcast of model (1) with data till April 2020. This approach yields -16% as our tax shortfall due to the pandemic in terms of annual percentage change from previous fiscal year, which is consistent with the revenue forecast we reported before.

5.3. Further Budgetary Use of Boosted Nowcasts for Fiscal Surveillance

Since the monthly tax variable is directly part of the target variable in current fiscal year, it is possible to turn nowcasts (i.e., at horizons shorter than 12) of annual total growths into estimated tax revenue to be collected in the remaining months. This is an important task the budget managers face to avoid revenue shortfall as the fiscal year comes close to its end constrained by a balanced budget rule. To do this, the nowcast of real annual growth is turned into nominal growth first, using CPI forecast from Blue Chip forecasts. The nowcast of nominal annual tax receipts is easily obtained since previous fiscal year value is known. The estimated revenue to be collected in the remaining months is computed by subtracting the sum of monthly taxes already collected in the current fiscal year from the total fiscal year nowcast amount. An illustration of such fiscal surveillance for the FY 2021 is reported in Table 8. For instance, in 10/2020, the nowcast for the tax revenues from October 2020 to March 2021 is $40.77 billion. Since the actual value of tax revenues in that period is $41.68, nowcast shortfall is $0.91 billion. This nowcast back-calculations are useful since it provides budget analysts a handy target number to be accomplished within the fiscal year, e.g., during mid-year budget updates.

6. Concluding Remarks

In this paper we have developed a mixed-frequency model to forecast state tax revenues using a number of machine learning techniques. The conventional revenue forecasting models have

been found to be particularly deficient during recessions that typically devastate state budgets. Also, they lack the capability of updating revenue streams in real time which necessitate the use of high-frequency data. In addition, state revenue models suffer from political influences, particularly due to the balance budget requirements that each state budget has to adhere. With these considerations in mind, we developed alternative machine learning models such that they are data driven, and require very little human judgment. In the current cataclysmic situation where the COVID-19 pandemic has dramatically affected the state economies and their budgets, our model could generate expected revenue shortfalls for the fiscal year in a transparent manner that is revised in real time as new information arrives at monthly frequencies.

In order to mimic the actual budget making process, we set out to forecast year-over-year growth in tax revenues using monthly and quarterly data over 2006-2020 beginning six months before the start of the fiscal year and ending when the fiscal year is over. Thus we generate a sequence of 18 monthly forecasts for each fiscal year. We have two broad sets of explanatory variables - the monthly taxes and a set of carefully selected leading macroeconomic indicators related to the U.S. and the New York economies. The latter set of macro indicators are brought in specifically to capture the effect of anticipated recessions on government revenues. We showed that for each target year and for each horizon the boosted forecasts are informationally efficient, performed better than an ADL-MIDAS benchmark, and the sequences of 18 multi-horizon forecasts converged to the final outturns fairly well.

We experimented with a number of boosting and LASSO algorithms to study which one would work best given our novel context and the data. We found that the conventional boosting with monthly lags of taxes and two macro common factors performed the best among other machine learning procedures. Since boosting may not chose the latest macroeconomic data, we also studied the advisability of restricting boosting to select the most recent macro variables to capture abrupt structural changes.

With this positive evidence in hand, we generated out-of-sample forecasts for fiscal year 2021 in the aftermath of the COVID-19 pandemic that upended all government budgets. In terms of real year-over-year growth, our boosting model with monthly tax variable and most recent macro variables revised its forecast from around -0.5% in the beginning months of the calendar year 2020 to an unprecedented -16.6% in May 2020. This dramatic decline is supported by the fact that in

April 2020, the yearly decline in New York tax revenue was 70%, and its total employment fell by 18.8% - the third worst decline amongst all states in the US. The nowcasts then started to improve to -10% with the arrival of the May data, as many of the macro indicators and the monthly tax variable improved with signs of continued recovery. The recovery in tax revenues is consistent with the recent NBER announcement that April 2020 is the last month of the U.S. recession that began in February 2020. With data till March 2021, our updated nowcast for revenue growth stood close to the year-ending actual value of -1.6%. It is noteworthy that NY Division of the Budget maintained their forecast for tax revenue growth between -11% and -12% throughout the period. Our forecasts correctly indicated a significantly better scenario for the FY 2021 tax revenues than those reported in the New York State Executive Budgets.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors appreciate helpful comments from two anonymous referees.

References

Andersen, A. L., Lassen, D. D., & Nielsen, L. H. W. (2012). Late budgets. *American Economic Journal: Economic Policy*, *4*, 1–40.

Anderson, J. E., & Shimul, S. N. (2018). State and local property, income, and sales tax elasticity: Estimates from dynamic heterogeneous panels. *National Tax Journal*, *71*, 521–546.

Andreou, E., Ghysels, E., & Kourtellos, A. (2013). Should macroeconomic forecasters use daily financial data and how? *Journal of Business & Economic Statistics*, *31*, 240–251.

Asimakopoulos, S., Paredes, J., & Warmedinger, T. (2020). Real-Time Fiscal Forecasting Using Mixed-Frequency Data. *The Scandinavian Journal of Economics*, *122*, 369–390.

Babii, A., Ghysels, E., & Striaukas, J. (2021). Machine learning time series regressions with an application to nowcasting. *Journal of Business & Economic Statistics*, (pp. 1–23).

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*, 191–221.

Bai, J., & Ng, S. (2009). Boosting diffusion indices. *Journal of Applied Econometrics*, *24*, 607–629.

Baicker, K., Clemens, J., & Singhal, M. (2012). The rise of the states: U.S. fiscal decentralization in the postwar period. *Journal of public Economics*, *96*, 1079–1091.

Bańbura, M., & Modugno, M. (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, *29*, 133–160.

Boyd, D. J., & Dadayan, L. (2014). *State tax revenue forecasting accuracy: Technical report*. Rockefeller Institute of Government. SUNY, Albany, NY. (Sept.).

Breiman, L. (1998). Arcing classifier (with discussion and a rejoinder by the author). *The annals of statistics*, *26*, 801–849.

Bühlmann, P., & Hothorn, T. (2007). Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science*, *22*, 477–505.

Bühlmann, P., & Yu, B. (2003). Boosting with the L_2 loss: regression and classification. *Journal of the American Statistical Association*, *98*, 324–339.

Buxton, E., Kriz, K., Cremeens, M., & Jay, K. (2019). An Auto Regressive Deep Learning Model for Sales Tax Forecasting from Multiple Short Time Series. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)* (pp. 1359–1364). IEEE.

Clemens, J., & Veuger, S. (2020). Implications of the COVID-19 pandemic for sate government tax revenues. *National Tax Journal*, *73*, 619–644.

Dadayan, L. (2020). State Revenue Forecasts Before COVID-19 and Directions Forward. Washington, DC: Urban Institute.

Davig, T., & Hall, A. S. (2019). Recession forecasting using Bayesian classification. *International Journal of Forecasting*, *35*, 848–867.

Delaney, J., & Bugdayci, O. (2016). Economic and Revenue Forecasting Model for New York State with Mixed-Frequency Data. Paper for MA Forecasting program, Department of Economics - University at Albany, SUNY.

Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic statistics*, *20*, 134–144.

Döpke, J., Fritsche, U., & Pierdzioch, C. (2017). Predicting recessions with boosted regression trees. *International Journal of Forecasting*, *33*, 745–759.

Eilers, P. H., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical science*, (pp. 89–102).

Fernández-Villaverde, J., & Jones, C. I. (2020). Macroeconomic outcomes and COVID-19: a progress report (No. w28004). National Bureau of Economic Research.

Foroni, C., Marcellino, M., & Schumacher, C. (2015). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *178*, 57–82.

Frankel, J. (2011). Over-optimism in forecasts by official budget agencies and its implications. *Oxford Review of Economic Policy*, *27*, 536–562.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, (pp.

1189–1232).

Ghysels, E., Grigoris, F., & Ozkan, N. (2020). Real-Time Forecasts of State and Local Government Budgets with an Application to COVID-19. Available at SSRN 3580363.

Ghysels, E., Kvedaras, V., Zemlys, V. et al. (2016). Mixed frequency data sampling regression models: the R package midasr. *Journal of statistical software*, *72*, 1–35.

Ghysels, E., & Ozkan, N. (2015). Real-time forecasting of the U.S. federal government budget: A simple mixed frequency data regression approach. *International Journal of Forecasting*, *31*, 1009–1020.

Ghysels, E., Sinko, A., & Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, *26*, 53–90.

Hofner, B., Mayr, A., Robinzonov, N., & Schmid, M. (2014). Model-based boosting in R: a hands-on tutorial using the R package mboost. *Computational statistics*, *29*, 3–35.

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical statistics*, *15*, 651–674.

Krol, R. (2013). Evaluating state revenue forecasting under a flexible loss function. *International Journal of Forecasting*, *29*, 282–289.

Lehrer, S., Xie, T., & Zeng, T. (2019). Does High-Frequency Social Media Data Improve Forecasts of Low-Frequency Consumer Confidence Measures? *Journal of Financial Econometrics*, .

Lewis, D., Mertens, K., & Stock, J. H. (2020). US economic activity during the early weeks of the SARS-Cov-2 outbreak (No. w26954). National Bureau of Economic Research.

Ludvigson, S. C., Ma, S., & Ng, S. (2021). COVID-19 and the Costs of Deadly Disasters. In *AEA Papers and Proceedings* (pp. 366–70). volume 111.

Makridis, C. A., & McNab, R. (2020). The Fiscal Cost of COVID-19: Evidence from the States. Mercatus COVID-19 Response Working Paper.

McCracken, M. W., & McGillicuddy, J. T. (2019). An empirical investigation of direct and iterated multistep conditional forecasts. *Journal of Applied Econometrics*, *34*, 181–204.

Megna, R., & Xu, Q. (2003). Forecasting the New York State economy: The coincident and leading indicators approach. *International Journal of Forecasting*, *19*, 701–713.

National Association of State Budget Officers NASBO (2013). *State Budgeting and Lessons Learned from the Economic Downturns: Analysis and Commentaries from State Budget Officers, Technical Report (Summer)*.

New York State Comptroller (2008). *Special Quarterly Report on the State Financial Plan Revenue Results and Trends-First Quarter State Fiscal Year 2008-09*.

New York State Division of the Budget (2019). *Budget Methodology Book*.

New York State Division of the Budget (2020a). *Executive Budget Financial Plan-Updated for Governor's Amendments and Forecast Revisions* .

New York State Division of the Budget (2020b). *FY 2021-Enacted Budget Financial Plan* .

New York State Division of the Budget (2020c). *FY 2021-First Quarterly Update*.

New York State Division of the Budget (2020d). *FY 2021 Mid-Year Update*.

Ng, S. (2014). Boosting recessions. *Canadian Journal of Economics/Revue canadienne d'économique*, *47*, 1–34.

Nordhaus, W. D. (1987). Forecasting efficiency: concepts and applications. *The Review of Economics and Statistics*, (pp. 667–674).

Onorante, L., Pedregal, D. J., Pérez, J. J., & Signorini, S. (2010). The usefulness of infra-annual government cash budgetary data for fiscal forecasting in the euro area. *Journal of Policy Modeling*, *32*, 98–119.

Paredes, J., Pedregal, D. J., & Pérez, J. J. (2014). Fiscal policy analysis in the euro area: Expanding the toolkit. *Journal of Policy Modeling*, *36*, 800–823.

Pedregal, D. J., & Pérez, J. J. (2010). Should quarterly government finance statistics be used for fiscal surveillance in Europe? *International Journal of Forecasting*, *26*, 794–807.

Pérez, J. J. (2007). Leading indicators for euro area government deficits. *International Journal of Forecasting*, *23*, 259–275.

Pew Center on States and Rockefeller Institute of Government (2011). States' Revenue Estimating: Cracks in the Crystal Ball. *Philadelphia, PA and Albany, NY. Pew Charitable Trusts and Rockefeller Institute of Government*, .

Piger, J. (2020). Turning Points and Classification. In *Macroeconomic Forecasting in the Era of Big Data* (pp. 585–624). Springer.

Rodgers, R., & Joyce, P. (1996). The effect of underforecasting on the accuracy of revenue forecasts by state governments. *Public Administration Review*, (pp. 48–56).

Simon, N., Friedman, J., Hastie, T., & Tibshirani, R. (2013). A sparse-group lasso. *Journal of computational and graphical statistics*, *22*, 231–245.

Sun, J. (2008). Forecast Evaluation: A Case Study. In J. Sun, & T. D. Lynch (Eds.), *Government Budget Forecasting: Theory and Practice* chapter 10. CRC Press.

Taieb, S. B., & Hyndman, R. J. (2014). A gradient boosting approach to the Kaggle load forecasting competition. *International journal of forecasting*, *30*, 382–394.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*, 267–288.

Vrontos, S. D., Galakis, J., & Vrontos, I. D. (2020). Modeling and predicting U.S. recessions using machine learning techniques. *International Journal of Forecasting*, . doi:https://doi.org/10.1016/j.ijforecast.2020.08.005.

Wohlrabe, K., & Buchen, T. (2014). Assessing the macroeconomic forecasting performance of boosting: evidence for the United States, the Euro area and Germany. *Journal of Forecasting*, *33*, 231–242.

Yousuf, K., & Ng, S. (2020). Boosting high dimensional predictive regressions with time varying parameters. *Journal of Econometrics*, . doi:https://doi.org/10.1016/j.jeconom.2020.08.003.

Zhao, B. (2020). Forecasting the New England States' Tax Revenues in the Time of the COVID-19 Pandemic. Current Policy Perspectives, Federal Reserve Bank of Boston Research Department.

Table 1: Individual taxes from NY Department of Tax and Finance

Original Name	Description
PITNet	Personal Income Tax
Sales	Sales Tax, Aggregate
Business	Business Tax, Aggregate
CorpMTANet	Corporate Metropolitan Transit Authority
MiscTax	Miscellaneous Tax
NYCAlcBev	New York City Alcoholic Beverage Tax
NYCStockX	New York City Stock Exchange Tax
TotStLoc	Net State and Local Receipts
TotalLocal	Total Locality Receipts
MotorFuel	Motor Fuel Tax
Cigarette	Cigarette Tax
AlcBev	Alcoholic Beverage Tax
Highway	Highway Tax
CorpUtil	Corporate Utility Tax
Banks	Banks Tax
Insur	Insurance Tax
PropGains	Real Property Gains Tax
EstGift	Estate nad Gift Tax
RETT	Real Estate Transfer Tax
PariM	Pari Mutuel
Racing	Racing Tax
CorpFran	Corporate Franchise Tax
Witholding	Witholding component of PIT
Estimated	Estimated Payments of PIT
Final	Final Payments of PIT
Delinquencies	Delinquencies of PIT
NycPit	New York City Personal Income Tax
NycSales	New York City Sales Tax
YonkersPIT	Yonkers Personal Income Tax
PITRef	Personal Income Tax Refunds
SalesRef	Sales Refunds
TotelRef	Total Refunds
MTax (monthly tax variable)	**Net Total Tax Receipts**

Table 2: U.S. Variables List: Part 1

Variable	Explanation	Frequency	Source
TCU	Capacity Utilization: Total Industry	Monthly	Federal Reserve Board of Governors
PersConsExpen_real	Personal Consumption Expenditures (real)	Monthly	US. Bureau of Economic Analysis
PersInc_real	Personal Income (real)	Monthly	US. Bureau of Economic Analysis
PersSaveR	Personal Saving Rate	Monthly	US. Bureau of Economic Analysis
LeadCompIndic	Leading Economic Index	Monthly	Conference Board
TSI_T	Transportation Service Index_total	Monthly	Bureau of Transportation Statistics
BlueC	Blue Chip Real GDP Forecast: Current Year	Monthly	Blue Chip Economic Indicators
BlueN	Blue Chip Real GDP Forecast: Next Year	Monthly	Blue Chip Economic Indicators
RECESS2	Probability of Decline in Real GNP/GDP in next quarter	Quarterly	Federal Reserve Bank of Philadelphia
SEN	Confidence Index: U.S.	Monthly	Organisation for Economic Co-operation and Development
EPU_3C	Overall Economic Policy Uncertainty Index	Monthly	Economic Policy Uncertainty
EMP_WRO	Total Employees: Whole Sales Trade, Retail Trade and Other Services	Monthly	Bureau of Labor Statistics
AAADGS10	Moody's Seasoned Baa Corporate Bond Minus 10-Year TreasuryConstant Maturity Rate	Monthly	Moody's; Board of Governors of the Federal Reserve System (US)
TVT	Travel in Millions of Vehicle Miles	Monthly	U.S. Department of Transportation
ISM_MEI	ISM Manufacturing Employment Index	Monthly	Institute for Supply Management
ISM_MII	ISM Manufacturing Inventory Index	Monthly	Institute for Supply Management
ISM_MPI	ISM Manufacturing Prices Index	Monthly	Institute for Supply Management
ISM_MNEOI	ISM Manufacturing New Export Orders Index	Monthly	Institute for Supply Management
ISM_MNIOI	ISM Manufacturing Imports Index	Monthly	Institute for Supply Management
ISM_NMNOI	ISM Non-Manufacturing New Orders Index	Monthly	Institute for Supply Management
ISM_NMEI	ISM Non-Manufacturing Employment Index	Monthly	Institute for Supply Management
FedTxWithold	Withheld Income and Employment Federal Taxes	Monthly	U.S. Treasury
CCLAIM	Continued Claims of Unemployment Insurance	Monthly	U.S. Employment and Training Administration

Table 3: New York and Regional Variables List: Part 2

Variable	Explanation	Frequency	Source
HX1M.CNER	Existing Home Sales: Single-Family Northeast Census Region	Monthly	National Association of Realtors
HX1MEDM.CNER_real	Median Sales Price Single Family Homes: Northeast Census Region (real)	Monthly	National Association of Realtors
HOFHOPIQ.NY	FHFA House Price Index, Existing Single-Family Homes: New York	Quarterly	U.S. Federal Housing Finance Agency
ELEC.NY	New York State Total Electric Generation	Monthly	New York Independent System Operator
NYOTOT_PC1_real	Personal Income (real): New York	Quarterly	U.S. Bureau of Economic Analysis
autotraf	Truck Traffic, Six PANYNJ Crossings	Monthly	New York State Port Authority
trucktraf	Auto Traffic, Six PANYNJ Crossings	Monthly	New York State Port Authority
SCI_NY	State coincident Index: New York	Monthly	Federal Reserve Bank of Philadelphia
RHPNRM.NY	Total Housing Permits: New York	Monthly	U.S. Census Bureau
NYICLAIMS	Initial Claims in New York	Monthly	U.S. Employment and Training Administration
BAFDINA	Business Leaders Survey: Future Business Activity	Monthly	Federal Reserve Bank of New York
NY_Bench	Early Benchmarked Employment: New York State	Monthly	Federal Reserve Bank of New York
ret_rec_work_NY	Community Mobility: Average of Retail/Recreation and Workplace-New York	Monthly	Google

Table 4: RMSE Ratios with Random Walk Model (with Drift) as Benchmark

Horizon	All Variables	**Factors**	**All Variables: Most Recent Macro**	LASSO	sg-LASSO
1	0.099	0.099	0.097	0.125	0.116
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
2	0.112	0.109	0.109	0.125	0.115
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
3	0.11	0.109	0.11	0.125	0.118
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
4	0.095	0.095	0.093	0.111	0.124
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
5	0.105	0.107	0.102	0.123	0.15
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
6	0.144	0.151	0.15	0.156	0.182
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)
7	0.145	0.151	0.147	0.154	0.173
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)
8	0.272	0.263	0.246	0.251	0.247
	(0.012)	(0.011)	(0.011)	(0.011)	(0.01)
9	0.353	0.317	0.317	0.319	0.359
	(0.011)	(0.011)	(0.011)	(0.011)	(0.012)
10	0.372	0.327	0.329	0.366	0.35
	(0.012)	(0.011)	(0.01)	(0.013)	(0.012)
11	0.601	0.466	0.499	0.66	0.587
	(0.04)	(0.016)	(0.022)	(0.094)	(0.046)
12	0.655	0.531	0.574	0.689	0.634
	(0.058)	(0.018)	(0.019)	(0.064)	(0.05)
13	0.852	0.62	0.683	0.77	0.789
	(0.301)	(0.089)	(0.111)	(0.174)	(0.213)
14	0.851	0.617	0.732	0.797	0.849
	(0.3)	(0.088)	(0.163)	(0.224)	(0.31)
15	0.837	0.611	0.739	0.758	0.852
	(0.28)	(0.083)	(0.176)	(0.21)	(0.314)
16	0.83	0.605	0.917	0.986	0.912
	(0.263)	(0.078)	(0.391)	(0.48)	(0.379)
17	0.772	0.59	0.741	0.812	0.851
	(0.2)	(0.073)	(0.187)	(0.253)	(0.299)
18	0.774	0.6	0.749	0.752	0.795
	(0.175)	(0.07)	(0.171)	(0.179)	(0.237)

Note: Column 1 shows the results for boosting with all variable lags directly. Column 2 shows the results for boosting when the first two extracted factors from macro leading variables together with monthly and annual taxes are used. Column 3 shows the results for boosting with all variables, but restricting high-frequency macro variables to be most recent. Column 4 shows the results for LASSO with all variable lags. Column 5 shows the results for sg-LASSO with all variable lags, where all high-frequency lags of the same variable are in one group. The values in parentheses are p values of one-sided Diebold-Mariano tests with null that the benchmark forecast is better. Methods in column 2 and 3 will be focused on when forecasting in COVID-19.

Table 5: RMSE Ratios with ADL-MIDAS Benchmark Model

Horizon	All Variables	**Factors**	**All Variables: Most Recent Macro**	LASSO	sg-LASSO
1	2.063	2.056	2.01	2.59	2.395
	(0.948)	(0.938)	(0.94)	(0.942)	(0.966)
2	1.174	1.143	1.145	1.313	1.213
	(0.922)	(0.897)	(0.916)	(0.888)	(0.83)
3	0.788	0.781	0.787	0.897	0.848
	(0.046)	(0.046)	(0.051)	(0.166)	(0.11)
4	0.51	0.506	0.499	0.595	0.663
	(0.001)	(0.002)	(0.002)	(0.005)	(0.007)
5	0.474	0.485	0.462	0.558	0.676
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
6	0.548	0.573	0.568	0.592	0.691
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
7	0.518	0.54	0.527	0.554	0.622
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
8	0.842	0.813	0.761	0.776	0.763
	(0.071)	(0.037)	(0.008)	(0.008)	(0.005)
9	0.992	0.89	0.892	0.897	1.009
	(0.482)	(0.209)	(0.263)	(0.27)	(0.522)
10	0.944	0.83	0.836	0.93	0.889
	(0.37)	(0.085)	(0.149)	(0.341)	(0.257)
11	1.33	1.032	1.104	1.461	1.299
	(0.896)	(0.555)	(0.692)	(0.915)	(0.903)
12	1.061	0.86	0.929	1.116	1.026
	(0.618)	(0.2)	(0.317)	(0.812)	(0.573)
13	0.998	0.726	0.8	0.902	0.923
	(0.495)	(0.059)	(0.107)	(0.27)	(0.305)
14	1.017	0.738	0.876	0.953	1.015
	(0.533)	(0.078)	(0.276)	(0.417)	(0.528)
15	1.013	0.74	0.895	0.917	1.032
	(0.525)	(0.085)	(0.302)	(0.361)	(0.559)
16	1.012	0.736	1.117	1.201	1.111
	(0.522)	(0.086)	(0.667)	(0.784)	(0.701)
17	0.955	0.73	0.917	1.005	1.053
	(0.42)	(0.086)	(0.357)	(0.514)	(0.596)
18	0.971	0.753	0.94	0.944	0.998
	(0.448)	(0.12)	(0.391)	(0.411)	(0.496)

Note: Column 1 shows the results for boosting with all variable lags directly. Column 2 shows the results for boosting when the first two extracted factors from macro leading variables together with monthly and annual taxes are used. Column 3 shows the results for boosting with all variables, but restricting high-frequency macro variables to be most recent. Column 4 shows the results for LASSO with all variable lags. Column 5 shows the results for sg-LASSO with all variable lags, where all high-frequency lags of the same variable are in one group. The values in parentheses are p values of one-sided Diebold-Mariano tests with null that the benchmark forecast is better. Methods in column 2 and 3 will be focused on when forecasting in COVID-19.

Table 6: Selected Lags: Boosting with All Variables

Horizon	Selected Variables
1	MTax_7, MTax_5, MTax_4, MTax_9, MTax_1, MTax_11, PersInc_real_2, HX1M.CNER_20
2	MTax_7, MTax_5, MTax_4, MTax_9, MTax_3, PersInc_real_2, HX1M.CNER_20, MTax_11
3	MTax_7, MTax_5, MTax_4, MTax_9, MTax_3, HX1M.CNER_20, MTax_11
4	MTax_7, MTax_5, MTax_4, MTax_9, HX1M.CNER_20, MTax_11, SEN_7, BAFDINA_8, NYOTOT_PC1_real_7
5	MTax_7, MTax_5, MTax_9, BAFDINA_8, PersSaveR_10, SEN_6, NYOTOT_PC1_real_7, PersSaveR_5, ELEC.NY_15
6	MTax_7, MTax_6, MTax_8, HX1M.CNER_20, BAFDINA_8, NYOTOT_PC1_real_7
7	MTax_7, MTax_8, RECESS2_2, HX1M.CNER_20, HX1M.CNER_18, NYOTOT_PC1_real_7, BAFDINA_8
8	MTax_8, MTax_9, ISM_NMEI_9, BAFDINA_8, NYICLAIMS_10, FedTxWithold_15, ISM_MNIOI_9
9	MTax_9, FedTxWithold_12, PersSaveR_18, ISM_MNIOI_9, NY_Bench_22, FedTxWithold_20
10	MTax_10, NYICLAIMS_17, NYICLAIMS_14, HX1MEDM.CNER_real_11, NYICLAIMS_15, NY_Bench_22, FedTxWithold_20
11	NYICLAIMS_14, MTax_11, FedTxWithold_12, BAFDINA_18, ISM_NMNOI_14, HX1MEDM.CNER_real_11, ISM_MNIOI_18, FedTxWithold_20, PersSaveR_18, MTax_23
12	NYICLAIMS_14, ISM_NMNOI_15, FedTxWithold_12, BAFDINA_18, NYICLAIMS_18, MTax_23, ISM_MNIOI_18, FedTxWithold_20
13	NYICLAIMS_14, ISM_MEI_15, NYICLAIMS_17, SEN_16, ISM_NMNOI_14, ISM_MII_14, ADL_3, MTax_23, HX1MEDM.CNER_real_13, FedTxWithold_22, BAFDINA_18, PersSaveR_18
14	NYICLAIMS_14, ISM_MEI_15, NYICLAIMS_17, SEN_16, ISM_NMNOI_14, ISM_MII_14, ADL_3, MTax_23, HX1MEDM.CNER_real_17, FedTxWithold_22, ISM_MNIOI_17, PersSaveR_18
15	SEN_16, NYICLAIMS_17, ISM_MEI_15, NYICLAIMS_15, ISM_NMNOI_17, SEN_15, ADL_3, MTax_23, HX1MEDM.CNER_real_17, PersSaveR_18, FedTxWithold_22, ISM_MNIOI_16
16	SEN_16, NYICLAIMS_17, ISM_NMNOI_17, ADL_3, MTax_23, ISM_MNIOI_16, PersSaveR_18, FedTxWithold_22
17	SEN_17, NYICLAIMS_17, ISM_NMNOI_17, ADL_3, MTax_23, ISM_MNIOI_17, HX1MEDM.CNER_real_17, FedTxWithold_22, PersSaveR_18
18	SEN_18, NYICLAIMS_18, MTax_23, ADL_3, HX1M.CNER_19, ISM_MNIOI_18, BAFDINA_18, PersSaveR_18

Note: "_i" following each monthly variable name represents the monthly lag index counting back from the last month of the annual target variable. "ADL" is the lagged annual target variable. The "_i" following "ADL" is the annual lag index counting back from the target. "MTax" is monthly tax variable. To save space, each variable lag is reported only once even if selected multiple times. First selected variable lags at each horizon are listed first.

Table 7: Selected Lags: Boosting with Factors

Horizon	Selected Variables
1	MTax_7, MTax_5, MTax_4, MTax_9, MTax_1, MTax_11
2	MTax_7, MTax_5, MTax_4, MTax_9, MTax_3, MTax_11, F1_3, F2_3
3	MTax_7, MTax_5, MTax_4, MTax_9, MTax_3, MTax_11, F1_3
4	MTax_7, MTax_5, MTax_4, MTax_9, MTax_11, F2_4, MTax_13
5	MTax_7, MTax_5, MTax_9, F1_5, ADL_2, F2_5
6	MTax_7, MTax_6, MTax_8, F2_6
7	MTax_7, MTax_8, F2_7
8	MTax_8, MTax_9, F1_8
9	MTax_9, F1_9, MTax_14
10	MTax_10, MTax_23, F1_10
11	MTax_11, F1_11, MTax_23, MTax_22
12	F1_14, MTax_23, MTax_22, MTax_13
13	F1_14, MTax_23, ADL_3, MTax_13, F1_13, MTax_22
14	F1_14, MTax_23, ADL_3
15	F1_15, MTax_23, ADL_3, F2_17
16	F1_16, MTax_23, ADL_3, F2_17
17	F2_17, MTax_23, ADL_3
18	F2_18, MTax_23, ADL_3

Note: F1 is the first factor. F2 is the second factor. "_i" following each monthly variable name represents the monthly lag index counting back from the last month of the annual target variable. "ADL" is the lagged annual target variable. The "_i" following "ADL" is the annual lag index counting back from the target. "MTax" is monthly tax variable. To save space, each variable lag is reported only once even if selected multiple times. First selected variable lags at each horizon are listed first.

Table 8: Nowcasts of Remaining Tax Receipts to be Collected: Fiscal Year 2021

Nowcast Date	Nowcast for Tax Revenues in Remaining Months ($Billion)	Actual Tax Revenues in Remaining Months ($Billion)
5/2020	64.30	76.89
6/2020	66.57	73.82
7/2020	60.47	66.81
8/2020	56.86	54.52
9/2020	49.12	50.26
10/2020	40.77	41.68
11/2020	35.65	37.47
12/2020	31.43	33.25
1/2021	23.58	25.03
2/2021	13.20	13.75
3/2021	7.26	6.85

Table 9: Forecasts by Different State Agencies using Data til January 2020: FY 2020 and FY 2021

	Forecaster	FY 2019 Actual Nominal (million)	FY 2020 Nominal (million)	Nominal Growth(%)	FY 2021 Nominal (million) January data	Nominal Growth (%) Jan.
1	NY DOB	75578	82390.00	9.01	87932.00	6.73
2	Senate Financial Committee	75578	83037.00	9.87	88432.00	6.50
3	Senate Minority	75578	83062.00	9.90	87973.00	5.91
4	Ways and Means	75578	83146.00	10.01	88750.00	6.74
5	Ways and Means Minority	75526	82671.00	9.39	88310.00	6.82

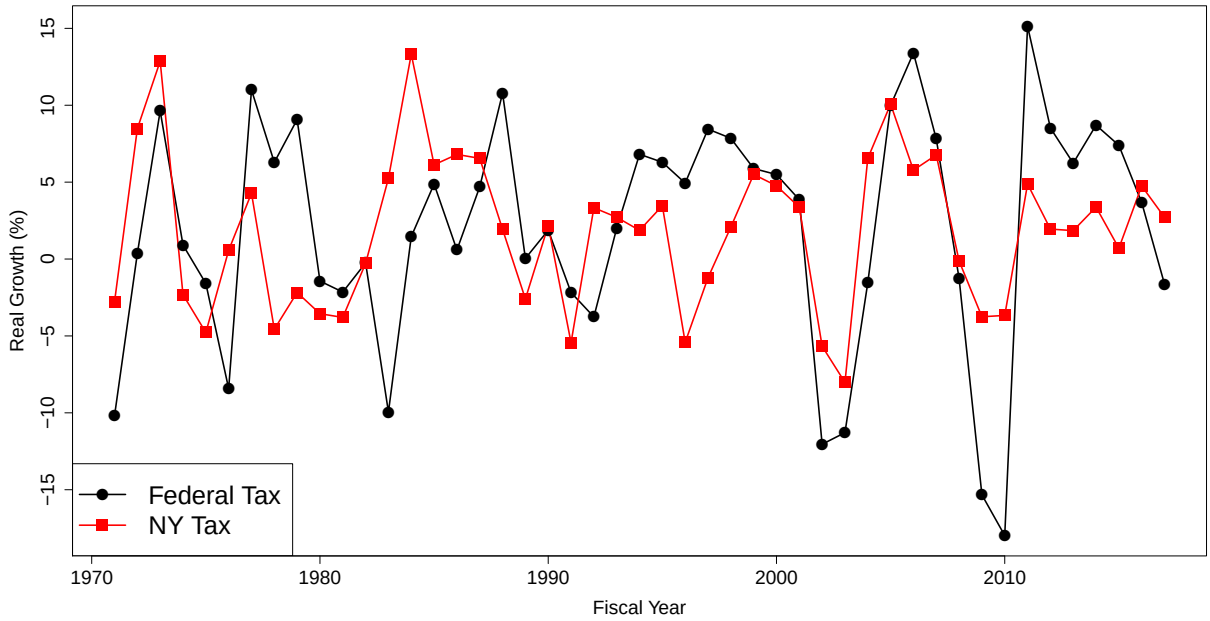
Note: All forecasts and their growth values in this table are nominal. Forecasts are made with data until January 2020, when executive budget proposals are made.

Table 10: Variables Selected by Boosting to Forecast/Nowcast FY 2021 under COVID-19: Horizons 12 to 8

Data Used	Selected Variables
March 2020	FedTxWithold_12, CCLAIM_12, ADL_2, PersConsExpen_real_12, MTax_23, MTax_13
April 2020	CCLAIM_11, MTax_11, trucktraf_11, FedTxWithold_11, MTax_22, HX1MEDM.CNER_real_11, BAFDINA_11
May 2020	MTax_10, CCLAIM_10, BAFDINA_10, FedTxWithold_10, MTax_23
June 2020	MTax_9, ISM_MNIOI_9, CCLAIM_9, FedTxWithold_9
July 2020	MTax_8, MTax_9, BAFDINA_8, ISM_MII_8

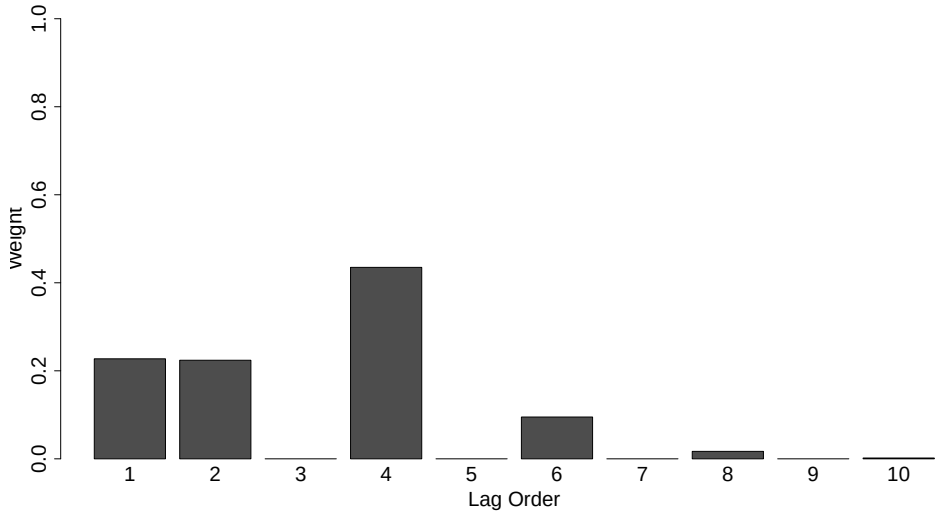
Note: Selected variables here are from the model which restricts macro variables to be the most recent.

Figure 1: State Tax in New York and Federal Tax

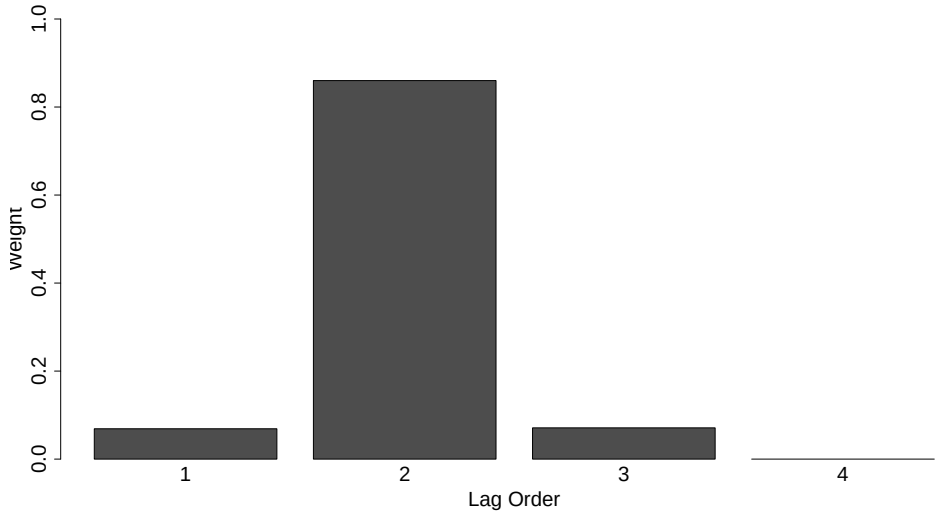


Note: Federal tax data is from U.S. Bureau of Economic Analysis. State tax data for New York is from Ways and Means Committee of New York Assembly. Its annual value is only slightly different from tax data of New York Department of Taxation and Finance, but goes earlier than 1986. We use data from Department of Taxation and Finance in forecasts even though it is shorter since it is most original source and is updated more timely every month. The values are in terms of year-over-year real growths. For the purpose of comparison, the annual values of federal tax receipts are in terms of New-York fiscal year: Federal tax receipts from the second quarter of current calendar year to the first quarter of next calendar year are added up to get one annual value.

Figure 2: Examples of Implied Lag Distributions for MTax from Boosting



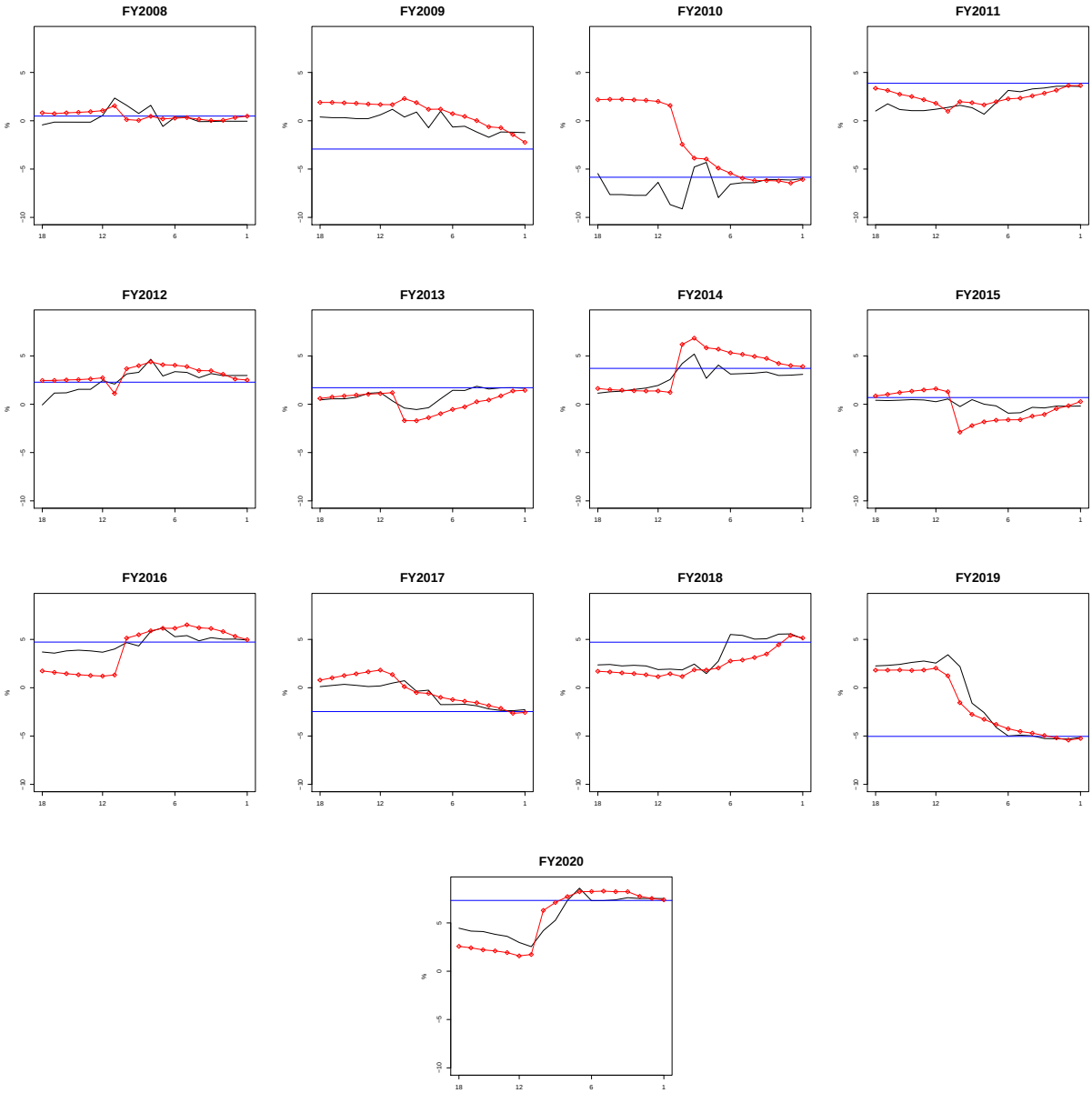
(a) 4 Months Ahead



(b) 6 Months Ahead

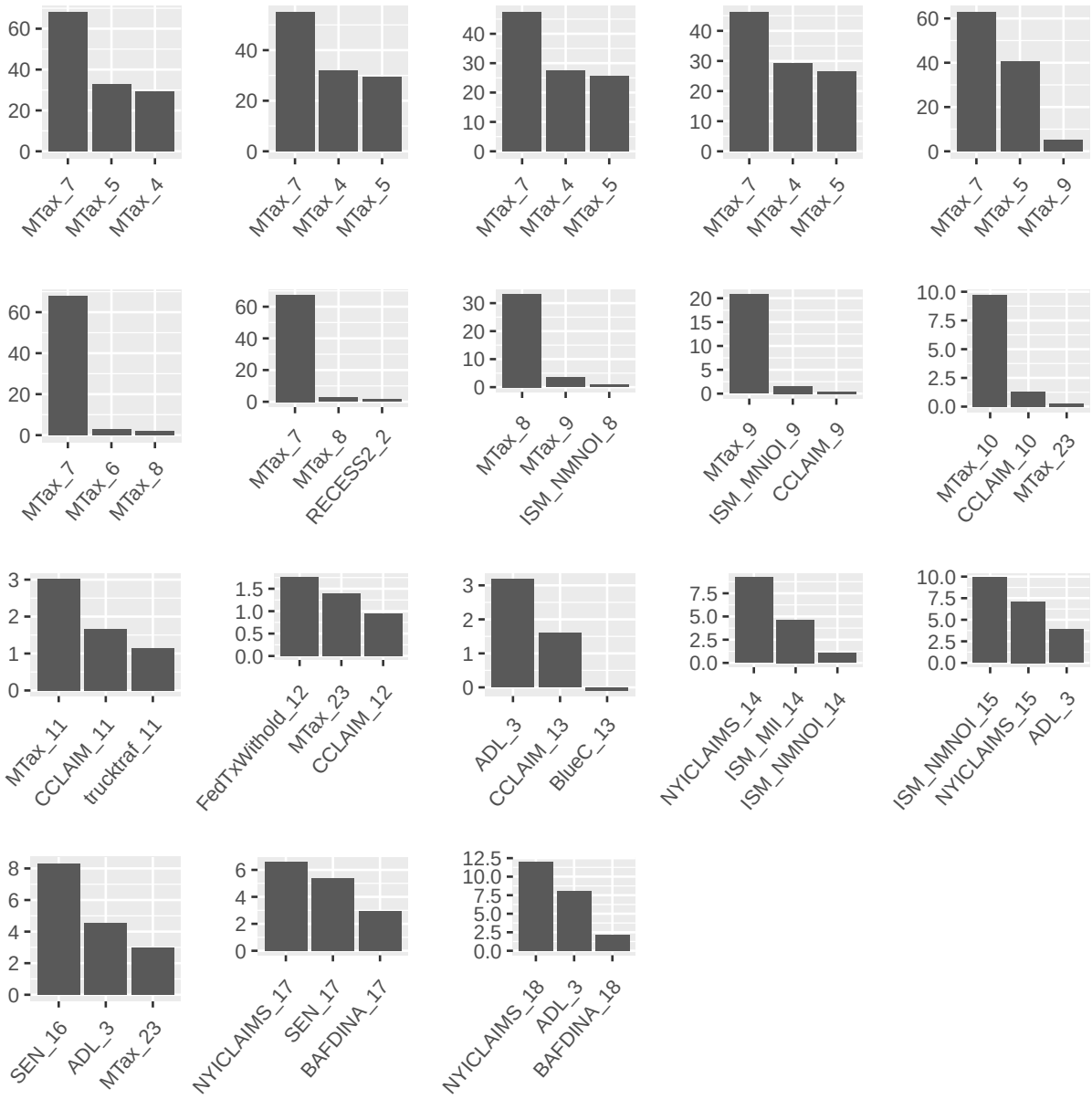
Note: This figure shows implied lag distribution for MTax (monthly tax variable) from boosting with factors at horizon 4 and horizon 6. The first lag at horizon 4 is November, which corresponds to "MTax_4" in Table 6 and 7. Similarly, the first lag at horizon 6 is September, which corresponds to "MTax_6". The distributions differ clearly from uniform distribution or any regular distributions with parameters in restricted MIDAS models.

Figure 3: Boosting Forecasts with Factors



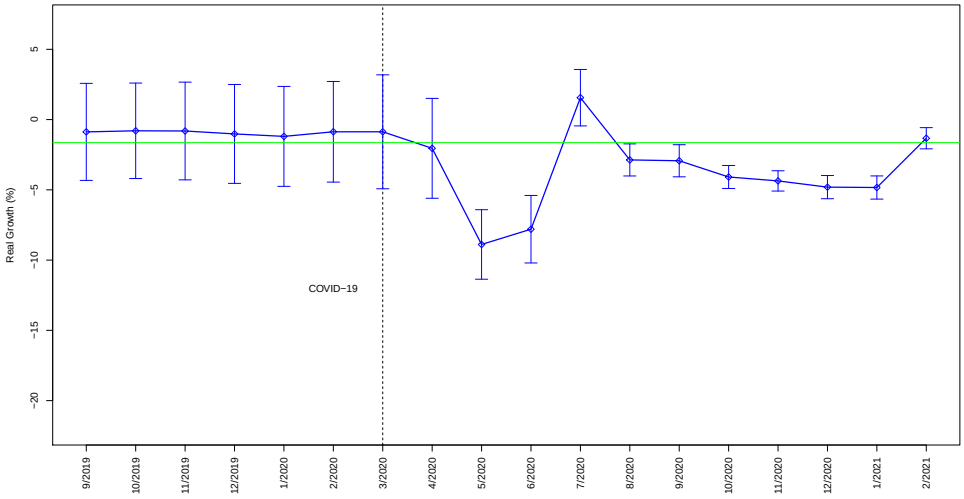
Note: Each graph represents boosting forecasts with factors for a specific target fiscal year. The horizon gets smaller from left to right, ranging from 18 to 1. Vertical axis is forecast/nowcast in terms of percentage change. Black lines are boosting forecasts and nowcasts. Red lines with squares are benchmark forecasts and nowcasts. Blue horizontal lines are actual values.

Figure 4: Relative Importance of Selected Variables

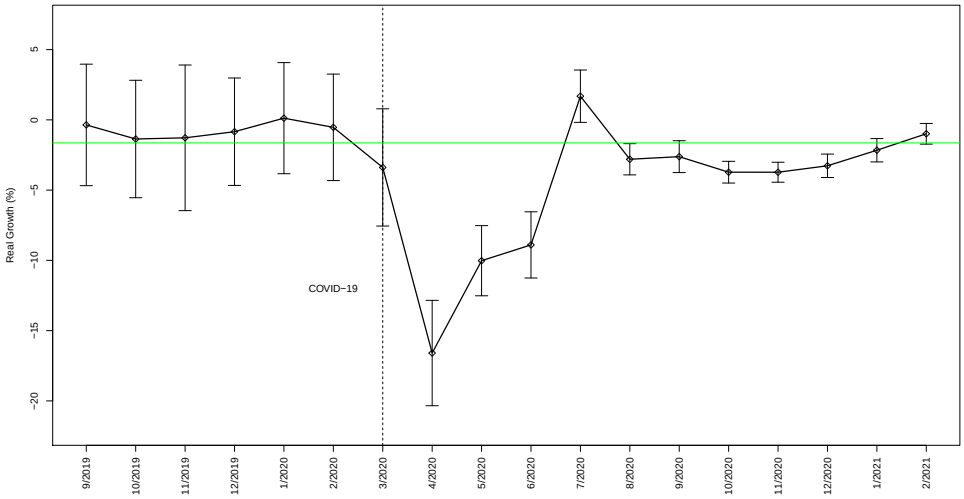


Note: From left to right, the graphs in the four rows are for horizons 1 to 5, 6 to 10, 11 to 15, and 16 to 18, respectively.

Figure 5: Forecasts for Fiscal Year 2021



(a) Boosting with factors



(b) Boosting with most recent macro variables

Note: The forecast started from 10/2019 with data of 9/2019, which was the 18-month-ahead forecast for Fiscal Year 2021 (April 2020-March 2021). Growth rates here are real. Data used are as of March, 2021, which is for February 2021. The green horizontal line is the actual value. The first figure represents boosting forecasts and nowcasts using factors. The second figure represents boosting forecasts and nowcasts with most recent macro variables. Both figures are with 75% confidence intervals. The confidence interval at each horizon is generated according to the root mean squared error of historical forecasts/nowcasts at that horizon. We leave a better estimation of forecast confidence interval of mixed-frequency-data boosting model for future study.