

Revenue Forecasting Methodologies

Single Exponential Smoothing

$$F_t = F_{t-1} + \alpha(y_{t-1} - F_{t-1})$$

- F_t is the forecast revenue for period t
- t is the time period
- α is the smoothing constant, $0 < \alpha \leq 1$
 - We set this coefficient by using solver to find the value that results in the lowest Mean Squared Error
- y_{t-1} is the previous period's actual revenue
- F_{t-1} is the previous period's forecast revenue
 - We set the value for the first forecast as equal to the previous month's actual

This method is generally best for short-term forecasts

- Because after actual values are no longer available, the last available forecast is used as the value for all future forecasts

12-Month Moving Average

To make a 12-month moving average forecast, we take the average of the last 12 actual or forecast values

$$F_t = \text{Average}(y_{t-12}: y_{t-1})$$

- F_t is the forecast revenue for period t
- t is the time period
- y_{t-12} is the actual revenue for the time period 12 periods before t
- y_{t-1} is the actual revenue for the time period immediately before t

$$F_t = \text{Average}(y_{t-12}: F_{t-1})$$

- F_{t-1} is the forecast revenue for the time period immediately before t

$$F_t = \text{Average}(F_{t-12}: F_{t-1})$$

- F_{t-12} is the forecast revenue for the time period 12 periods before t

This method is best for relatively short-term forecasts, as the forecast values flatten after we reach 24 months beyond the last actual value

Simple Trending

This method treats the data as linear, and uses linear regression to determine the average slope of the data over time

$$F_t = m * t + b$$

- F_t is the forecast for period t
- t is the time period
- m is the slope
 - To find this, use the SLOPE function in excel, with actuals as y-values and t as x-values
- b is the intercept
 - To find this, use the INTERCEPT function in excel, with actuals as y-values and t as x-values

This method allows us to make forecasts far into the future, but is limited because it does not account for seasonal fluctuation

Time Series Regression

This is the most sophisticated methodology used in our forecasting

We use a centered moving average to divide the data into a de-seasonalized component and a seasonal factor for each month

- Then we apply linear regression to the de-seasonalized component to identify the trend
- Then we re-apply the monthly seasonal component to the de-seasonalized trend

This method allows us to forecast more long-term, and takes seasonality into account

Time Series Regression-Formulas

12-month moving average: $MA_t = average(y_{t-5}:y_{t+6})$

- MA_t is the moving average for period t
- t is the time period
- y_{t-5} is the actual value for the time period 5 periods before t
- y_{t+6} is the actual value for the time period 6 periods after t

Centered moving average: $CMA_t = average(MA_t:MA_{t+1})$

- CMA_t is the Centered Moving Average for period t

Extracting seasonality and irregularity: $S_t, I_t = y_t/CMA_t$

- S_t, I_t is the seasonal and irregularity component for period t

Extracting just seasonality: $S_t = average(all S_{tmonth}, I_{tmonth})$

- S_t is the seasonal component for period t
- S_{tmonth}, I_{tmonth} is the seasonal irregular component for all time periods that occur during the same month of the year. For example, to find the seasonal component for December, average the S_t, I_t for all Decembers. Then, use this averaged value as the seasonal value for each December

Time Series Regression-Formulas

De-seasonalizing the data: $DeseasonalizedY_t = Y_t/S_t$

- $DeseasonalizedY_t$ is the deseasonalized value for period t

Finding the trend component: $T_t = m * t + b$

- T_t is the trend component for period t
- m is the slope
 - To find this, use the SLOPE function in excel, with $DeseasonalizedY_t$ as y-values, and t as x-values
- b is the intercept
 - To find this, use the INTERCEPT function in excel, with $DeseasonalizedY_t$ as y-values, and t as x-values

Making the forecast: $F_t = T_t * S_t$

- F_t is the forecast for period t

Compound Annual Growth Rate

Compound Annual Growth Rate (CAGR) allows us to find the annual rate of growth over a given time period, and use this to create annual forecasts

$$CAGR = \left(\frac{EV}{BV}\right)^{\frac{1}{t}} - 1$$

- EV is the ending value for your time period (for our forecasts, the last available actual)
- BV is the beginning value for your time period
- t is the number of years covered by your time period

$$F_t = CAGR * F_{t-1}$$

- F_t is the forecast for period t
- F_{t-1} is the forecast (or actual) for the previous time period

Modified Compound Annual Growth Rate

Several modified CAGR methods were created during the time of COVID-19

This method aims to forecast recovery from COVID-19, and the associated social distancing policies

- Phased social distancing is necessary during COVID-19 outbreaks to contain the spread of the virus, but these policies mean substantial reductions in several revenue sources and potential long-term increases in unemployment

3-Year CAGR

- Assumes permanent downward shift in demand
- Compound annual growth rate is applied to 2021 budget value

3-Year CAGR 2019 Base

- Assumes full recovery in 2022
- 2022 estimate is greater than the 2019 estimate by the compound annual growth rate, as though 2020 and 2021 did not happen

3-Year CAGR 2019 Base Ramp Up

- Assumes gradual recovery until 2025
- Combination of 3-Year CAGR and 3-Year CAGR 2019 Base

3-Year CAGR 2019 Base Ramp Up

$$F_{2022} = (CAGR * F_{2021}) + (((CAGR * Y_{2019}) - (CAGR * F_{2021})) * .5)$$

- F_{2022} is the forecast for 2022
- $CAGR$ is the compound annual growth rate, calculated as shown on the compound annual growth rate slide above
- F_{2021} is the 2021 estimate being used
- Y_{2019} is the 2019 actual

$$F_{2023} = (CAGR * 2 * F_{2021}) + (((CAGR * 2 * Y_{2019}) - (CAGR * 2 * F_{2021})) * .66)$$

$$F_{2024} = (CAGR * 3 * F_{2021}) + (((CAGR * 3 * Y_{2019}) - (CAGR * 3 * F_{2021})) * .85)$$

$$F_{2025} = (CAGR * 4 * F_{2021}) + ((CAGR * 4 * Y_{2019}) - (CAGR * 4 * F_{2021}))$$

R Regression

More complex linear regressions are completed using the statistical software R

- This provides more processing power than Excel

This method is similar to the “simple trending” method described previously, but it takes into account seasonality and economic variation, in addition to the trend over time

$$F_t = a + b * t + c * e + m_i$$

- F_t is the forecast for period t
- a is the intercept
- b is the coefficient for time, and t is the time period
- c is the coefficient for the economic indicator, and e is the value of the economic indicator
 - This economic indicator will be the unemployment rate, GDP, etc
 - Up to three economic indicators could be included
- m_i is the coefficient for month
 - There are 12 m_i values, one for each month. Select the one corresponding to the appropriate month.
- The best values for a, b, c, m_i calculated by an automated computer program since the math is too complex to conduct in excel

Model fit can be evaluated using an R-squared value, which represents the percentage of variation in revenue that can be explained by the equation

Gross Domestic Product (GDP)

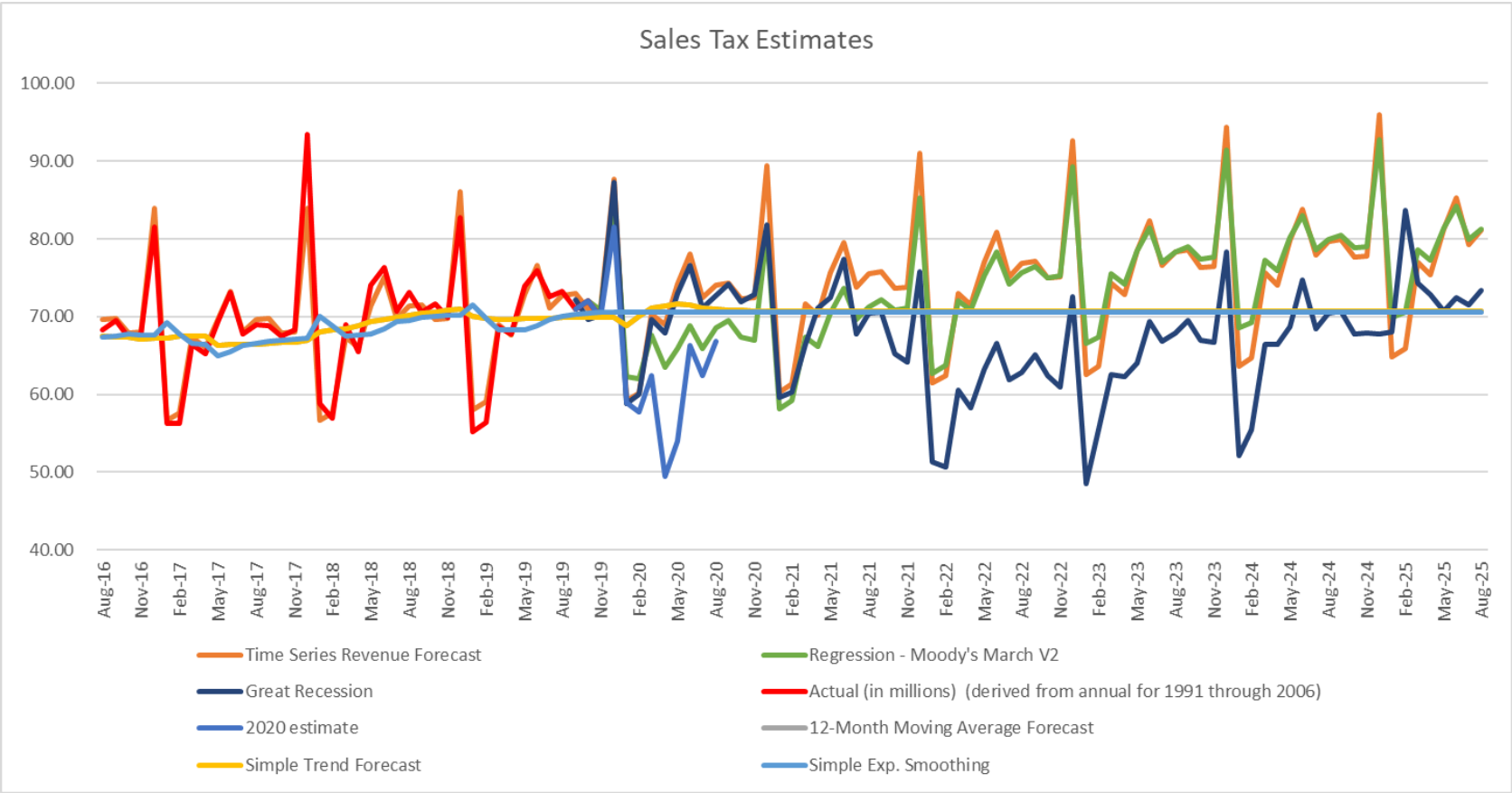
The GDP methodology makes the assumption that many revenue streams will grow or decline at a similar rate to the US economy in the long run. Under this methodology, we start with the 2021 budget value and grow a given revenue at the same rate as forecasted GDP growth.

Firstly, numerous US GDP growth rate forecasts from reputable sources were collected. Both the most conservative forecast, and the average of all forecasts, were used. Real GDP is adjusted for inflation to represent actual output changes independent of price changes. This is appropriate for quantity-based taxes (e.g. Gasoline Tax of 6 cents per gallon). However, percentage-based tax revenue (e.g. Sales Tax of 1.75%) will grow with price increases, so nominal GDP is more appropriate.

$$F_t = F_{t-1} * (1 + GDP_t)$$

- F_t is the forecast for period t
- F_{t-1} is the forecast (or actual) for the previous time period
- GDP_t is the predicted GDP growth rate in period t, either real or nominal depending on the revenue

Example: Sales Tax



Interpolation Methodology

Due to the COVID-19 pandemic, many revenue streams had abnormally low values in 2020. This posed a problem for future long term forecasting efforts. Since another pandemic is not anticipated, statistical extrapolations should not be influenced by this one time shock. To solve this issue, 2020 data was interpolated for some revenue streams that had significant losses.

The interpolation technique used was linear with seasonality. This involved first de-seasonalizing the actuals using the seasonality calculated for the time series forecast. Then, the slope between the de-seasonalized months before and after the pandemic was calculated to create a linear fit for the “missing” data. This de-seasonalized interpolated data is then re-seasonalized and used in standard statistical extrapolation methods.

$$\text{Slope} = (Y_{2.20}/S_{2.20} - Y_{2.20+i}/S_{2.20+i}) / (Y_{2.20}/S_{2.20}) / i$$

$$\text{Interpolated}_t = S_t * \text{Interpolated}_{t-1} * (1 + \text{Slope})$$

- $Y_{2.20}$ is the actual from February 2020, before the pandemic
- $Y_{2.20+i}$ is the actual from i months after February 2020, after the pandemic (varies by revenue)
- S_t is the time series seasonality for period t

Naïve Forecasting

Naïve: $Y_{t+1} = Y_{t+n} = Y_t$

- Take the most recent value in the time series and propagate it forward as many prediction steps (n) as desired

Seasonal Naïve: $Y_{t+1} = Y_{t-h+1}$ for a time series with seasonality h

- Take the most recent value of the same seasonality
- For monthly data, take the value 12 months prior. E.g December 2021 will be predicted as the value of December 2020

ETS Models

Apply exponential smoothing on Error, Trend, and Seasonal (ETS) components (with cycle length L)

Initialization: $s_0 = x_0$

Data Smoothing: $s_t = \alpha(x_t - c_{t-L}) + (1 - \alpha)(s_{t-1} + b_{t-1})$

Trend Smoothing: $b_t = \beta (s_t - s_{t-1}) + (1 - \beta)b_{t-1}$

Seasonal Smoothing: $c_t = \gamma(x_t - s_{t-1} - b_{t-1}) + (1 - \gamma)c_{t-L}$

Forecast (m steps): $F_{t+m} = s_t + mb_t + c_{t-L+1+(m-1)}$

- For an additive model
- If you wanted a multiplicative model, use the same smoothing just multiply in F_{t+m} instead of adding
- Additive: Seasonality has absolute change
- Multiplicative: Seasonality has factor (%) change
 - E.g: December is \$10,000 higher than November → Additive
 - E.g: December is 10% higher than November → Multiplicative

Parameters fit by minimizing AIC

AutoRegressive Integrated Moving Average (ARIMA) Part 1

Given in the form ARIMA(p,d,q), where p is the order of the autoregression, d is the number of differencing steps, and q is the order of the moving average

Autoregression AR(p): $Y_{t+1} = \alpha_t Y_t + \alpha_{t-1} Y_{t-1} + \dots + \alpha_{t-p} Y_{t-p}$

- Regressing variable on its own past values

Moving Average MA(q): $Y_{t+1} = \epsilon_t + \theta_{t-1} \epsilon_{t-1} + \dots + \theta_{t-q} \epsilon_{t-q}$

- ϵ are error terms
- Regressing variable on its own past errors

Integration I(d): $Y'_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) - \dots - (Y_{t-d-1} - Y_{t-d})$

- Makes mean constant

Reference 1: <https://otexts.com/fpp3/arma.html>

Reference 2: https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average

ARIMA Part 2

Differenced (Integrated) data is then used to fit an ARMA(p,q) model

$$\text{ARMA}(p,q): Y_t = \alpha_{t-1}Y_{t-1} + \alpha_{t-2}Y_{t-2} + \dots + \alpha_{t-p}Y_{t-p} + \epsilon_t + \theta_{t-1}\epsilon_{t-1} + \dots + \theta_{t-q}\epsilon_{t-q}$$

- Adding AR(p) and MA(q) models

To choose p, d, q: Fit multiple models and choose based on lowest information criterion, usually AICc

$$\text{AICc: } AIC - \frac{2k(k+1)}{n-k-1} = -2 * \text{Log}(\text{Likelihood}) + 2k - \frac{2k(k+1)}{n-k-1}$$

- k is the number of parameters
- n is the sample size
- Likelihood is the probability of the data arising from the proposed model
- AICc is a common estimator for prediction error, essentially choosing model with the lowest prediction error

Reference 1: https://en.wikipedia.org/wiki/Autoregressive-moving-average_model

Reference 2: https://en.wikipedia.org/wiki/Akaike_information_criterion

Seasonal and Trend Decomposition with Loess (STL)

$Y_v = T_v + S_v + R_v$ where T:=Trend, S:=Seasonality, R:=Remainder for observations 1...v

Loess: Create a smooth function by taking all possible x_i within the region of q points, then fit a d degree polynomial to that region

Core Algorithm:

- De-Trend series: $Y_v - T_v$
- Smooth cycle-subseries using loess. Number of cycles is a parameter. Result is C_v
- Filtering cycle-subseries: apply two moving averages of length n_p , which is a parameter, then a moving average of length 3, then apply loess smoothing again. Result is L_v
- De-Trend smoothed cycle sub-series: $S_v = C_v - L_v$
- De-Seasonalize series: $Y_v - S_v$
- Smooth De-Seasonalized series with loess.

This process is repeated with different “robustness” weights, which determine how influential outliers are. The different values are iteratively updated for a predetermined number of iterations

Reference 1: <https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/stl-a-seasonal-trend-decomposition-procedure-based-on-loess.pdf>

Reference 2: <http://www.gardner.fyi/blog/STL-Part-II/>

Reference 3: <https://otexts.com/fpp3/stl.html>