# Dr. Z. Radovilsky
# BAN 673 Time Series Analytics: Summary of Process Steps

| *Step* | *Name* | *Main Decisions* | *Comments* |
|---|---|---|---|
| 1 | Define Goal | • **Descriptive and/or predictive goal**<br>• **Forecasting horizon** (long-term vs. short-term).<br>• **Forecasting usage** (stakeholders and periodic adjustments).<br>• **Forecast expertise** (in-house or outsourcing; how many and how often to forecast; software for time series forecasting). | • Forecasting is a combination of science and art.<br>• Popular software in business time series forecasting: **Excel** (for simple time series forecasting), **R** (for simple and advanced forecasting methods), **Python** (for simple and advanced forecasting methods), **end-user general statistical software** (e.g., JMP, Stata, SPSS, Minitab), **professional forecasting software** in enterprise resource planning (ERP) systems (e.g., Oracle, SAP, Microsoft, etc.) |
| 2 | Get Data | • **Data quality**<br>• **Temporal frequency**<br>• **Series granularity**<br>• **Domain expertise** | • In business, typical frequencies of historical data are monthly, quarterly, daily, weekly, and yearly.<br>• **Minimum data requirements in time series analytics**: at least 36 monthly periods (3 years) for monthly data; at least 16 quarterly periods (4 years) for quarterly data; at least 104 weekly periods (2 years) of weekly data.<br>• Sometimes, it is very useful to aggregate the data and then forecast it to achieve better accuracy results. |
| 3 | Explore & Visualize Series | • **Summary statistics of historical time series** (mean, median, standard deviation, etc.).<br>• **Time series patterns (components):** Level (Stationary), Trend, Seasonality, and Noise (Randomness).<br>• **Additive vs. multiplicative components** (specifically, for seasonality).<br>• **Autocorrelation** in various lags.<br>• **Visualize series** with a data plot to check for data series patterns (components). | • Not all time series have trend and/or seasonality, but **all contain level and noise (randomness).**<br>• **Significant autocorrelation coefficients**: lag 1 --> trend; lag 4 -> seasonality in quarterly data; lag 12 -> seasonality in monthly data; autocorrelation coefficients decline rapidly to being insignificant (0) after 2$^{nd}$ or 3$^{rd}$ lags -> level in time series data.<br>• To remove trend and/or seasonality from time series apply differencing method.<br>• For **data predictability** evaluation: ***Approach 1*** -> hypothesis testing of the autoregression coefficient in *AR(1)* model, *Ho*: β = 1. If *Ho* is rejected (p-value is <= 0.05), this time series is predictable, otherwise – it is random walk; **Approach 2** -> consider autocorrelation coefficients for first differenced data. If the balk |

| | | | |
|---|---|---|---|
| | | • **Evaluate data predictability** (time series is 'random walk' -> forecasting methods other than naïve forecast may not be useful; time series is not 'random walk' -> various forecasting methods may be used) | portion of them is random, specifically in lags 1-4 and seasonality lags, this is random walk; otherwise – time series is predictable. |
| 4 | Pre-Process Data | • **Detect and clean potential issues in time series data**: outliers, missing values, unequal spacing, irrelevant periods, etc. | • Apply domain expertise to remove or keep outliers in time series historical data.<br>• Utilize a standard imputation approach to replace missing values (similar to Data Mining/Machine Learning imputation).<br>• If unequal spacing is expected in a data set (e.g., in daily stock market data), consider using the data frequency of 1.<br>• Remove records with irrelevant periods. |
| 5 | Partition Time Series | • **Training partition (period)** is used for model development.<br>• **Validation (test) partition (period)** is used for model testing. | • **Nonrandom selection** of partition period.<br>• **Early period** in historical data -> **training partition** with 70-80% of historical data.<br>• **Later period** in historical data -> **validation partition** with 20-30% of historical data.<br>• The smaller the time series, the less data will go into validation period.<br>• **Recombine partitions to the entire data set for the final forecasting model development** and forecasting into the future.<br>• If data patterns in the validation period are significantly different from those in the training period, the data partitioning may not be useful -> apply the entire data set for model development and forecasting. |
| 6 | Apply Forecasting Method(s) | • **Model-driven methods** (e.g., linear/non-linear regression, autoregressive models, ARIMA)<br>• **Data-driven methods** (e.g., naïve and seasonal naïve trailing and weighted moving average, simple and advanced exponential smoothing).<br>• **Two-level forecasting** (e.g., Level 1 -> regression model with trend and seasonality; Level 2 -> simple | • Always **select several alternative models** (at least 2-3) for development in training period and performance comparison in the validation period.<br>• Examples of alternative models for time series with **level component** only: naïve, trailing MA, simple exponential smoothing (SES), ARIMA, and auto ARIMA.<br>• Examples of alternative models for times series with **level and trend components** only: naïve, regression with linear/nonlinear trend, Holt's model, AR, non-seasonal ARIMA, auto ARIMA, two-level, and ensembles. |

| | | | |
|---|---|---|---|
| | | exponential smoothing for regression errors; Two-level forecast -> combination of Level 1 and 2 forecasts).<br>• **Ensemble methods** (resulting forecast is the average of several forecasting methods in respective periods; can be weighted average with specific forecast's weight proportional method performance).<br>• **Additional methods** | • Examples of alternative models for times series with **level and seasonality components** only: seasonal naïve, regression with seasonality, Holt-Winter's, AR, seasonal ARIMA, auto ARIMA, two-level, and ensembles.<br>• Examples of alternative methods for times series with **level, trend, and seasonality components**: naïve, seasonal naïve, regression with linear/nonlinear trend and seasonality, Holt-Winter's, AR, trend and seasonal ARIMA, auto ARIMA, two-level, and ensembles.<br>• Develop each model using training partition (period) and test it in validation partition (period).<br>• For the final forecast into the future, use the entire data set to develop the model or several alternative models.<br>• For additional time series methods consider Hyndman, R.J. & Athanasopoulos, G. Forecasting: Principles and Practice, 2nd Ed., https://otexts.com/fpp2/. |
| 7 | Evaluate & Compare Performance | • **Common accuracy performance measures** (ME, RMSE, MAE, MPE, MAPE, MASE, ACF1, Theil's U).<br>• Used to **select the best model for forecasting**.<br>• **Prediction (confidence) intervals** to identify forecast range in the forecasting period. | • The lower the value of each accuracy measure, the more accurate the associated model.<br>• Usually compare performance measures in validation period and/or entire data set.<br>• **MAPE and RMSE should be considered as the preferred performance measures** for model selection, with MAPE being of higher importance than RMSE in business time series forecasting.<br>• Typical prediction intervals are based on 95% confidence level probability (in R, default to 80% and 95% confidence level). |
| 8 | Implement Forecast/ System | • Incorporate forecast into the existing enterprise IT system.<br>• **Create and maintain IT support system**.<br>• **Periodic review and reevaluation of forecast**. | • As new historical data becomes available, consider reevaluating the forecasting model.<br>• The reevaluation interval should be approximately every 3-6 months (from one quarter to semi-annual), and should include steps 2-8 of the time series analytics process. |