Time Series:
Improving Your Models
Today’s Speakers

Jess Lin  
Data Scientist, DataRobot

Taylor Larkin  
Data Scientist, DataRobot
Time Series Overview
Time Series Modeling
AI-Driven Forecasting for a Real World, Not an Ideal World

Unmatched Accuracy, Out of the Box
Human-friendly visual insights, and unmatched accuracy, out of the box with automated advanced algorithms, time-aware feature engineering and backtesting

Scale & Automation to Forecast Millions of Items
AI that learns from and forecast millions of items at once, and can scale and automate forecasting millions of items

End-to-End Support for Real World Business Constraints
End-to-End support for real world business constraints, highly-configurable specify horizons, constraints, known in advanced or custom features

Deep Trust, Insights, and Integration to Put Models in Production
Complete automation from data to value Deep Trust, Insights, and Integration to Put Models in Production (MLops, Data Management, API, Dedicated Predictions)
Now What?
1. **Problem Framing**
   - Validate time series approach
   - Match modeling approach to business goals

2. **Time Series Data Prep**
   - Include important predictors
   - Confirm data quality over time
   - Assess seasonality patterns

3. **Adjusting Project Settings**
   - Match project settings with business goals
   - Incorporate domain knowledge
   - Determine impact of behavior across other groups

4. **Advanced Workflows**
   - Isolate distinct groupings of behavior
   - Handle periods of no behavior
   - Establish a hierarchy to the data
   - Investigate other advanced techniques

5. **Modeling Evaluation and Iteration**
   - Compare against a simple baseline model
   - Plot model error with respect to time
   - Analyze if certain series behave differently than others
   - Simplify your models (when possible)
   - Build multiple models for multiple periods of time
Problem Framing
Prioritizing the Right Time Series Use-Cases

- Personal Impact
  - Low Impact
  - High Impact

- $ Impact
  - Low Value
  - High Value

**IS THERE AN ESTIMATED VALUE IN REDUCING THE ERROR?**
- No
- Yes

**IS THERE A QUANTIFIABLE ERROR IN YOUR FORECAST?**
- No
- Yes

**IS THERE SCALE?**
- No
- Maybe?
- Yes, Definitely

**High Impact**
- High Value
Is it a **time series** problem?

- **Prediction Type**
  - Direct, forecast, nowcast
- **Row Spacing**
  - Irrelevant
  - Irregular, semi-regular, regular
- **Target Transforms**
  - Differencing
  - Log

- **Inputs**
  - Multivariate, univariate, interactions
- **Partitioning**
  - Cross validation, group
  - OTV
- **Series**
  - Single or multi series

➡️ Just because you have a timestamp doesn’t mean it is a time series problem!
What level of aggregation should I model?

1. What is the business question?
2. Should we predict the mean or the total?
3. Explore adding features such as mean, total, min, max, slope while aggregating

MORE Granularity
- Less stable target
- More data to train models
- Higher likelihood time series models will capture interesting dynamics
- Potentially too many zeros to model data well

LESS Granularity
- More stable target
- Less data to train models
- Dynamics are often damped
- Tend to be worse ML problems
Attributes of Single vs. Multiseries Modeling

Single series models
- **Pros**
  - All aspects of the model are estimated for the exact series of interest
- **Cons**
  - Feature selection is limited to just what the individual series has encountered
  - Scale - if you have >100 products, the number of models is prohibitive

Multi series models
- **Pros**
  - The model often learns features that single series models miss
  - Scale - in DataRobot we can model 1M series per project
  - We can also add cross-series features under advanced settings
- **Cons**
  - The best result is selected for the aggregate, not for any one series
Cold Start, Warm Start, Ongoing

- **Cold Start**
  - New series model (no history)
- **Warm Start**
  - Model(s) for series with less than the FDW of data
- **Ongoing**
  - Model(s) for SKUs with more data than the FDW

- Prepare a dataset of all metadata at launch for all previous new series.
  - Examples include clothing size, category, color, materials
- Consider including the prediction for the category from an ongoing model to establish a baseline for category activity
- Build multiple models for each day until you reach the FDW
- Use the Cold start Dataset, with the addition of lags like day 1 sales, day 2 sales
- Run time series models as expected
Data Prep
Add (and Check) New Data Sources
Data Quality

Leading or trailing zeros

Zeros padded to fill in missing values of a SKU. As a rule, never fill in missing values with zeros unless you know it’s actually zero.

Irregular spacing

Check if the frequency of data matches across series. If it doesn’t, consider imputing or handling separately.

Duplicate Dates

Series C has multiple observations for July 1st and 4th. Drop or aggregate these duplicates prior to uploading.
What about planned events? - KIAs

Known in Advance (KIA) variables are things we do know in the future, like a planned store closure for renovation.
Add Cross Series Features

Consider historical observations across series to better capture target signal

- Extract rolling statistics on the total target across all series in a regression project.
  - For example, if we are forecasting sales across all stores, cross series feature include:

<table>
<thead>
<tr>
<th>Feature</th>
<th>Cross-series statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales (total)</td>
<td>28 day mean</td>
<td>Total sales across all stores within a 28 day window</td>
</tr>
<tr>
<td>Sales (total)</td>
<td>1st lag</td>
<td>Latest value of total sales across all stores</td>
</tr>
<tr>
<td>Sales (total)</td>
<td>Naive 7 day seasonal value</td>
<td>Total sales 7 days ago</td>
</tr>
<tr>
<td>Sales (total)</td>
<td>7 day diff - 28 day mean</td>
<td>Average of 7-day differenced total sales in a 28 day window</td>
</tr>
</tbody>
</table>

Consider extracting cross-series information for features other than the target, or grouping on different levels of aggregation.

Take known seasonality into account - daily data often has weekly and monthly cycles.
Time Series Learning Curves

Validation Error vs. # of rows (age of data)

- Traditional: More data is redundant
- Time Series: More data is better, more data is harmful, more data is redundant
Project Settings
Explore different FDWs and FWs

- Models spanning long horizons may not perform as well as models focused on specific forecast distances
- Generating lags across shorter or longer rolling windows of time each provide different information to the model, just find the right balance

![Diagram showing past, present, and future windows with associated terms like Blind History Gap (BHG) and Can't Operationalize Gap (COG).]
Link Backtesting Scheme to Business Goal

Make sure the time period you will use the model in is represented in the training and validation across backtests.
“All backtests” should capture performance on entire period of interest
Exponential Trends and Differencing

**Treat as exponential trend?**
DataRobot can automatically handle exponential trends using techniques like log transformations and log link functions, or you can explicitly specify the action.

- Treat as exponential trend is disabled when the target feature has zero or negative values.
  - Auto-detect
  - Yes
  - No

**Apply differencing?**
If DataRobot detects that the target is non-stationary, it can automatically apply differencing from previous values to make it stationary. Or, you can explicitly specify the action.

- Auto-detect
- Simple
- Seasonal: e.g., 2 months
- No
Advanced Workflows
Model Diversity

Integrated Models

ARIMA, ETS, RNNs, etc.

Per Forecast Distance Models

Xgboost, elastic-net, etc.

Trends and Decomposition Models

Fourier models, linear trends, decompositions, etc.
Plotting average target value over time can blur potential clusters in your data.
Certain clusters have distinct time trends, for example only Cluster 1 (orange) has a sharp sales spike in August each year. This is strong indication you may want to build separate projects per cluster!
Cluster & Build One Model per Cluster

DataRobot does some of this (performance and similarity clustering), but that’s after differencing!
What about when you have a lot of zeros?

- **Background**
  - With time series modeling there are cases where the target contains zeros
  - Examples:
    - There are a few zeros in a short interval because the store is closed on Sunday
    - There are a few non-zeros in a short interval (e.g. Rainfall of Australia)
    - There is a long zero gap (e.g. Store is closed during winter)
Hierarchical modeling can improve performance for multi-series data where many of the series are very low-volume.
What if all I care about is deep learning...

- Some problems may require the complexity that deep neural networks can handle.
- It’s important to **always benchmark** against simple alternatives to understand if something more complex is needed.
Model Evaluation and Iteration
How do we assess quality?

- First, assess performance of either a naive baseline or an existing model
- Determine how much better the best model performs in comparison
- Are we beating the baseline prediction? Is the difference significant?

<table>
<thead>
<tr>
<th>Model Name &amp; Description</th>
<th>Feature List &amp; Sample Size</th>
<th>Backtest 1</th>
<th>All Backtests</th>
<th>Holdout</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elastic-Net Regressor (mixing alpha=0.5 / Least-Squares Loss) with Forecast Distance Modeling</td>
<td>No Differencing 1 year 3 months 17 days</td>
<td>1041.8127</td>
<td>1421.7788</td>
<td></td>
</tr>
<tr>
<td>Baseline Predictions Using Most Recent Value (periodicity=7 days)</td>
<td>Baseline Only (7 days) 1 year 3 months 17 days</td>
<td>1880.2038</td>
<td>2398.1527</td>
<td></td>
</tr>
</tbody>
</table>
How do we assess quality? | Accuracy vs Time

Explore how well the model is predicting the actuals

- Are there key dates that are missed?
- Do we have all the “Known in Advance” variables?
- Are periods of poor performance correlated with important aspects of your business not in the dataset?
How do we assess quality? | Stability

Explore performance across back tests

- How similar are the scores?
- Look at date ranges, do we expect difference in performance?
- e.g., holiday volume may be higher, so higher error may be ok
## Check Series Accuracy

Are any of the series showing very different accuracy compared to the others?

<table>
<thead>
<tr>
<th>Store</th>
<th>Total Length (Rows)</th>
<th>Backtest 1</th>
<th>All Backtests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltimore</td>
<td>714 rows</td>
<td>6059.5242</td>
<td>16280.0525</td>
</tr>
<tr>
<td>Columbus</td>
<td>714 rows</td>
<td>5239.8470</td>
<td>11535.3396</td>
</tr>
<tr>
<td>Detroit</td>
<td>714 rows</td>
<td>9614.4266</td>
<td>21212.7358</td>
</tr>
<tr>
<td>Lancaster</td>
<td>714 rows</td>
<td>8120.8212</td>
<td>22004.6043</td>
</tr>
<tr>
<td>Louisville</td>
<td>714 rows</td>
<td>29568.4672</td>
<td>7874.5997</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>714 rows</td>
<td>12895.6853</td>
<td>34457.5533</td>
</tr>
<tr>
<td>Portland</td>
<td>714 rows</td>
<td>7467.4478</td>
<td>21164.2404</td>
</tr>
<tr>
<td>Richmond</td>
<td>714 rows</td>
<td>3782.2621</td>
<td>8163.9739</td>
</tr>
<tr>
<td>San Antonio</td>
<td>714 rows</td>
<td>5481.3616</td>
<td>12719.4853</td>
</tr>
<tr>
<td>Savannah</td>
<td>714 rows</td>
<td>33515.7735</td>
<td>85862.4197</td>
</tr>
</tbody>
</table>

- The difference in errors suggests the model isn't doing a consistent job across series.
- This is a sign you need to break these out - maybe into their own project.
Explore how performance changes as you forecast further into the future

- How similar are the scores?
- At some point, is the forecast accuracy comparable with the baseline model?
Blend Models by Forecast Distance

- Error starts around 550
- Error starts around 500
- Error peaks around 500
- Error peaks around 550
- Error peaks around 600
- Error peaks around 700
Discussion and Q&A

Jess Lin
Data Scientist, DataRobot

Taylor Larkin
Data Scientist, DataRobot

If we miss your question, please ask us here: community.datarobot.com
Engage, learn, and accelerate your AI/ML journey

Connect with peers to find solutions to AI challenges

Explore helpful content to take your AI to the next level

Build your brand as an AI expert & thought leader

Join your peers today at community.datarobot.com

Questions: aisuccess-webinars@datarobot.com
Thank You