Probabilistic Methods in Time Series Analysis: A Case Study

Sarah Cameron, Valerie Schnapp, Lindsay Truong

Booz | Allen | Hamilton

Introduction

Modeling real-world time series data poses unique challenges such as:

- Infrequent Reporting
- Data Inconsistences
- Nonlinear Behavior

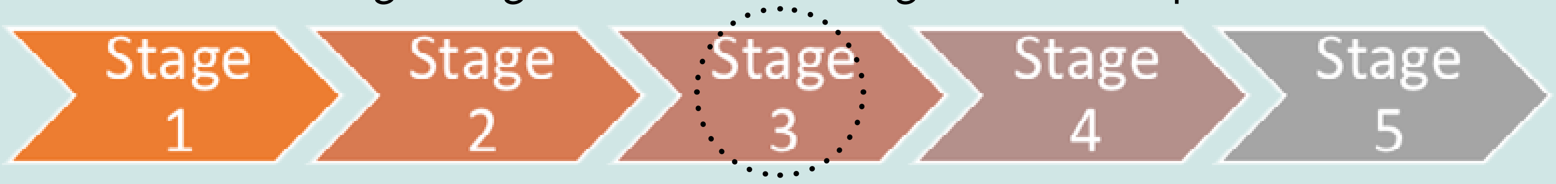
A probabilistic approach helps circumvent these issues by natively handling missing, noisy data while providing a result with measurable uncertainty.

Statement of Purpose

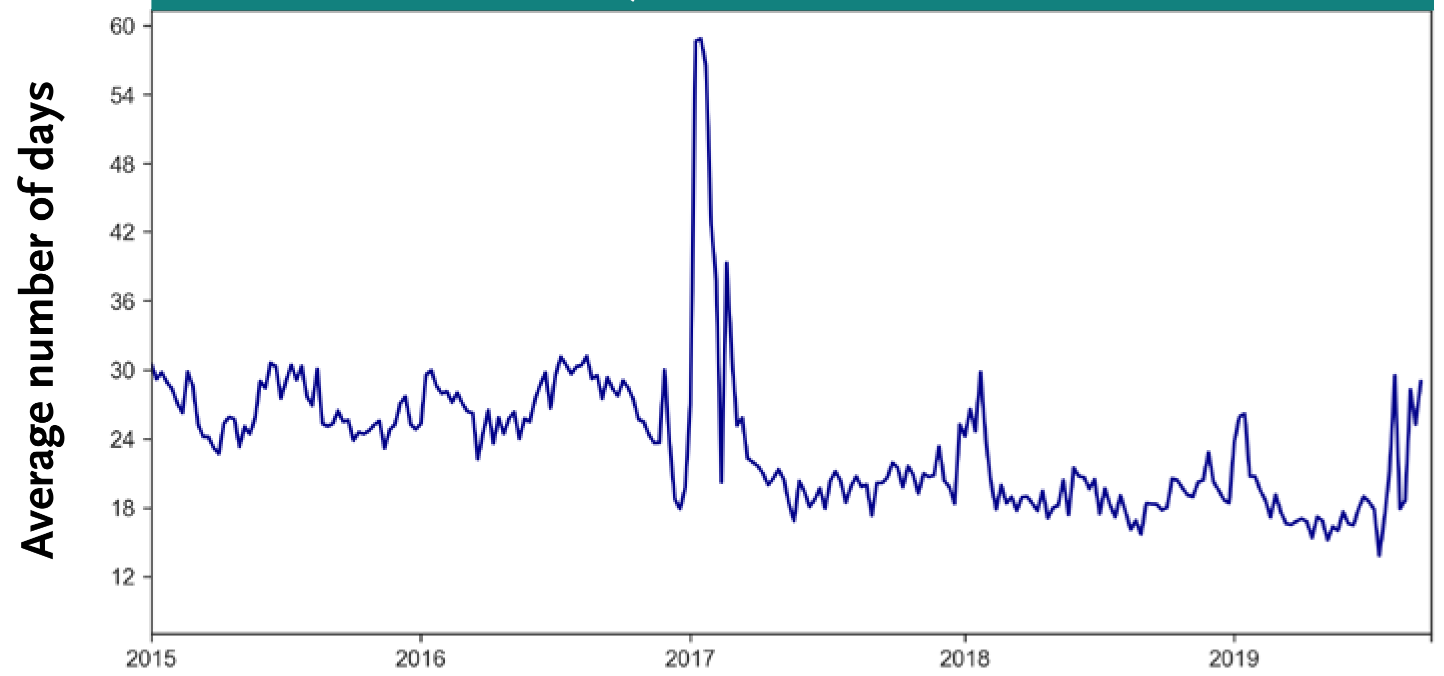
Utilize quantitative methods to enable data-driven policy decisions

Dataset

Analyzed **timeliness**: the number of days spent in a single stage within a multistage evaluation process



Weekly Process Timeliness



Weekly aggregation revealed a seasonal pattern and was the basis for all analysis

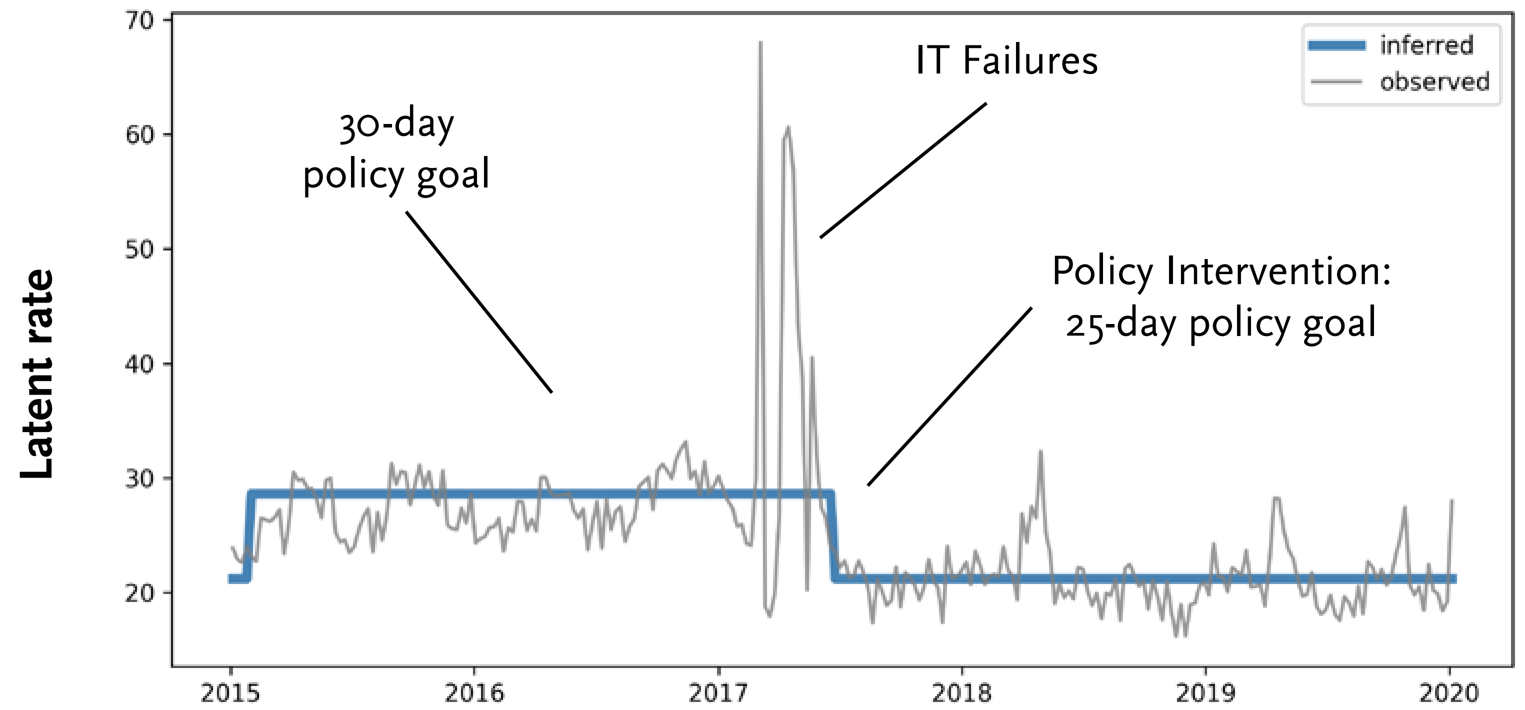
Objectives

1. Forecast future timeliness
2. Derive a data-driven timeliness policy goal
3. Characterize process behavior
4. Detect performance outliers

Model Technique 1: Change Point Analysis

Change Point Analysis was used to determine the effect of policy change on process and stage timeliness.

Change Point Analysis uses **Hidden Markov Model (HMM)** to detect structural changes in behavior. HMM uses hidden layers to detect the change points in the data and an observed layer for the time series.



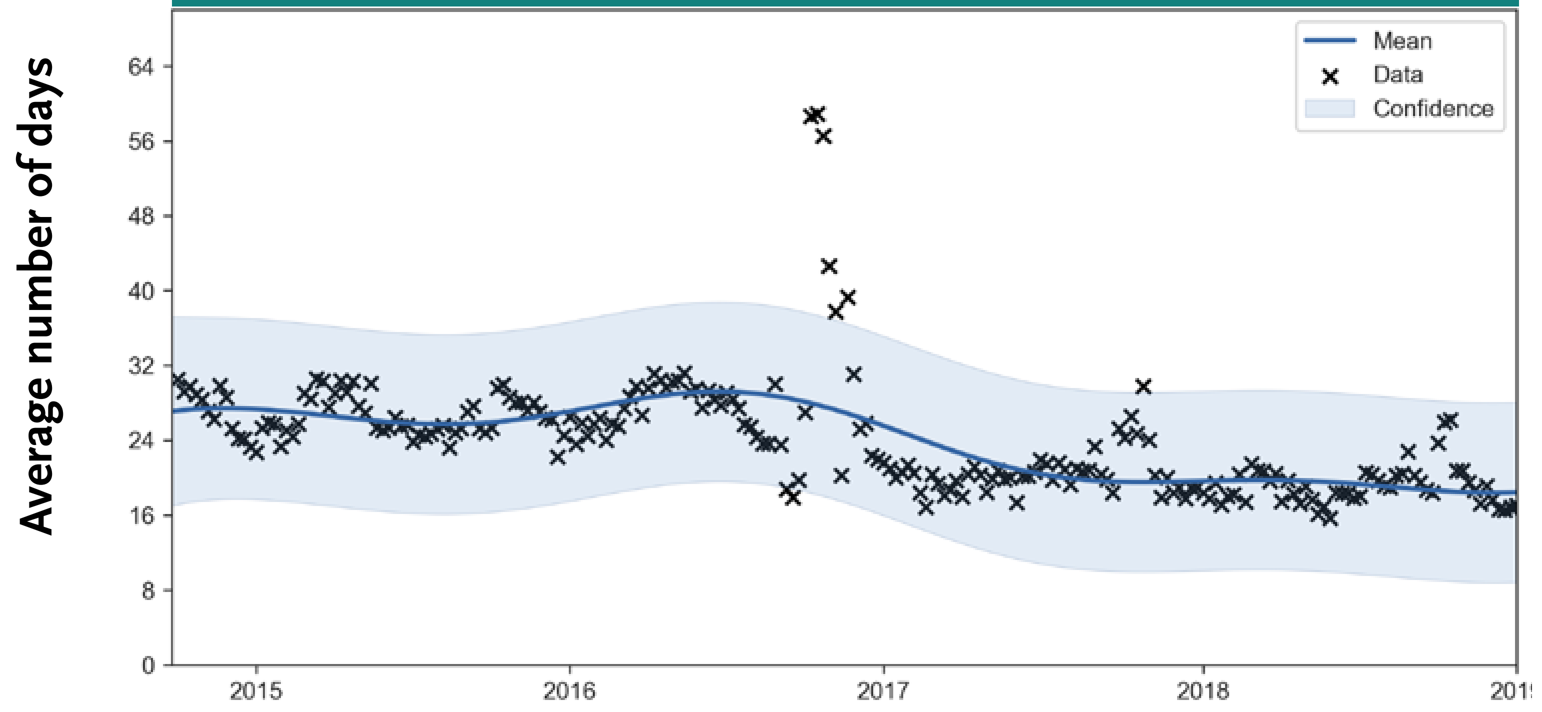
Conclusion: A noticeable shift in timeliness behavior was noticed after a policy change that was enacted in 2017.

Model Technique 2: Gaussian Process Regression

Gaussian Process Regression was used to characterize process behavior and identify performance outliers.

Gaussian Process Regression (GPR) is a Bayesian regression technique that flexibly adapts to the data, assumes noisy data and finds a distribution that best characterizes the data.

Timeliness Regression



Conclusion: Based on the overall characterization of timeliness behavior, a data- driven policy goal of 25 days was derived.

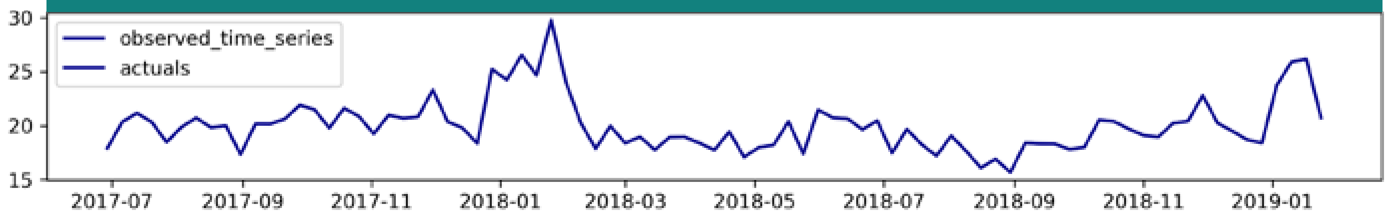
Model Technique 3: Structural Time Series

Structural Time Series was used to forecast timeliness behavior and anticipate backlog.

Structural Time Series (STS) expresses a time series as a *sum* of individual effects, including seasonality, autoregressive, local linear trends, and external influencers.

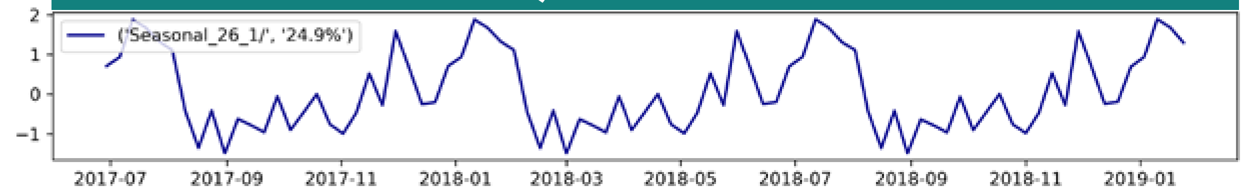
$$f(t) = f_1(t) + f_2(t) + \cdots + f_n(t) + \varepsilon;\ \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Time Series

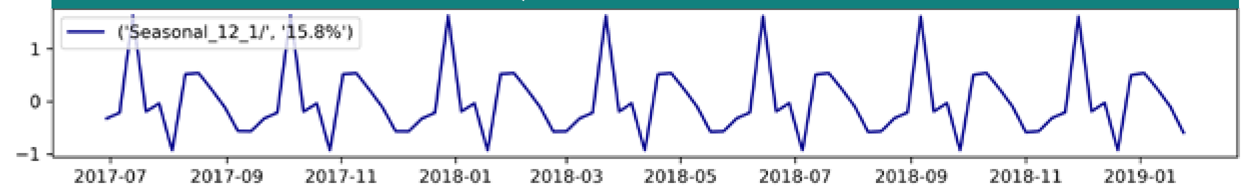


Component Decomposition

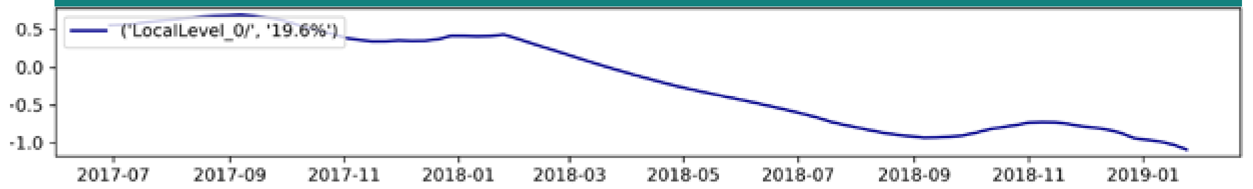
Weekly Seasonal Effect



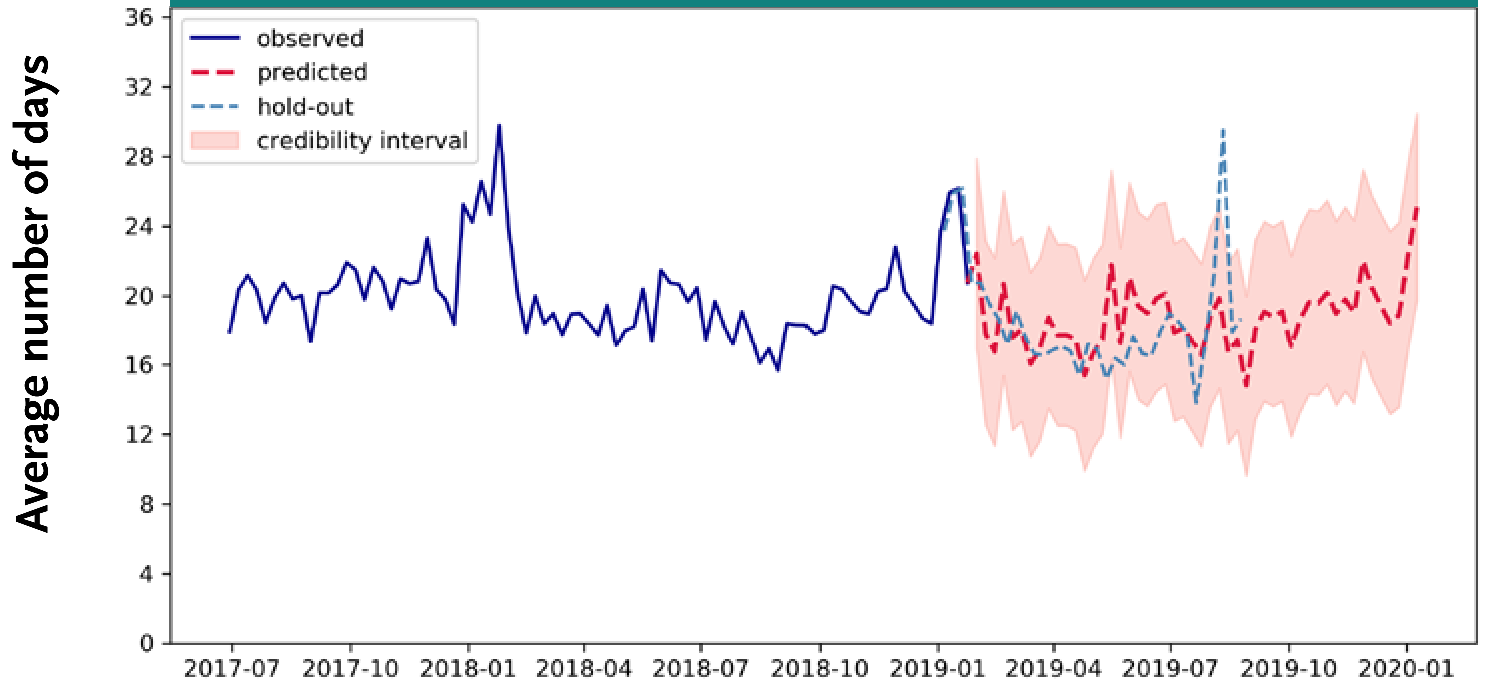
Daily Seasonal Effect



Local Linear Effect



Weekly Timeliness Forecast



Conclusion: By incorporating relevant time series components, timeliness was forecasted with narrow confidence.

The authors would like to acknowledge Chris Skiscim, Strider McGregor, Trami Pham and Will Davis for their contributions to this analysis