National Health Time Series
Methodology section

https://parleyyang.github.io/AIHACK2020/

Parley R Yang, Luis Chaves, Billy Zhao, Anthony Jancso

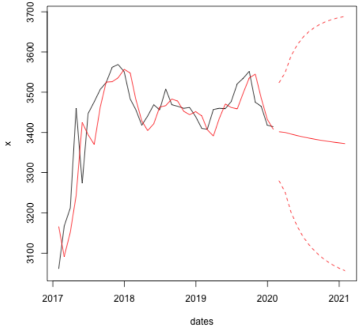
29 Feb — 1 Mar, 2020
AIHACK, Imperial College London

Set-up

- Small number of time-series observations on registered patients: up to 36 month of history available.
- Indigenous variations, e.g. patient movement, patient discharged, noises, etc.
- $\{y_{p,t}\}_{t=1}^{36}$ for each $p \in \{1, ..., 176\}$, want to forecast $\{y_{p,36+k|36}\}_{k=1}^{12}$ and rank them to give policy suggestions.

Traditional Method

- Use individual series $\{y_{p,t}\}_{t=1}^{36}$ to compute information criterion for model selection, then forecast.



- **Issue: lack of data to have large models extrapolating seasonality or ARIMA processes.**

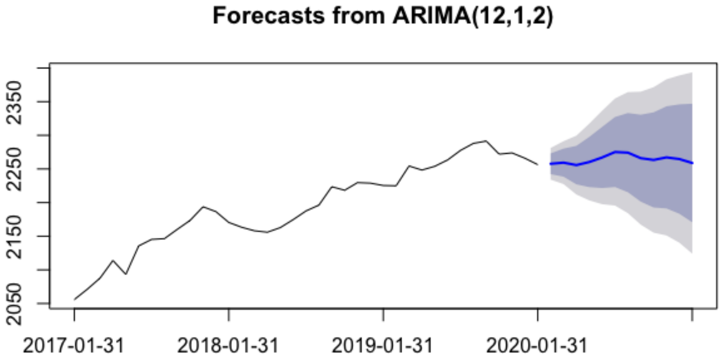
Our proposal

- Concept: Smart time series modelling using learning theory so that convergence speed is faster with small observations.
- Related time series literature: Yang (2020), Journal of Forecasting, `doi.org/10.1002/for.2676`
- Two-step method:
  1. Train model over CCG-averaged series $\{\overline{y_t}\}_{t=1}^{36}$ $\circlearrowleft$.
     Get model $f(\overline{y_t}) = g(\varepsilon_t)$
  2. Use $f(\cdot), g(\cdot)$ with constrained functional form and reasonable penalisation to train on CCG-level series, in particular,

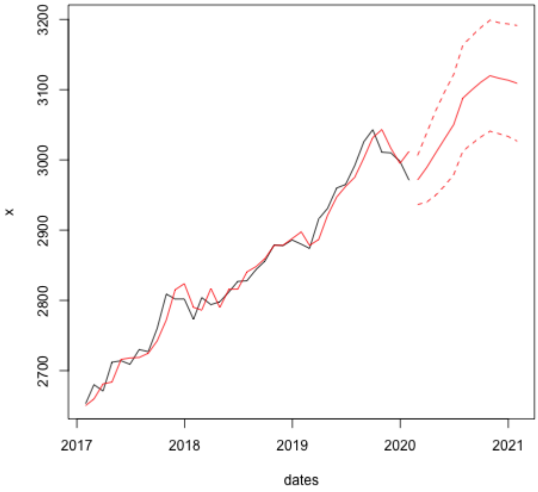
$$f_p(y_{t,p}) = g_p(\varepsilon_t)$$

     where f_p, g_p are estimated by the constrained optimisers with some penalties of $O(||f - f_p||, ||g - g_p||)$.

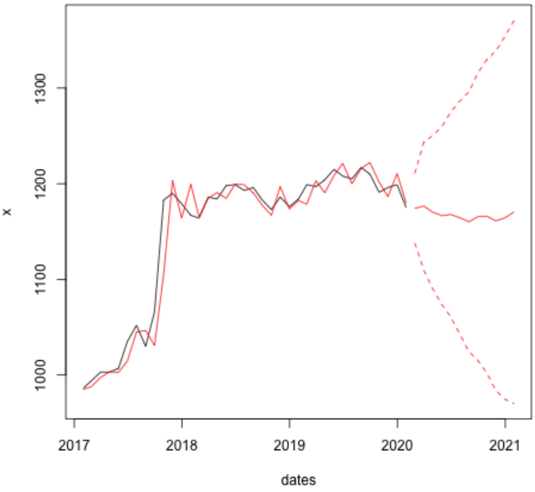
First stage


Forecasts from ARIMA(12,1,2)
2050
2150
2250
2350
2017-01-31
2018-01-31
2019-01-31
2020-01-31

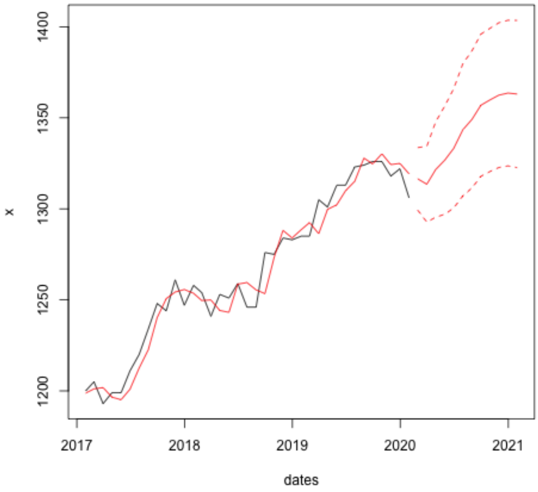
Second stage & Results



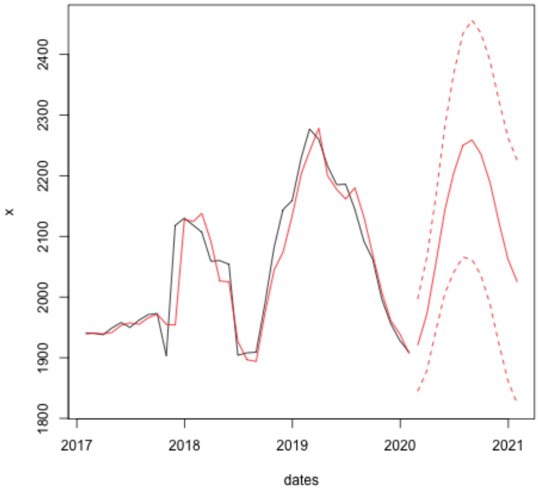
Second stage & Results



Second stage & Results



Second stage & Results



More about the application:
https://parleyyang.github.io/AIHACK2020/