Final Project

Stochastic Control and Forecasting

DHR model generalization behavior

Tim Rosenflanz

1. Summary

In this project we have explored generalization behavior of DHR model introduced by Young[1]. We construct several models on the dataset of restaurant visitors [2] and analyze generalization errors and different methodologies of extrapolating beyond the provided data. Although some serious downsides of DHR models are present, we found that DHR could consistently provide forecasts that do not show significant overfitting errors if sufficient data is provided. However, extrapolating to other similar sequences remains a challenge.

2. Approach

This paper will consist of three parts:

1. Data Construction

In this section, we will provide summary of the dataset used in the construction of the models and will justify some of the choices of the data subsetting we have done.

1. Exploration of generalization within the sample

We construct four DHR models introduced by Young [here and forth 1] that forecast number of restaurant visitors of four different restaurants. These models are then used to explore the generalization errors within the sample

1. Exploration of generalization outside the sample

We reuse insights from the previous section to create a generalized approach of constructing DHR model on the smaller samples. Specifically we explore different methodologies of averaging NVR data obtained in the previous section to predict restaurant visitors of the fifth restaurant that has much less training data and performance of such models on the unseen data.

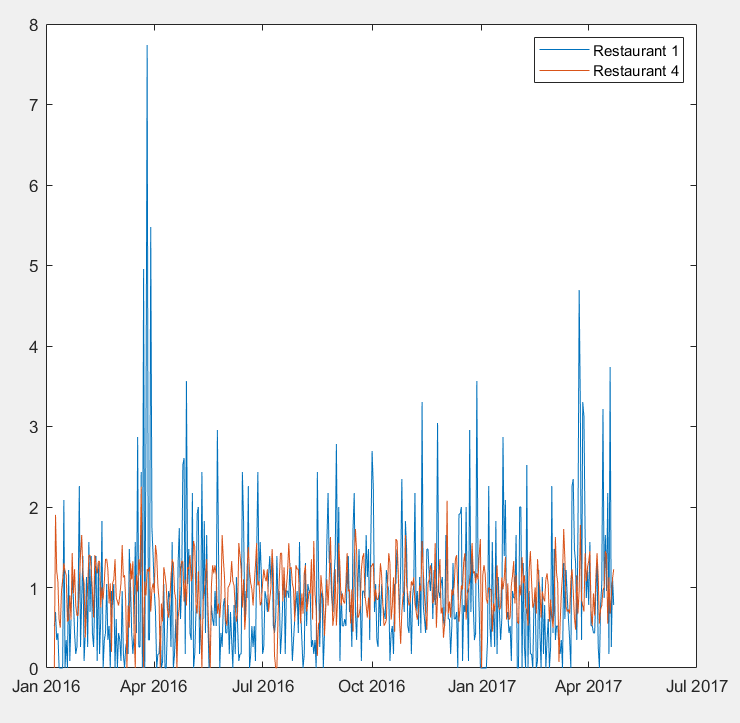
You can find all of the supporting code in the project repository[3].

3. Data Construction

For this paper we used a subset of the restaurant visitor volume forecasting from Kaggle[2]. Although the dataset includes a very large number of distinct restaurants, we will only extract five of them that have approximately the same cuisine, region, and average number of visitors during opened days. This is consistent with the approach a restaurant manager would employ when forecasting demand since restaurant location and size can have significant impact on the visitor behavior. We will not use any of the reservation data provided with the dataset and will only focus on using actual number of visitors as predictors.

Although DHR model is capable of interpolating over missing data, we fill all of the empty days as 0 visitors since in majority of cases missing data corresponds to the restaurant being closed.

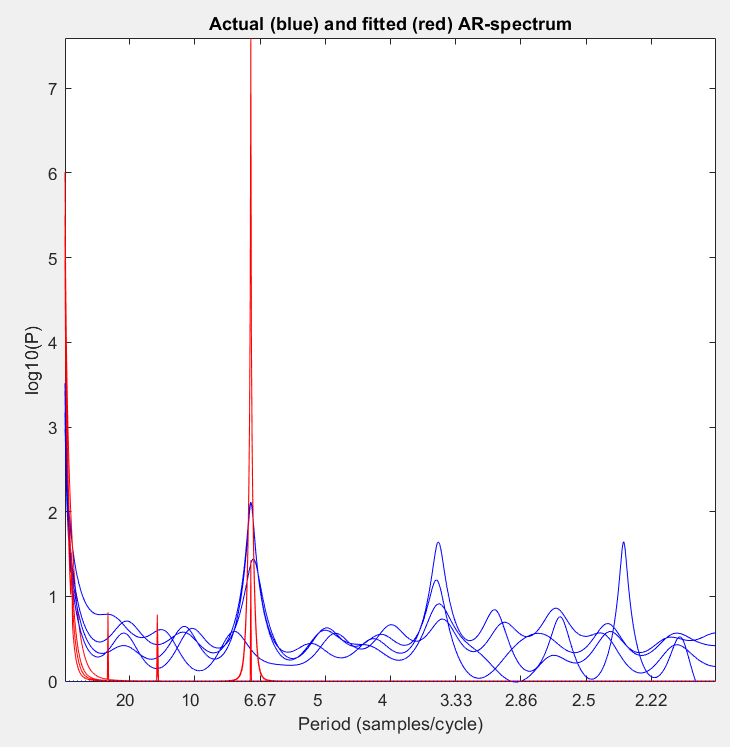
Resulting dataset consists of 5 sequences of numeric values (4 training, 1 holdout) indicating number of restaurant customers during the day and spanning up-to 471 days from January 8th 2016 through April 22nd 2017. All of the restaurants are located in Tōkyō-to Shinjuku-ku Kabukichō area and have Izakaya as their genre (it is a subtype of Japanese pub).



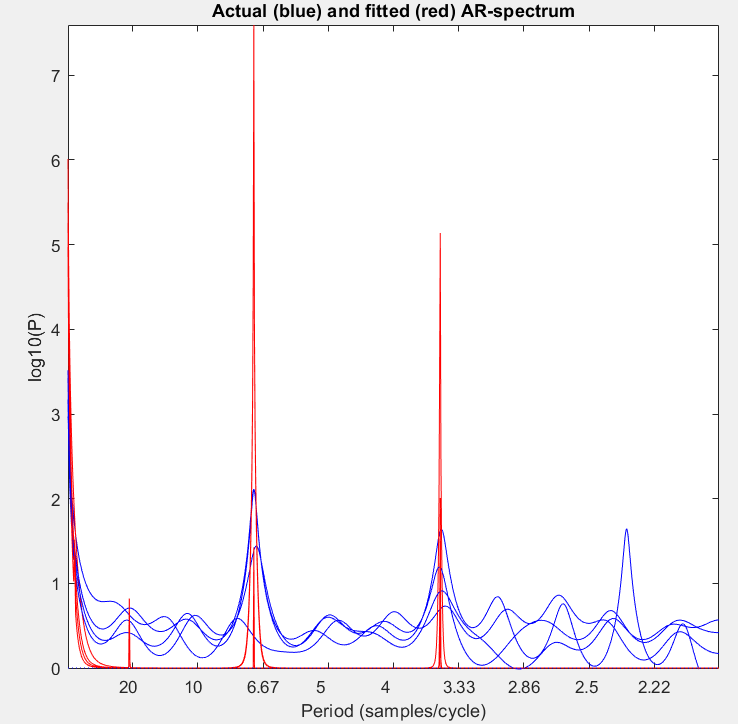
4. Exploration of generalization within the sample

Young had already explored generalization/forecasting power of the DHR model and showed that it can efficiently predict into the future when sufficiently robust model is constructed. However, we found the Air Passenger dataset to be fairly smooth without extreme values, thus it is of interest to us if DHR can also generalize to noisier data. From the plot of the data, we can see that our dataset contains some extreme outliers, significant variability from day to day and somewhat irregular components and thus it should be more challenging for DHR to fit and forecast over it.

In this section, we explore how DHR performs on each of the four training datasets in both fitting and forecasting behavior. To simplify the analysis we standardize all of the data to the same scale by dividing it over the mean number of customers during open days for each respective restaurant. We then takeout 100 last observations for each location as holdout and optimize the dhr spectra using dhropt command from CAPTAIN Toolbox [4] with trend and initial periods of 180,90, 30, 14 and 7 days. We have experimented with different settings of AR spectrum and found that default estimation method of AR-24 provides one of the best fit. Our finding is consistent with Young’s observation that increasing AR order to large values provides very noisy spectrum and makes analysis virtually impossible.

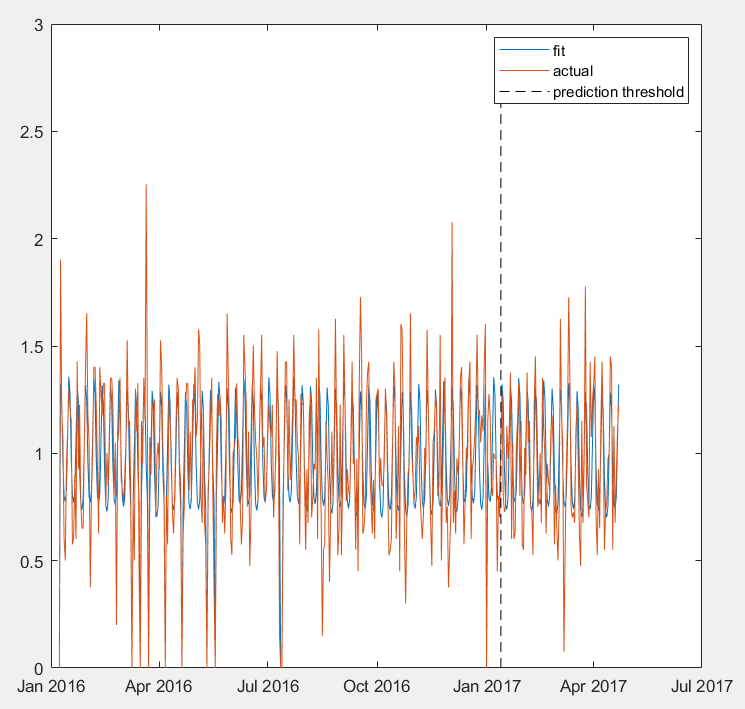
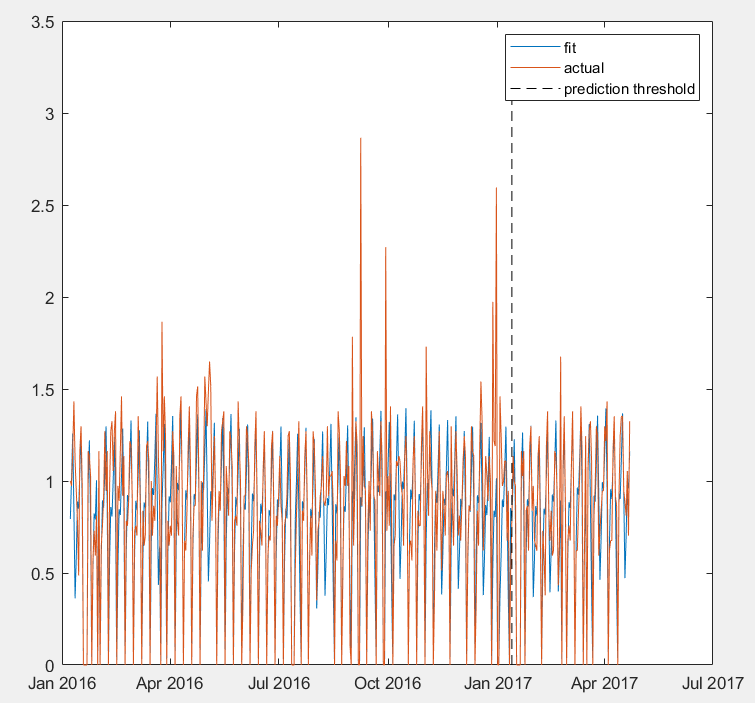
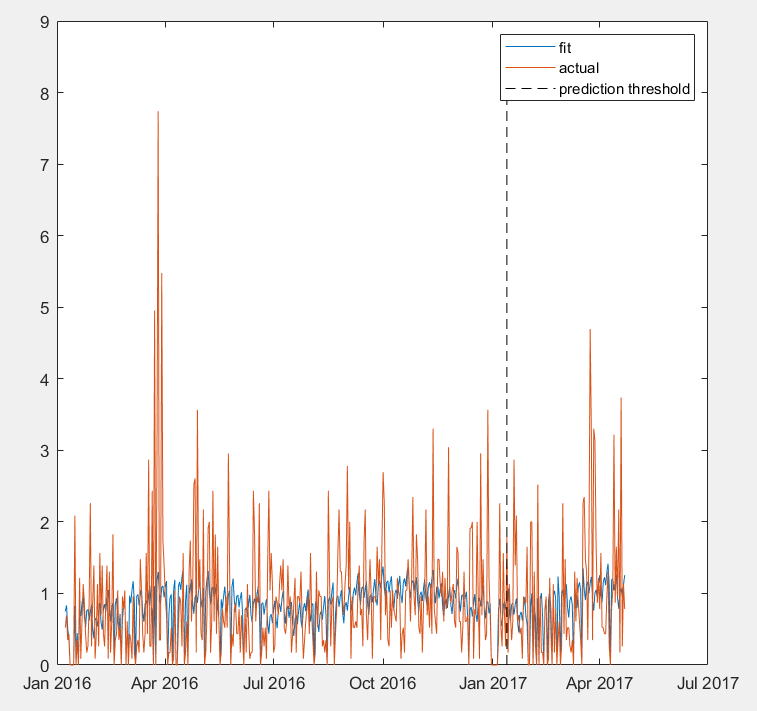


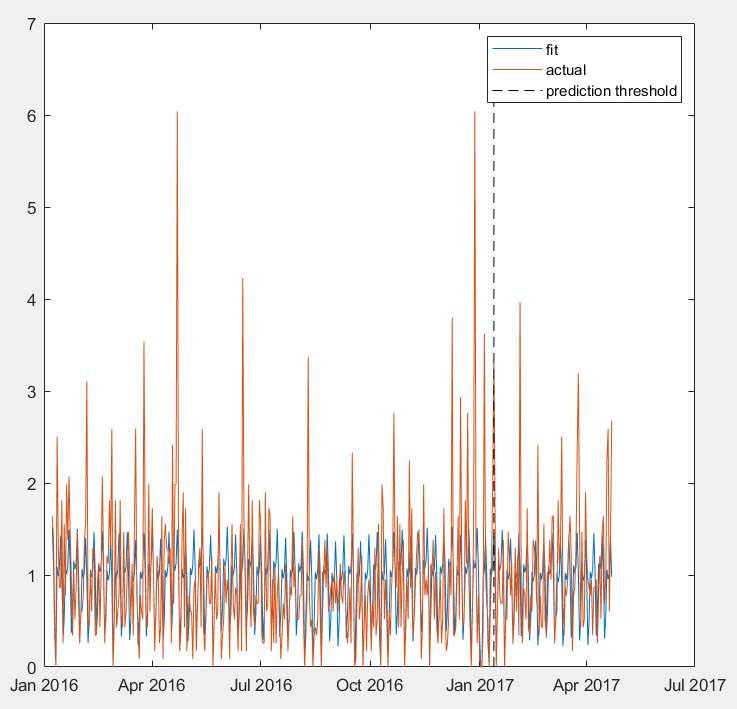
After analysis of the combined plots, we revised our periods to be 180, 21, 7 and 3.5 days, which provided a much better fit between fitted and actual AR Spectra. We hypothesize that we could find longer cyclical components with more years of data and averaging some of the days together to denoisify the data.



Careful analysis of the four sets of NVRs showed that there was some significant variability in shorter components but longer components were no more than a magnitude away.

We now fit four DHR models and analyze their output against the ground truth data. Since absolute majority of restaurant closures are planned we will simply replace fit values for closed days with 0s – one could argue that we should delete these days from the data but it will make consolidating the results together challenging since different restaurants might have different schedules.

Following graphs and tables summarize the fit model and the forecasted values:

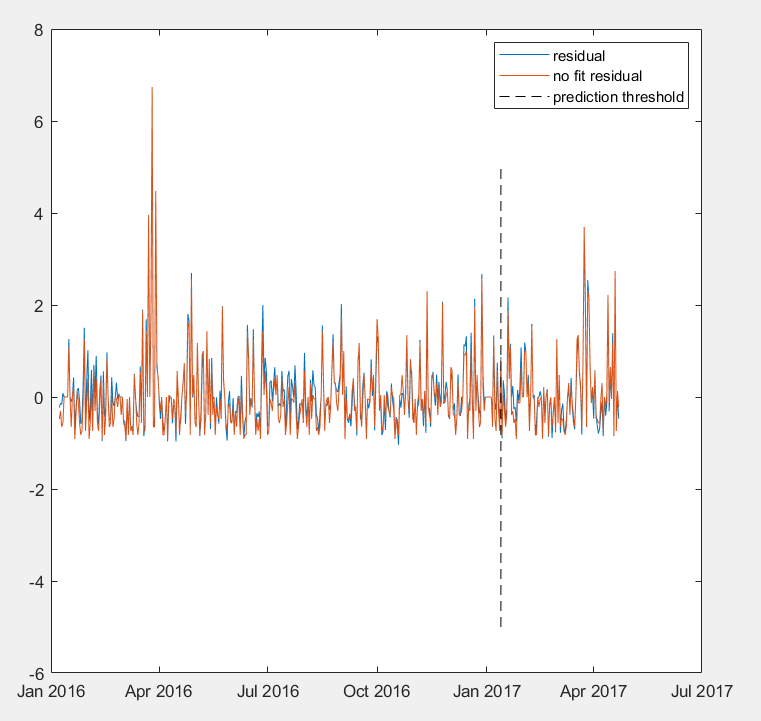
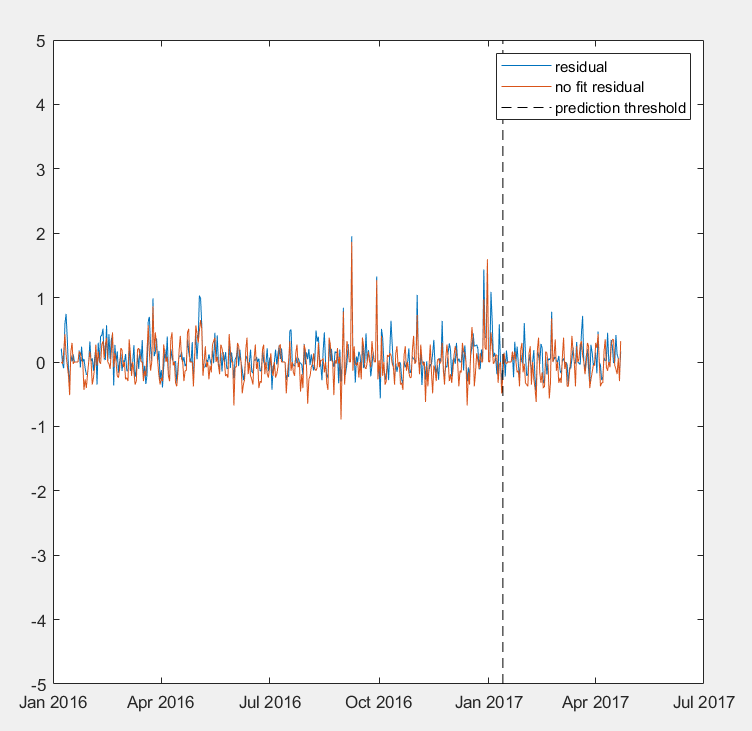
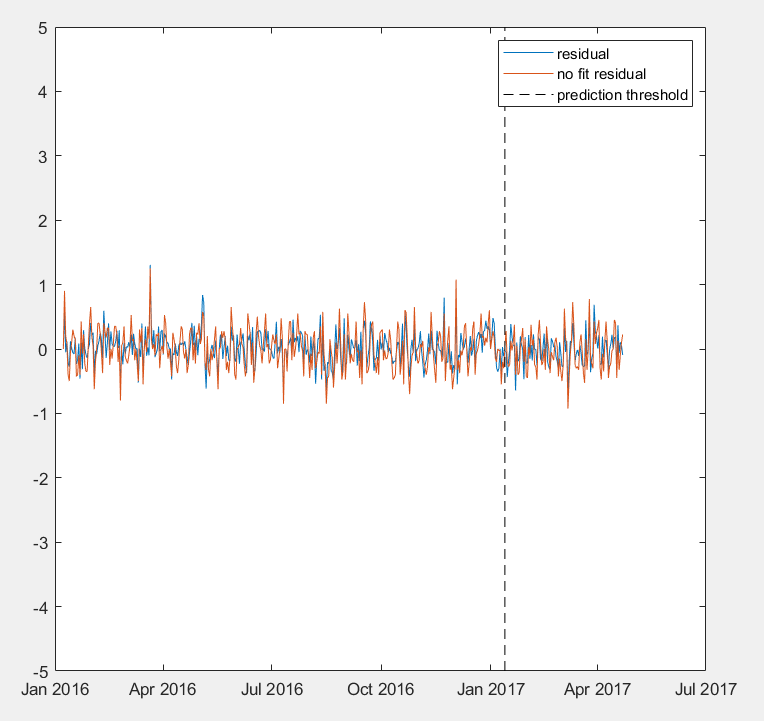


Fit/Forecast vs Actual values for Restaurant 2

Fit/Forecast vs Actual values for Restaurant 1

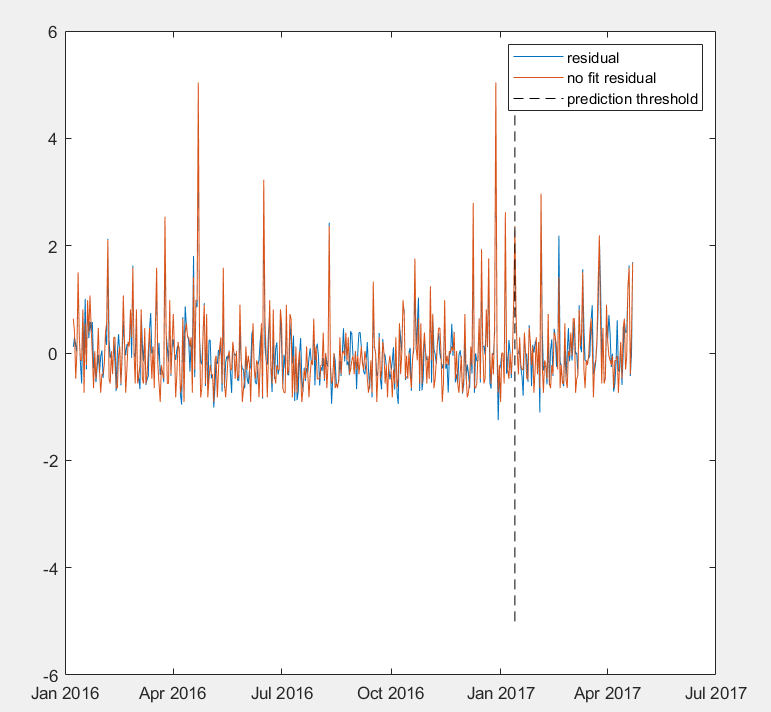
Fit/Forecast vs Actual values for Restaurant 3

Fit/Forecast vs Actual values for Restaurant 4



Residuals vs Predicting Average for Restaurant 2

Residuals vs Predicting Average for Restaurant 1



Residuals vs Predicting Average for Restaurant 4

Residuals vs Predicting Average for Restaurant 3

|  |  |  |  |
| --- | --- | --- | --- |
| Restaurant | MAE on Training data | MAE on Forecasting | MAE if no fitting is done |
| 1 | .502 | .582 | .539 |
| 2 | .187 | .147 | .202 |
| 3 | .434 | .471 | .504 |
| 4 | .187 | .185 | .253 |

We choose to report Mean Average Precision (MAE) as the metric of choice since it is less sensitive to severe outliers in the data compared to such metrics as Root Mean Squared Error and Mean Squared Error. We also compare fit models’ performance to just predicting average number of customers on open days and 0 on closed.

It is apparent that the model can generally follow the visitor pattern for each restaurant even though there is significant daily variation but heavily under predicts when extreme values are present. This is unsurprising since DHR relies on cyclical components and there is no mechanism that could predict outliers even though they follow some underlying logic that can explain these outliers (such as last week of December having many more customers than other weeks). One could argue that logarithmic transform of the data could smooth these outliers but we found that DHR model fit on log of the data performed worse than on original scale – this is unsurprising since Young created DHR to explicitly not require any logarithmic transformation.

Overall, we found that all four models performed better than predicting average number of customers on the fit data and 3/4 were better in forecasting than predicting average. Moreover, restaurants 2 and 4 showed *better* forecasting performance than on the fit data –possibly due to lower number of severe outliers in the forecasted period. Upon analyzing the data closer, it appears that when daily variation in number of visitors is low (restaurants 2 and 4) DHR performs much better than naïve fit and close to naive when the variation is higher (restaurants 1 and 3)

All of our findings are consistent with theoretical understanding of DHR: optimization and comparison to actual AR spectrum can uncover significant cyclical components missed by the analysts, modeling can capture consistent cyclical/seasonal components in the data, and forecasting is done with fairly small reduction in performance. Extreme value prediction remains a challenge but this is ubiquitous to time series models that rely on posing outliers as statistical anomalies rather than predictable events.

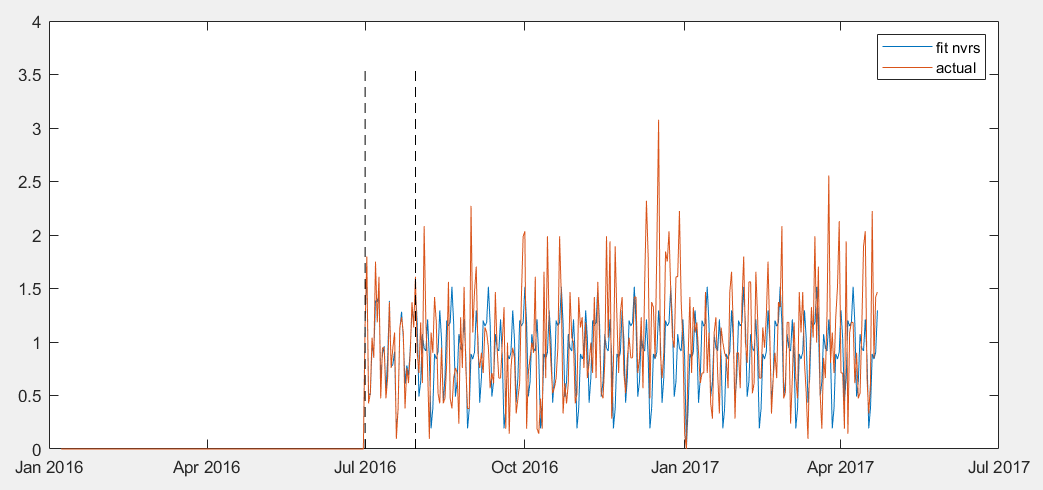
5. Exploration of generalization outside the sample

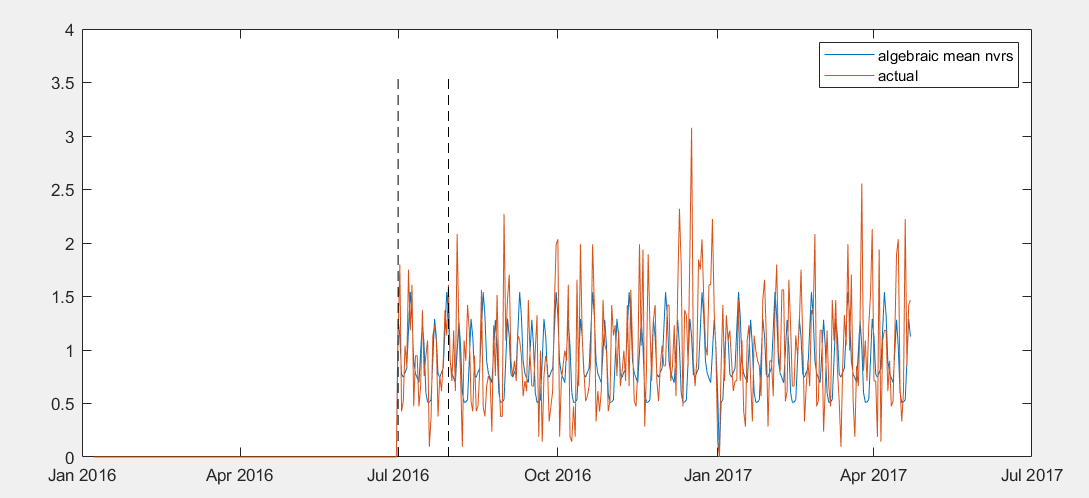
We have found that DHR can generalize/forecast well when provided with sufficient information from the sequence; however, it remains to be seen if we can extract sufficient information from the constructed models to forecast on sequences coming from similar source but with much less data. To explore this behavior, we will use just 30 days of the 5th holdout restaurant for which we have not previously constructed a DHR model and compare model constructed via fitting AR spectrum on just 30 days versus using a combination of spectra from the previous section. Theoretically, one could see this as asking “If we didn’t close restaurant after 30 days, what would be its daily visitors based on what we know about 4 other restaurants”. We require 30 days of data as initial seed to construct a baseline model.

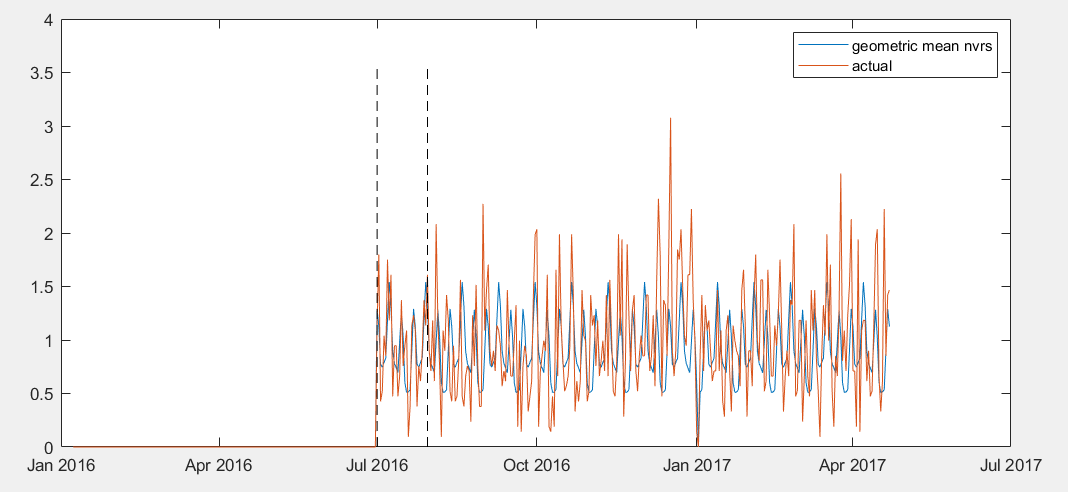
Such analysis is straightforward via averaging NVRs of the four already constructed models and using them to run a DHR model. Since DHR model fitting relies on the NVR values for the estimation, it is essentially equivalent to extracting Spectra information from the 4 constructed models. It is not unreasonable to drop long term cyclical component (180 days in this case) since DHR cannot effectively extract cyclical component that is so much longer than observed values.

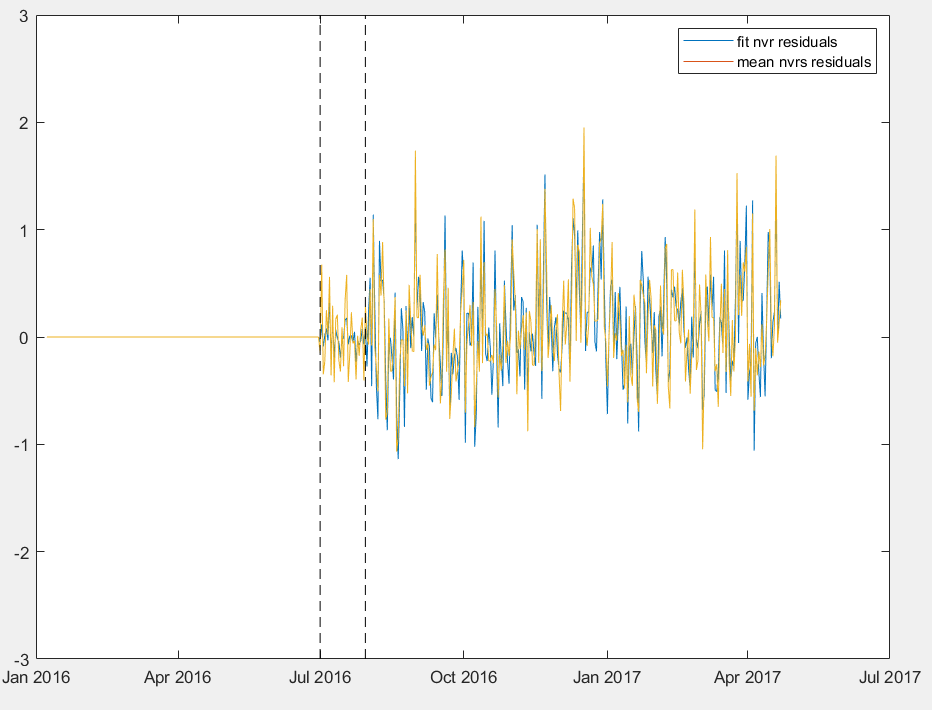
|  |  |  |  |
| --- | --- | --- | --- |
| Component Length | NVRs fit to holdout data | Algebraic Mean of NVRs from 4 training model | Geometric Mean of NVRs from 4 training model |
| 0 days (trend) | 9.38e-09 | 2.33e-05 | 3.46e-05 |
| 21 days | 1.14e-08 | 1.93e-12 | 2.09e-12 |
| 7 days | 2.36e-13 | 5.19e-10 | 6.38e-07 |
| 3 days | 0.9570 | 1.02e-10 | 1.96e-08 |

Training NVR means appear to be considerably different from the ones obtained by fitting the data to the 30 days of holdout data. This drastically changed DHR models fit to the holdout data. Graphs summarizing 3 models fit to the holdout data are presented below:









|  |  |  |
| --- | --- | --- |
| NVR type | MAE on Fit data | MAE on Forecasting |
| Extracted | 0.0917 | 0.4048 |
| Algebraic Mean | 0.2646 | 0.3870 |
| Geometric Mean | 0.2646 | 0.3870 |
| Predicting Average/No fit  (Forecasting/Fit combined) | 0.2547 | 0.2547 |

Surprisingly, models fit using algebraic and geometric means of NVRs were almost identical even though the NVRs themselves were quite different. Nonetheless, they performed much worse on the fit data compared to model constructed using AR spectra extracted from the data. This is consistent with our expectations since extracting NVRs directly from holdout data provides much more insight into the available information. However, mean NVR models were better at forecasting than the extracted NVR model. We hypothesize that these mean NVR serve as a regularizing factor since they bring the fit values closer to the trends and cycles observed in other restaurants thus minimizing the variance of data.

We hypothesize that regularizing effect of Training NVRs can be exploited to create models that combine extracted AR spectra from both holdout and training data. Our initial experiments provided some evidence towards this hypothesis, however further experiments are necessary. Moreover, it is reasonable to believe that it is possible to pose this combination of NVRs as a control problem and develop an algorithm that dynamically changes NVR combination for different cyclical periods.

Unfortunately, when compared to just predicting average number of customers none of the created DHR models are satisfactory and thus we are forced to conclude that DHR cannot generalize well with low number of samples and high variance even when some auxillary data from other sequences is provided.

6. Conclusions

In this paper, we explored DHR model behavior on data with high variance and a number of extreme values. Our findings are summarized below:

* DHR generalizes well to forecasting data when sufficient amount of data is present for fitting
* Outliers are not forecasted well by DHR and logarithmic transforms do not produce better models
* Longer cyclical components have lower variability of NVRs than shorter ones
* NVRs extracted from longer sequences can be helpful in regularizing DHR model fit on shorter sequences. This behavior could be explored further by combining NVRs in different ways
* Short sequences do not generalize well to forecasting but provide excellent fit to provided data. I.e. DHR models are prone to overfitting on low number of samples
* Predicting mean value is better in forecasting than constructing a model on small sample.

References

1. Young, P. C., Pedregal, D. J. and Tych, W. (1999), Dynamic harmonic regression. J. Forecast., 18: 369-394. doi:[10.1002/(SICI)1099-131X(199911)18:6<369::AID-FOR748>3.0.CO;2-K](https://doi.org/10.1002/(SICI)1099-131X(199911)18:6%3C369::AID-FOR748%3E3.0.CO;2-K)
2. “Recruit Restaurant Visitor Forecasting | Kaggle.” Countries of the World | Kaggle, [www.kaggle.com/c/recruit-restaurant-visitor-forecasting/data](http://www.kaggle.com/c/recruit-restaurant-visitor-forecasting/data).
3. Tim Rosenflanz, DHR Generalization Github, <https://github.com/tRosenflanz/dhr_generalization>
4. Young, and Lancaster. *THE CAPTAIN TOOLBOX*, 4 Apr. 2018, captaintoolbox.co.uk/Captain\_Toolbox.html/Captain\_Toolbox.html.