Real-Time Reporting – An Evaluation of Approaches to Prediction of Mature Monthly Metrics

Maria L. Joseph-King, Ph.D.[1], Brian O'Donnell, Ph.D[1], Michelle Roozeboom-Baker, Ph.D[1], Eric Rollins[2], Geoff Gerhardt[2]

[1]General Dynamics Health Solutions

[2]Centers for Medicare & Medicaid Services

Motivation

The Centers for Medicaid & Medicare Services (CMS) Chronic Condition Warehouse (CCW) is used as the data source for several CMS real-time reporting initiatives. Measures of quality, cost and utilization are computed based on all claims for a given month. Unfortunately, due to the nature of receiving, processing and finalizing claims data, the mature monthly metric value may not be observed until a number of months after the service month. The non-mature monthly metric value that we observe in more recent months is an error-prone estimate of the mature monthly metric value.

Research Objective

To characterize the relationship between observed and mature monthly metric values, and to evaluate approaches to predict mature values prior to maturity.

Data

We use fee-for-service claims on a 100% sample of Medicare beneficiaries. Monthly metrics are computed subject to processing dates between JAN 2008 and DEC 2014. The interval of time between the last date of service on a claim and when it is processed is called "run-out". We identify a claim as "mature" at 12 months of run-out. Observed metrics at zero to 11 months of run-out, and mature metrics are available for service months between JAN 2008 and DEC 2013. There are hundreds of metrics summarizing quality, utilization, and cost for several service months, service categories, and enrollment statuses. The metric of interest for this analysis is the total Medicare allowed amount on acute inpatient claims for beneficiaries with Parts A & B coverage.

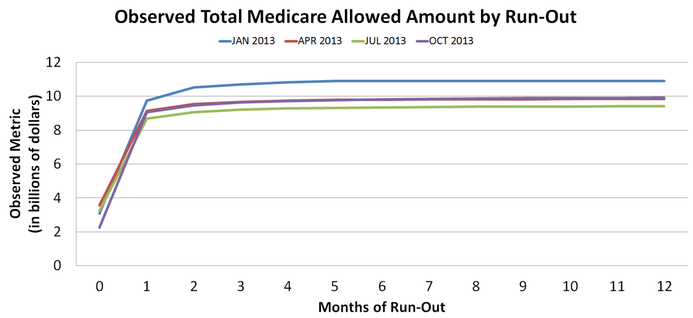

Figure 1. Observed metric at various months of run-out for select service months in 2013.

Methodology

Linear Regression

We use a simple linear regression model to characterize the relationship between the mature and the observed metric values at various months of run-out. Twenty-four months of data were used to train the model.

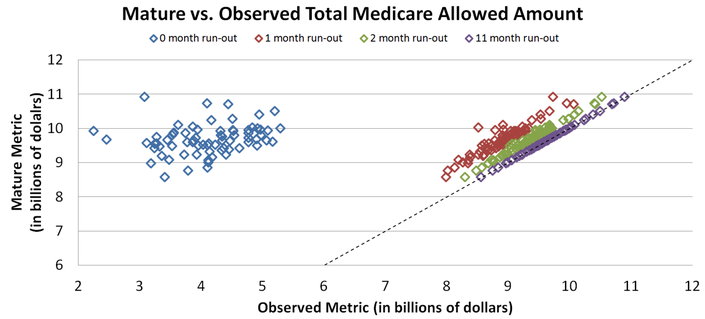

Figure 2. Mature metric values against observed metric values at various months of run-out.

Seasonal Auto-Regressive Integrated Moving Average (SARIMA)

We use a SARIMA(1,1,0)×(12,0,1,0) time series to model the trend in the historic mature metric values, and to predict future mature metric values. Thirty-six months of data were used to train the model.

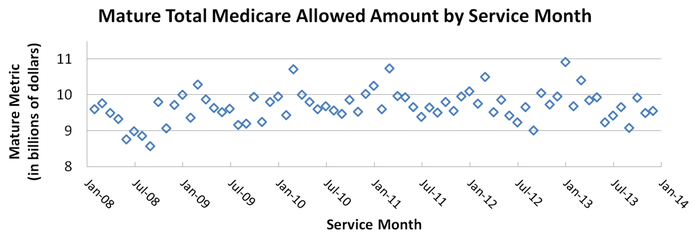

Figure 3. Mature metric value over time.

Kalman Filter

We use a Kalman filtering algorithm to update the predicted mature metric from the SARIMA model, by incorporating the observed non-mature metrics. The weight allotted to each the SARIMA and observed predictions in the Kalman model is determined by prediction error variance of each component.

Results

We fit the three models to subsets of the analytic sample, defined by the month in which the data were processed (JAN 2012 to DEC 2013). We used each model to predict the mature metric value for each service month at zero to 11 months of run-out. The following SAS procedures were used to compute the regression, SARIMA, and Kalman predictions, respectively: PROC REG, PROC ARIMA, PROC IML.

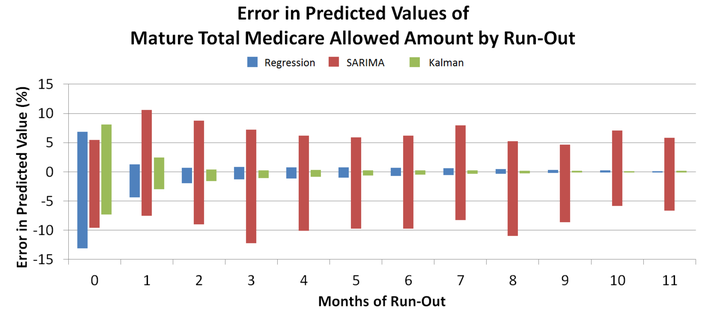

Figure 4. Difference in predicted and mature metric value at various months of run-out for Regression, SARIMA, and Kalman approaches. Error in predicted value is 100×(Predicted Value - Mature Value)/Mature Value. Each bar summarizes the range of the error in predicted values for at least 24 service months.

Discussion

The three models result in little to no bias in the predicted values of mature total Medicare allowed amount, and perform comparably at zero months of run-out. Months of run-out have little influence on the SARIMA predictions, while the Regression and Kalman predictions improve in precision as more months of run-out are observed.

At one or more month or run-out the Regression and Kalman predictions outperform the SARIMA predictions with respect to precision. Although the Kalman and SARIMA models are more sophisticated, they require specification of several parameters which can be cumbersome when several hundreds of metrics are of interest. The Regression model is preferred for our analysis due to its simplicity and comparability to the Kalman model.

● This material was developed under contract with the Centers for Medicare & Medicaid Services ●