Exploratory Data Analysis - Final Project

Ayesha Bhimdiwala, Harsh Mehta, Khusaal Giri, and Vaishnavi Srinivasan

04/13/2018

Introduction:

The United States national debt amounting to $20 trillion has Health Care spending as its driving factor. The national debt is projected to soar to an unsurmountable $92 trillon over the next three decades.

Medicaid is a joint federal-state program that provides health coverage or nursing home coverage to certain categories of low-asset people, including children, pregnant women, parents of eligible children, people with disabilities and elderly needing nursing home care. [Credit: Medicaid]

Federal Medicaid spending has risen to $389 billion since 2008 and is projected to soar to $650 billion in a decade. This expenditure is driven by a system that allows for the states to determine the eligibility and its benefits.

Medicare is the federal health insurance program for:

- People who are 65 or older
- Certain younger people with disabilities
- People with End-Stage Renal Disease (permanent kidney failure requiring dialysis or a transplant, called ESRD) [Credit: Medicare]

Medicare is a federal driven program and Medicaid is a state and federal driven program that provides health coverage for 74.6 million low-income and disabled Americans, as described above. [Credit: Medicare vs Medicaid]

The Trump presidency is looking to curb the funds for certain group of people whom they claim are physically capable of earning an income above the federal poverty line. As of December, 18 states were waiting for CMS to decide whether to approve substantive waivers to reshape their Medicaid programs, many of them by making it harder for people to qualify for or maintain their benefits. Based on rate per discharge categorized by the disease group, the government allocates the Medicare funds for the hospitals covered under this program. This led us to explore the Medicare insurance data to analyze the government spending patterns on the type of diseases the government invests in and if it has a relationship with the number of discharged people with that disease. [Credit: Governing.com]

Research Question - Does increase in the discharge count increase the Average Medicare Payment Amount when categorized by disease group?

To answer this, we focus on performing data analysis on the Provider Utilization and Payment Data Inpatient Public Use File (referred to as "Inpatient PUF") prepared by the Centers for Medicare & Medicaid Services (CMS). It contains hospital-specific charges for more than 3,000 U.S. hospitals. As described above, the hospitals receive Medicare based on a rate per discharge categorized by the Medicare Severity Diagnosis Related Group (MS-DRG). The Inpatient PUF is available for the years 2011 through 2015. The MS-DRG are limited to the top 100 most frequently billed discharges.

Data Description

We will consider three explanatory variables, DRG Description, Provider State, and Total Discharges, and one response variable, Average Medicare Payments, for our analysis.

- DRG_Description: Comprises the code & description identifying the Diagnosis Related Group(DRG). DRGs are a classification system that groups similar clinical conditions (diagnoses) and the procedures furnished by the hospital during the stay. In simple terms, they are different categories for various classes of diseases.

- Provider_State: State in which the provider (hospital) of the Medicare Services is physically located.

- Discharges: The number of discharges billed by the provider for inpatient hospital services.

- Average_Medicare_Payments: The average amount that Medicare pays to the provider for Medicare's share of the MS-DRG. Average Medicare payment amounts include the MS-DRG amount, teaching, disproportionate share, capital, and outlier payments for all cases. Medicare payments DO NOT include beneficiary co-payments and deductible amounts nor any additional payments from third parties for coordination of benefits.
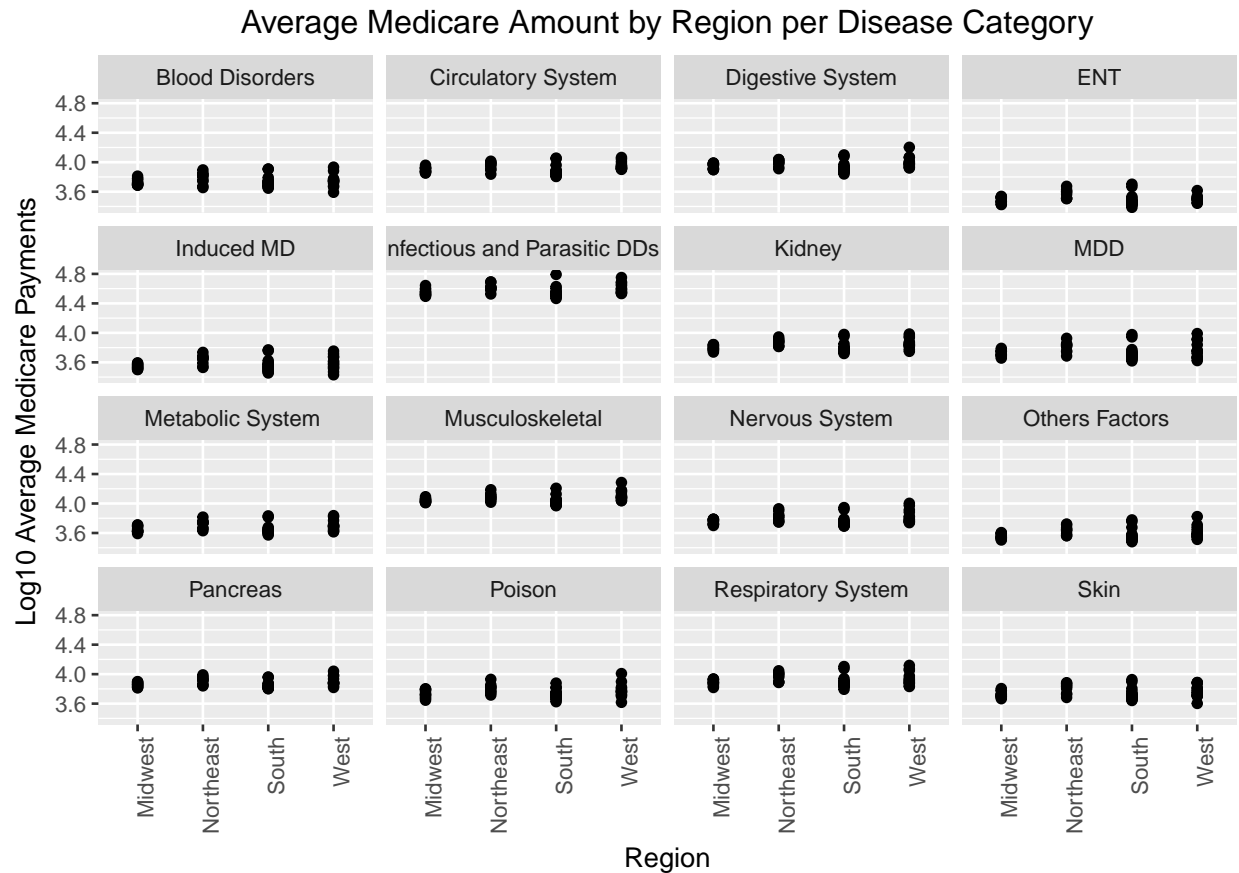
[Credit: CMS]

Data Preprocessing:

- We have consildated the 100 DRG description into one of 25 Major Diagnostic Categories (MDC) that generally correspond to a single organ system. [Credit: MDC]

- We added a new factor variable to categorize States by their Regions.

- Referencing state population defined in Wikipedia, we added a new explanatory (continuous) variable, "Proportion of Discharges per 10000"" of the state population.

Descriptive Analysis:

Average Medicare Payments in different US regions

As the first step of our analysis, we plot the Average Medicare Payments against different US regions and facet it by different diseases. We have transformed the average medicare payment to a log (base 10) scale to increase their variability.

Average Medicare Amount by Region per Disease Category

We see that among all the diseases for all the regions,

- Infectious and Parasitic Diseases show the highest Average Medicare Payments.
- Induced Mental Disorder (Induced MD) has the lowest Average Medicare Payments.
- Midwest has the lowest Average Medicare Payments.
- West has the highest Average Medicare Payments.

Infectious disease covers dieases caused by parasites that generally detoriate the hosts body to survive. These diseases can cause damages to the functioning of multiple organs in a human body, such as kidney failure or digestive system issues. They also have a high cost associated with its tests and diagnosis. The average number of people diagnoised is high for the Southern region as its states are more prone to such parasitic diseases compared to other regions. Parasitic infections have been shown to flourish in settings of warm climate and extreme poverty found in southern states of the US. [Credit: NTD in USA]

The medicare aid is offered to people over 65 years of age. Between 2013 and 2015, an estimate of 54.4 million US adults (22.7%) annually have been diagonised with some form of arthritis. Thus, it's standing as the second highest average Medicare aid makes sense. [Credit: CDC arthritis]

Most of the Americans are affected by diseases related to the circulatory organs, and so their higher discharge rate is quite reasonable. Based on CDC statistics in United States,

- About 610000 people die of heart disease every year–1 in every 4 deaths
- Coronary heart disease (CHD), most common heart disease, kills over 370000 people annually.
- Every year about 735,000 Americans have a heart attack.

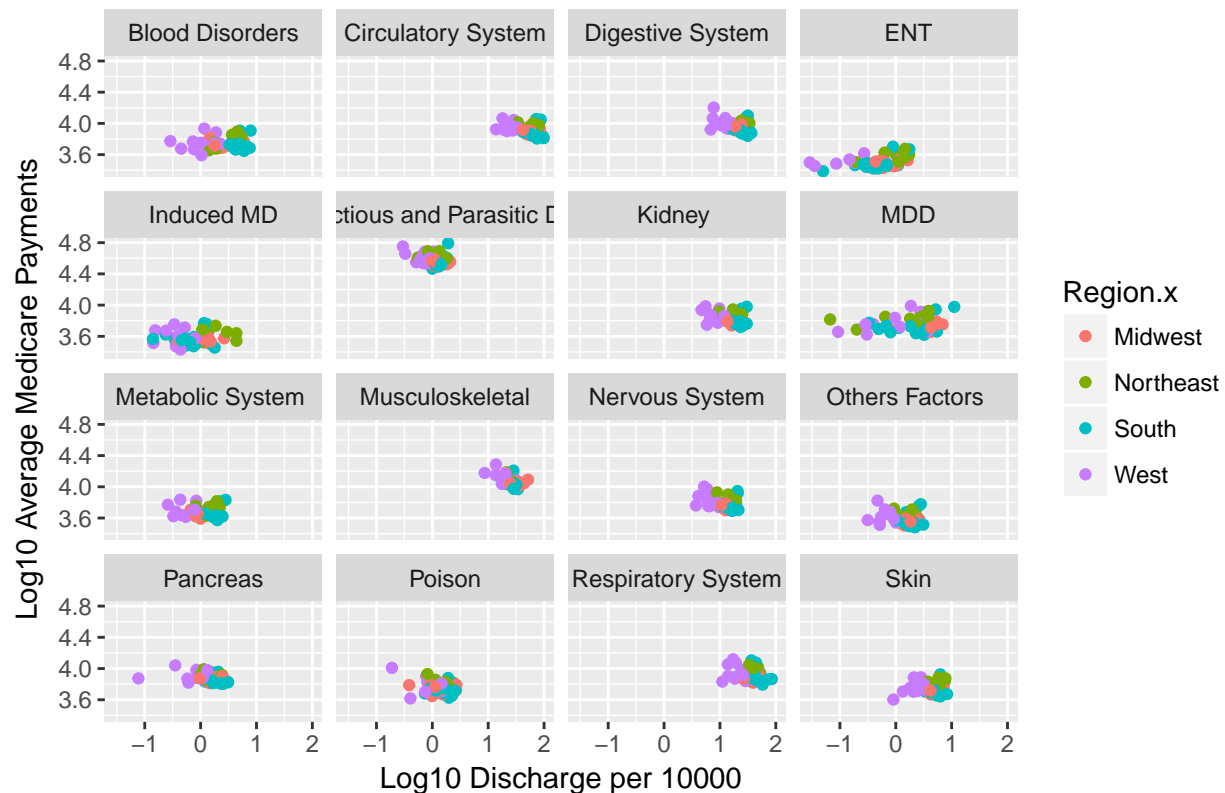And so, it has a high average medicare payment amount. [Credit: CDC]

This low value of the Average medicare payments for Induced MD (Alchohol or Drug Induced Mental Disorders) is straightforward as insurance policies do not generally cover or cover a base amount for the

beneficiaries who tend to be alcoholic or addicts.

Average Medicare Payments by Discharge per 10000

Next, we move onto studying the relationship between average medicare payments and discharges per 10000 categorized by diseases. Again, we have transformed the average medicare payment and discharges per 10000 to a log (base 10) scale to increase their variability.

Average Medicare Amount by Discharge per 10000 per Disease Category
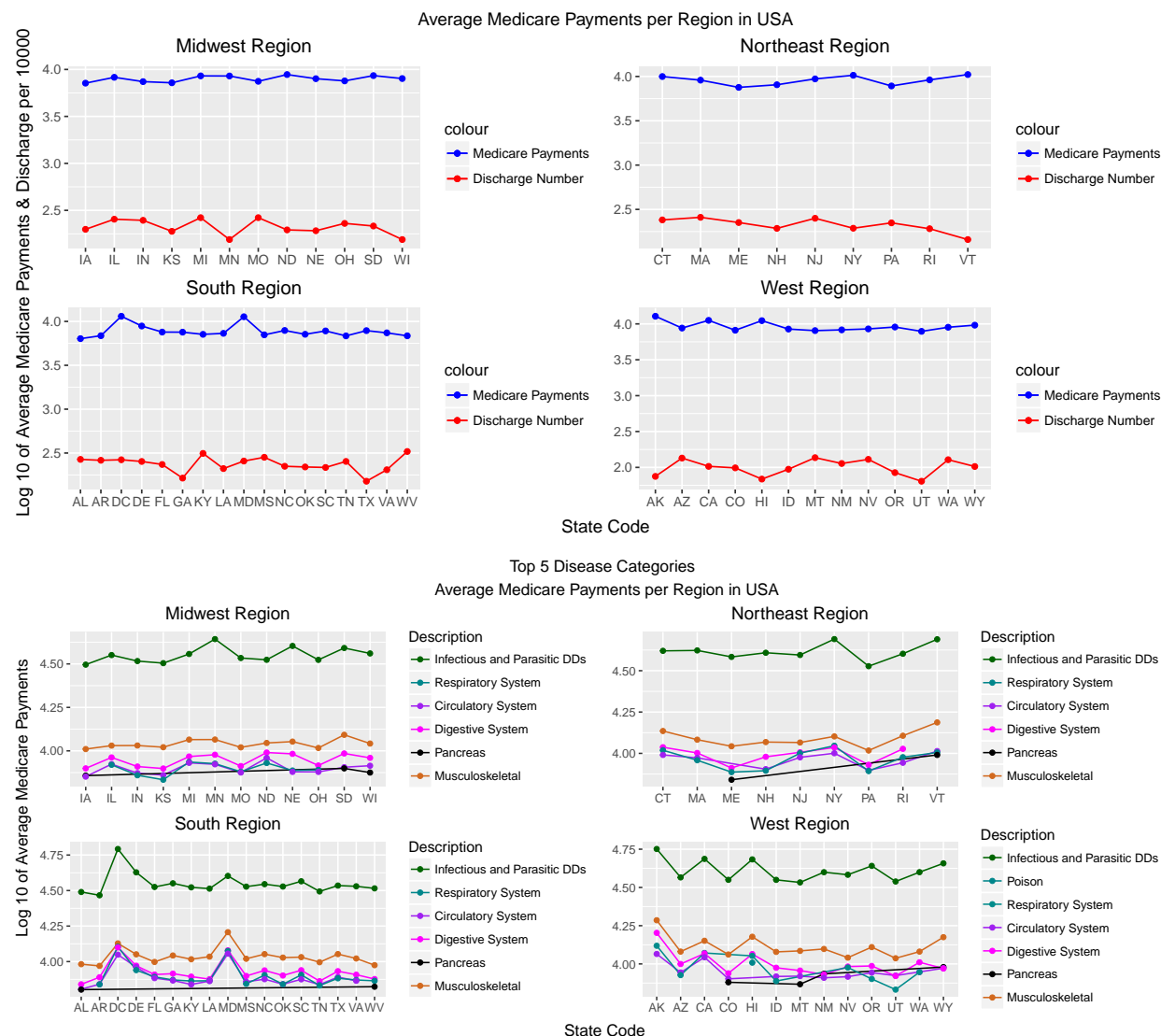
From the above graph, we have the below observations.

Among all the regions: - West region has higher Average Medicare Payments. This is pretty obvious as the cost of living along the West Coast if higher. It has the smallest Discharges per 10000. - South and Northeast have the highest Discharges per 10000, with South being on top for most of the diseases.

Across all the diseases: - ENT has lowest Average Medicare Payments and Discharges per 10000. - Circulatory, Digestive, and Respiratory Systems have a higher Discharges per 10000.

As detailed in the previous analysis (Average Medicare Payments in different US regions), the United States of America has the highest percentage of people suffering with heart related diseases when compared to other disease categories. And so it seems reasonable that cardiovascular diseases and musculoskeletal systems have higher discharge rates.

Last but not the least, Infectious and Parasitic Diseases doesn't have equally overwhelming Discharges per 10000 although it has the highest Average Medicare Payments.

Average Medicare Payments by Discharge per 10000 & Top Five Diseases analyzed by each state

Average Medicare Payments per Region in USA

Top 5 Disease Categories
Average Medicare Payments per Region in USA

Now we analyze the average payments in each state of the regions in USA - its average medicare payments and discharge rate and compare it with its top five diseases.

West Region:

- Minnesota has the lowest discharge rate and highest average medicare payment. This can be attributed to the Infectious and Parasitic disease, as it receives the highest payment for Infectious and Parasitic disease. In 2016, Sexually transmitted diseases and vector-borne diseases were reported to be most common infectious diseases in Minnesota, along with other diseases such as foodborne diseases and antibiotic-resistant organisms and rare diseases such as severe acute respiratory syndrome (SARS), Ebola virus disease, and Zika virus disease.

- Wisconsin, South Dakota and Ohio, even though have a downward slope for discharge rate, they have higher average medicare payments for infectious, musculoskeletal, digestive and circulatory systems disease compared to other states. The average medicare payments of these states tend to mimic the trend of top five disease.

Northeast Region:

- New Jersey has the highest discharge rate but does not have the highest average medicare payment compared to other states. Whereas New York has a lower discharge rate but has the highest average Medicare payments compared to the other states. Again, this trend tends to follow the top five diseases in that state. As New York tends to pay out more on infectious, musculoskeletal, digestive and circulatory systems diseases its state average is the highest and New Jersey comparatively spends lower, its state average medicare payment is also lower.

- Pennsylvania, Rhode Island and Vermont, even though have a downward slope for discharge rate, they have a higher slope for average Medicare payments for infectious disease, musculoskeletal, digestive and circulatory systems compared to other states. The average Medicare payments of these states tend to mimic the trend of top five diseases. Rhode Island and Vermont are also states having the highest cost of living in the region, they tend to pay more for hospital charges.

South Region:

- Georgia and Texas have the lowest discharge rate as compared to the other states but they have comparable average Medicare payments. This also follows the trend of average Medicare payments of the top five diseases.

- Maryland and District of Columbia, even though have a lower discharge rate, have the highest average Medicare payments. They also have the highest average Medicare payments for the top five diseases as comapred to other states. Their cost of living is also the highest in Southern region.
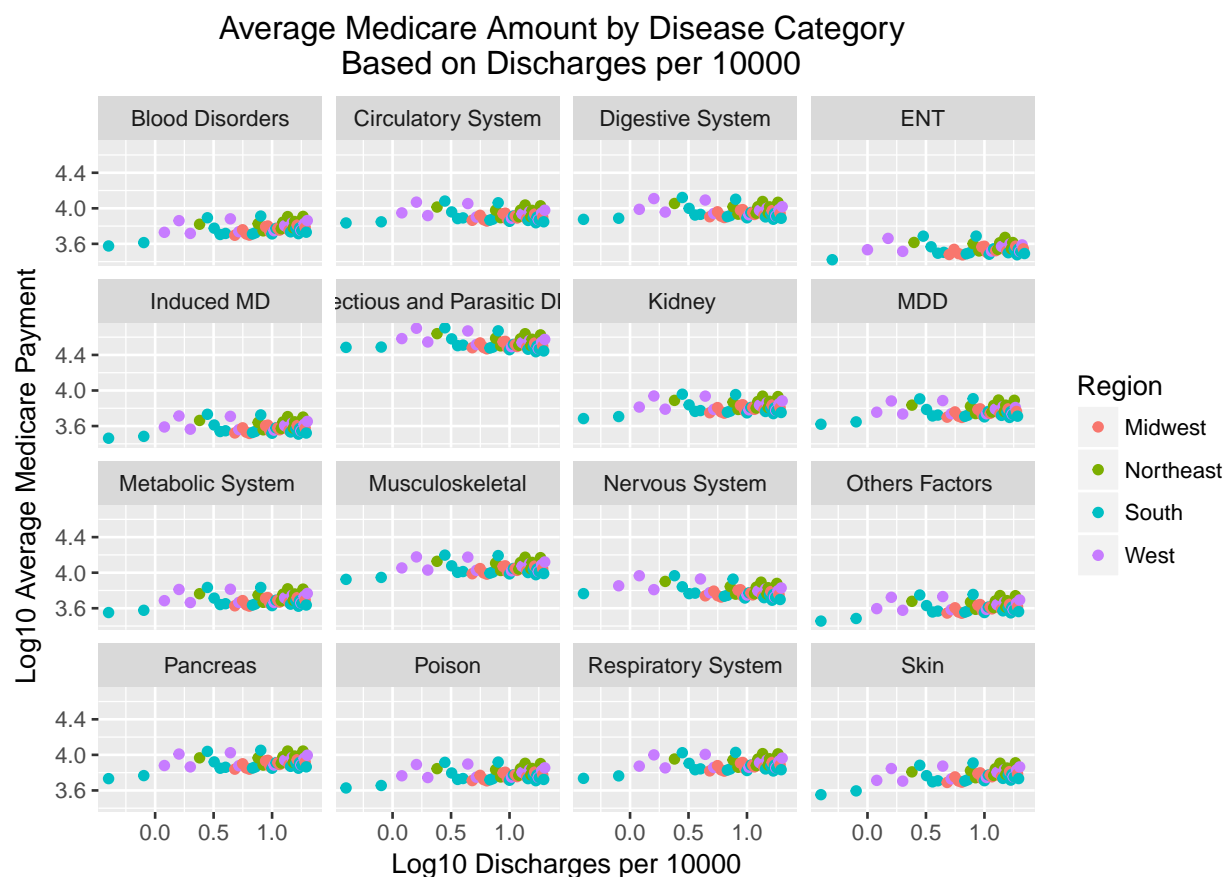
West Region:

- Hawaii and Arkansas have lower discharge rate but highest average Medicare payments. They have the highest average medicare payments for the top five diseases as compared to other states.

- Utah has one of the lowest discharge rate and lower average Medicare payments for the top five diseases. Thus, it has the lowest average medicare payments as compared to other states.

And so, we find that the average Medicare payments is more dependent on the disease category and the state but lesser dependent on the discharge rate.

Predictive Analysis:

Since the average Medicare payment is a continous variable, we fit a linear model. Based on the AIC, we have used discharge rate, State and the disease category to predict average Medicare payment. We also decided to include an interaction between discharge rate and the disease category. We have transformed the average medicare payment and discharge rate to a log (base 10) scale to increase their variability for fitting the model and prediction.

Our model has captured 98.82719 % of the variation in the Average Medicare Payments.

Average Medicare Amount by Disease Category
Based on Discharges per 10000

Northeast and Midwest are on the higher end of the discharge rate as compared to the other regions. West and South has more number of states having more discharges with few states having lower dsicharge rate. Infectious and parasitic diseases have the highest average Medicare payments. This is followed by musculoskeletal, digestive, circulatory diseases.

Conclusion:

Initially we had hypothesized that average Medicare payments are more dependent on discharge rate as mentioned by the CMS. But analyzing the graphs we find that they are more dependent on the disease category in each state. Apart from this, the presence of other confounding variables like cost of living, most commonly occuring diseases, climate of the state, living conditions, etc also affect the average medicare payments.

For future analysis, we can also include referral description for our analysis. This explanatory variable describes which hospitals have lesser medical facilities available for a category of disease and may refer the patient to other multispeciality hospitals within or outside that state. This can also affect the discharge rate and average Medicare payment of that state. The average Medicare payment may also be more for a particular city and may sway the state average to higher values. Performing a city wise analysis will also aid in predicting the average Medicare payments. The type of hospitals (eg. multispeciality having advanced equipments) and type of patients treated (eg. more patients treated who are below poverty line) may help predict a hospitals average Medicare payment.

Appendix:

Fitted vs Residuals