An Analysis of CMS Average Payments with a Focus on Internal Medicine

Marshall Tuck

12/14/2019

Overview

Medicare.data.rdata is an R data file consisting of three data frames: 'claims', 'hcpcs', and 'providers'. All are publicly available through the Centers for Medicare and Medicaid Services. This paper will provide an analysis of these three data tables in the following sections titled Data Overview, Exploratory Data Analysis and Research Questions, Statistical Analysis, and Conclusions and Future Work.

In the Data Overview Section, I will explore each data frame provided in the .RData file.

In Exploratory Data Analysis and Research Questions, I will conduct exploratory analysis and share summary data regarding each data frame provided in the.RData file. I will also explore the following relationships: the effect of 'entity.code' and 'provider.type' on 'avg.chrg.amt', 'allowed', and 'avg.payment'

In Statistical Analysis, I will conduct statical analysis on the impact of provider type and entity code on average charge amount. I will develop an SLR model for medicare payments 'avg.payment' using 'entity.code' and 'prov.type'. We will also explore how mean payments for a common prov.type have progressed from 2012-2016, and then make predictions for 2017. I will then compare the predicted value to the actual value.

In Section 3 and 4 I will also research the following question: which medical licensure (for example, M.D, D.O., PT, etc.) provides the highest financial reward based on the Medicare data?

Section 1: Data Overview

Claims

Claims data presents nearly 57 million claims for Medicare reimbursement for the following key variables: year, NPI (a provider identifier), zip code of provider, state of provider, HCPCS service code, service count, average allowed amount, average charge amount, and average payment.

HCPCS

HCPCS presents 7527 hcpcs codes, which are unique treatment identifiers, and the description of the treatment within that code.

Providers

Providers presents the >1.2 million providers eligible to submit Medicare claims, including their NPI number (a unique provider identifier), the licensing credentials, gender of the provider, entity code, and provider type/specialty.

Approach to Data

A sample data population was created that is based on a sample without replacement of 10,000 claims from the claims data set. This was then joined with hcpcs data set and providers data set to get a full picture of the sampled data. All references to data, summary statistics, and prediction in this paper is based on the sampled data of 10,000 claims records.

Problems in manipulating the data were due to the .RData size and the need to clean the data. ~57 million claims would not quickly or easily merge with the provider data set or hcpcs data set. In addition, credential data was not consistent; for example, all combinations of "M.D.", "M.D", and "MD" were coerced to "MD". This pattern was repeated for "DO". Due to outside research "PA-C" and "PA" are equivalent titles and were coerced to "PA". Like data munging occurred for other credentials.

Section 2: Exploratory Data Analysis and Research Questions

In our first data summary below, we see no clear concentration of NPI frequency in our data. The top frequency NPI number only occurrs three times in our sample of 10,000 claims. Many NPI numbers occur only once in 10,000 claims.

NPI	Percentage_NPI	count
1114906245	0.03	3
1205059763	0.03	3
1346285988	0.03	3

NPI	Percentage_NPI	count
1003000449	1e-04	1
1003012063	1e-04	1
1003014705	1e-04	1

Below in Figure 1 is time-series analysis showing the Total Reimbursement per year for I and O entity.codes. In Figure 2 is time-series analysis showing the top 5 prov.types in historical reimbursement money per year.

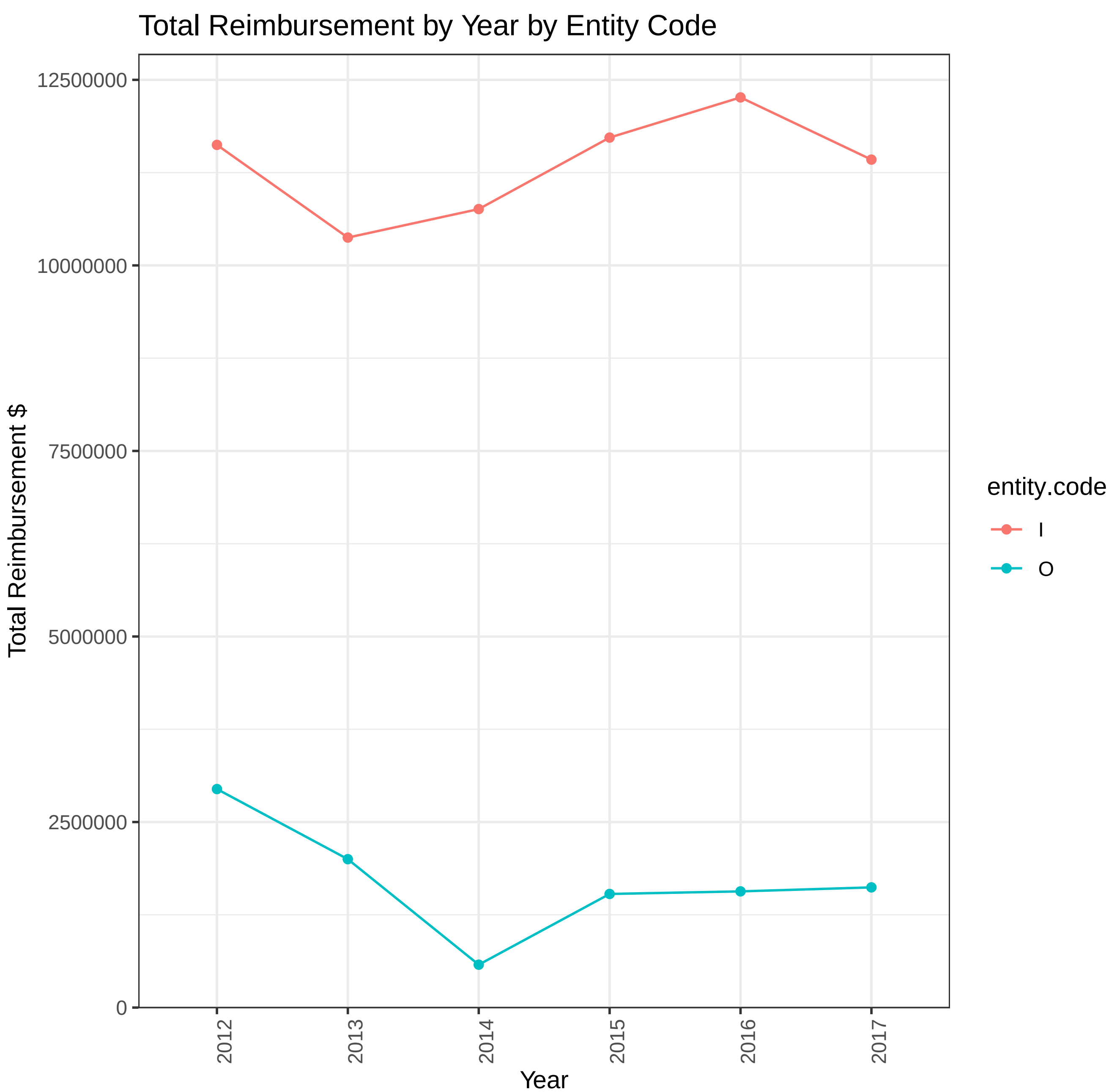


Figure 1

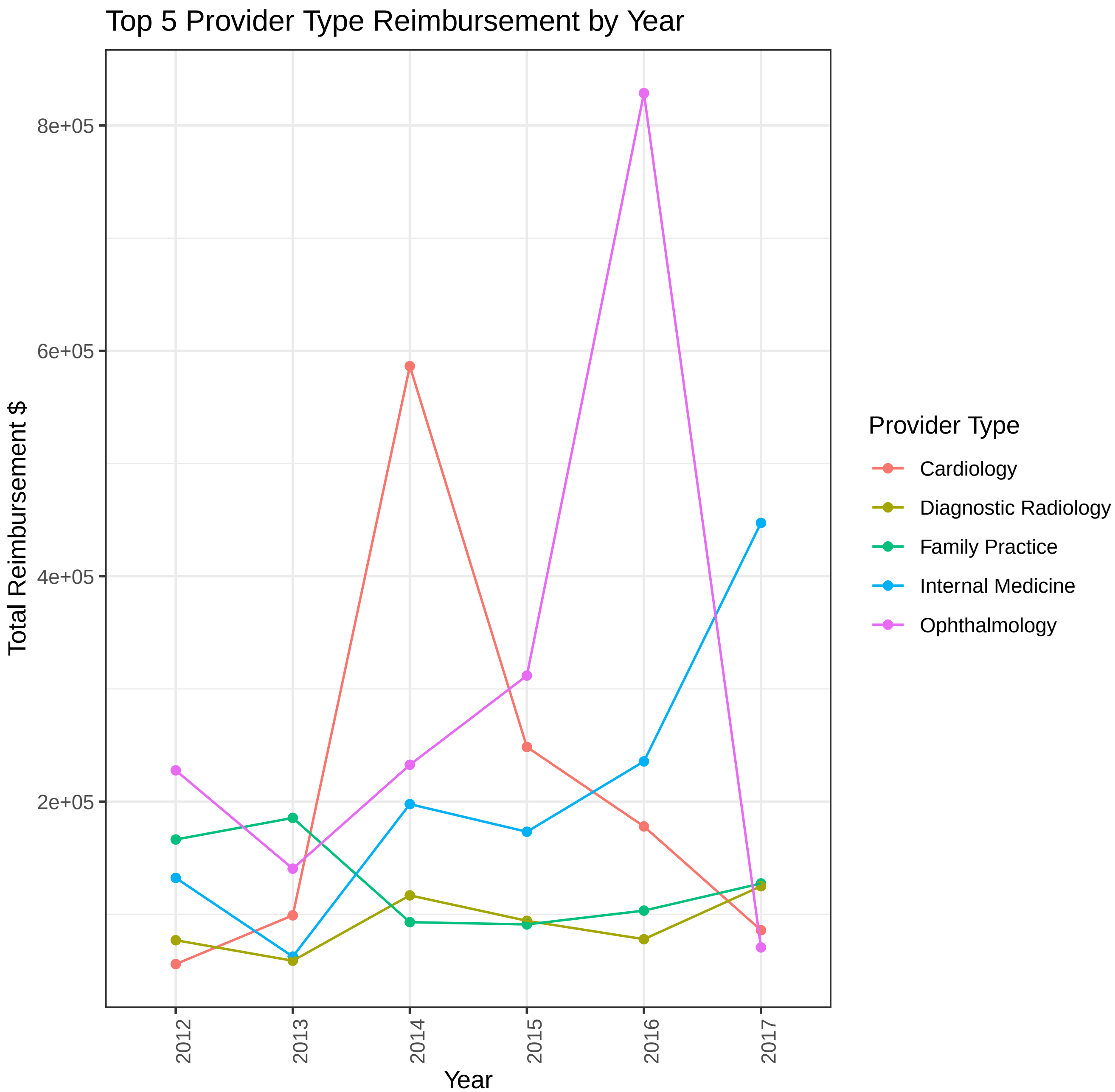


Figure 2

Of the most frequent treatments performed for Medicare reimbursement, code 99213 (15 minute office visit) was performed the most with a mean payment of $46.72, and the most expensive procedure occurring one time was code J7192 (Factor viii - antihemophilic factor) with a mean payment of $16474.

hcpcs.code	hcpcs.description	freq_hcpcs	Mean_Payment
99213	Established patient office or other outpatient visit, typically 15 minutes	425	46.72110
99214	Established patient office or other outpatient, visit typically 25 minutes	418	68.58015
99204	New patient office or other outpatient visit, typically 45 minutes	198	112.62190

hcpcs.code	hcpcs.description	avg.payment	freq_hcpcs	Mean_Payment
J7192	Factor viii (antihemophilic factor, recombinant) per i.u., not otherwise specified	16473.720	1	16473.720
37238	Insertion of intravascular stents in vein, open or accessed through the skin, with radiological supervision and interpretation	3175.375	1	3175.375
J2505	Injection, pegfilgrastim, 6 mg	3094.557	1	3094.557

Further reviewing summary statistics, we see that MDs receive the highest total payment from Medicare ($53543291), factoring in all types of procedures. M.B.B.S receive the lowest average payment from Medicare ($18.15) factoring in all procedures performed.

credentials	total_payment_amount
MD	53543291
NA	11476350
DO	3780607

credentials	total_payment_amount
MD, DO	74.33
M.D., PHD, FACC	63.48
M.B.B.S.	18.15

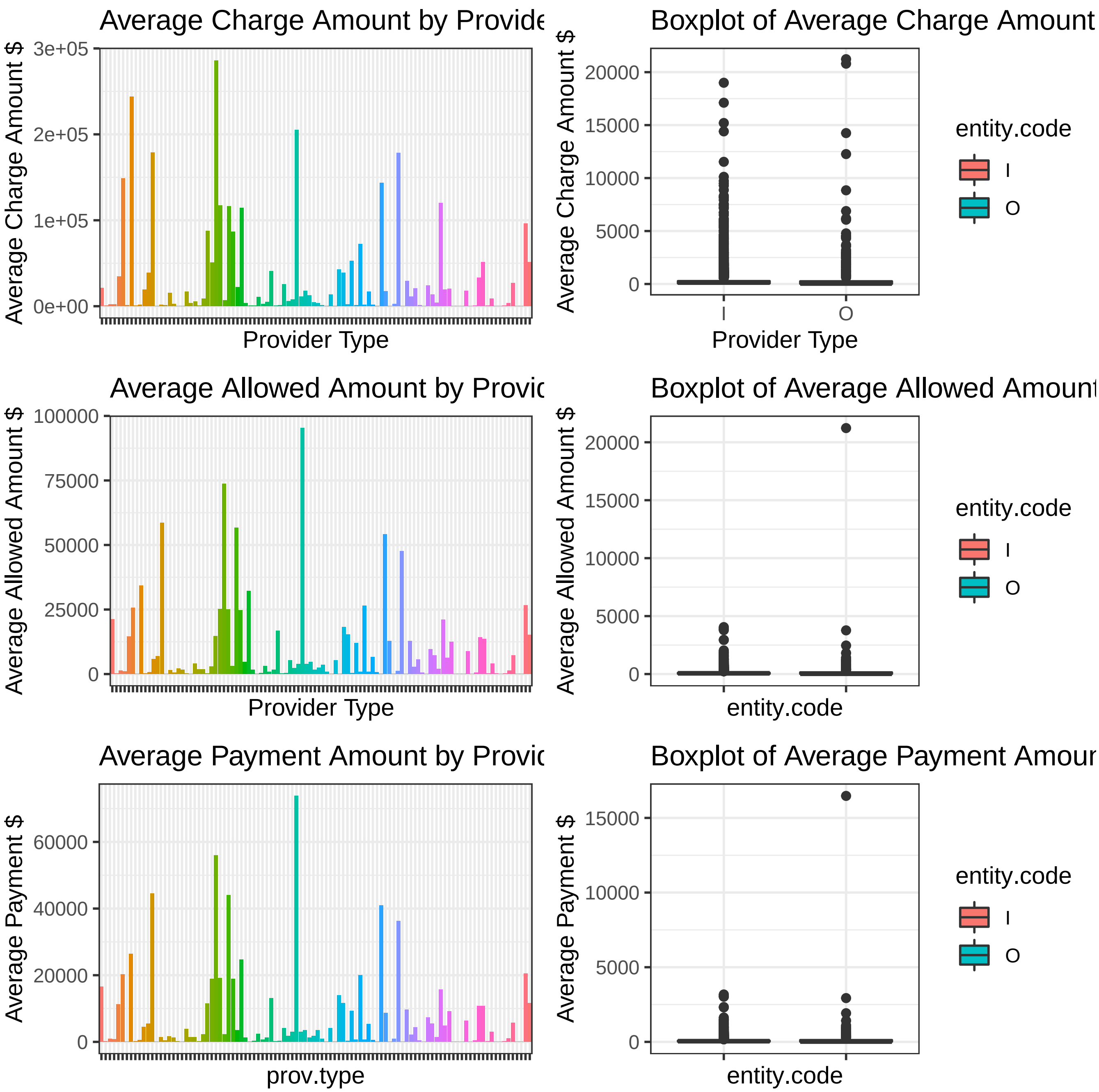
Below shows summary statistics of the top ten greatest counts of credentials, provider type, and entity code in our sample. The most frequent credential is MD, the most frequent prov.type is Diagnostic Radiology followed by Internal Medicine, and the most frequent entity.code is I.

credentials	credential_count
MD	7005
NA	656
DO	621
PA	247
DPM	110
O.D.	106
PT	105
CRNA	96
NP	87
D.P.M.	50

prov.type	prov_count
Diagnostic Radiology	1285
Internal Medicine	1247
Family Practice	1008
Cardiology	569
Nurse Practitioner	464
Physician Assistant	322
Orthopedic Surgery	313
Ophthalmology	260
Anesthesiology	255
Emergency Medicine	247

entity.code	entity_count
I	9580
O	420

The grid arrange below shows the effect of provider type and entity code on 1 of 3 response variables: Average Charge Amount, Average Allowed Amount, and Average Payment. On the left shows a concentration of charges, allowed amounts, and payment amount among a few providers by type. On the right, we see a narrow box-and-whiskers plot, with high outliers for only a few Entity Codes of type “Entity”.



Further, a research question I posed in the Overview asks which is the best credential to pursue. Based on the table below, MD appears to be the most financially rewarding choice - MD have the highest total reimbursement of any credential.

credentials	Total_Reimbursement
MD	53543291
NA	11476350
DO	3780607

However, this does not factor in the cost of performing the services (which we can assume is related to the Amount Charged Value). In finding the difference between the payment received from medicare and the charged amount from the provider, we see that all credentials have higher amounts charged than reimbursed. The credential that minimizes this loss is combined MSN, FNP. Below, the data is represented as

$$(avg.payment * countservices) - (avg.charge * countservices)$$

Charged amount will always be equal to or greater than payment amount. So, below, the negative value closest to zero is the value that minimizes the “losses” to provide the services.

credentials	Total_Profit_Loss
PA-C, PH.D.	-12.48
RN, PHN, NP-C, MSN	-11.30
MSN,FNP	-8.46

Section 3: Statistical Analysis of Internal Medicine Provider Type

In order to create an SLR model relationship, I first reduced the prov.types in my data into two values: "Internal Medicine" and "Not Internal Medicine". This would allow us to clearly state our coefficients and predict values for only Internal Medicine in a focused way.

In an explanatory model of relating avg.payment to entity.code and prov.type, we find an ANOVA analysis calculating a significant P Value for prov.type (.005383). and a signficant P Valye for entity.code (1.358e-14). We conclude that there is sufficient sample evidence to show prov.type and entity.code have a significant impact on avg.payment.

```
Analysis of Variance Table

Response: avg.payment
              Df    Sum Sq Mean Sq F value    Pr(>F)
prov.type      1    345377  345377  7.7493  0.005383 **
entity.code    1   2650649 2650649 59.4735 1.358e-14 ***
Residuals   9997 445552132   44569
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Formulating a model to show the effect of prov.type and entity.code on avg.payment, we look to reseach the below relationship:

$$avg.payment \sim prov.type + entity.code$$

In running summary data for this linear fit, in relation to Entity Code and provide type, we find the following coefficients:

```
Call:
lm(formula = avg.payment ~ prov.type + entity.code, data = fullsample)

Residuals:
    Min      1Q  Median      3Q     Max
 -154.4   -53.9   -27.1    12.0 16319.2

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                     59.183      5.978   9.900  < 2e-16 ***
prov.typeNot Internal Medicine  13.882      6.410   2.166   0.0304 *
entity.codeO                    81.420     10.558   7.712 1.36e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 211.1 on 9997 degrees of freedom
Multiple R-squared:  0.006679,  Adjusted R-squared:  0.006481
F-statistic: 33.61 on 2 and 9997 DF,  p-value: 2.829e-15
```

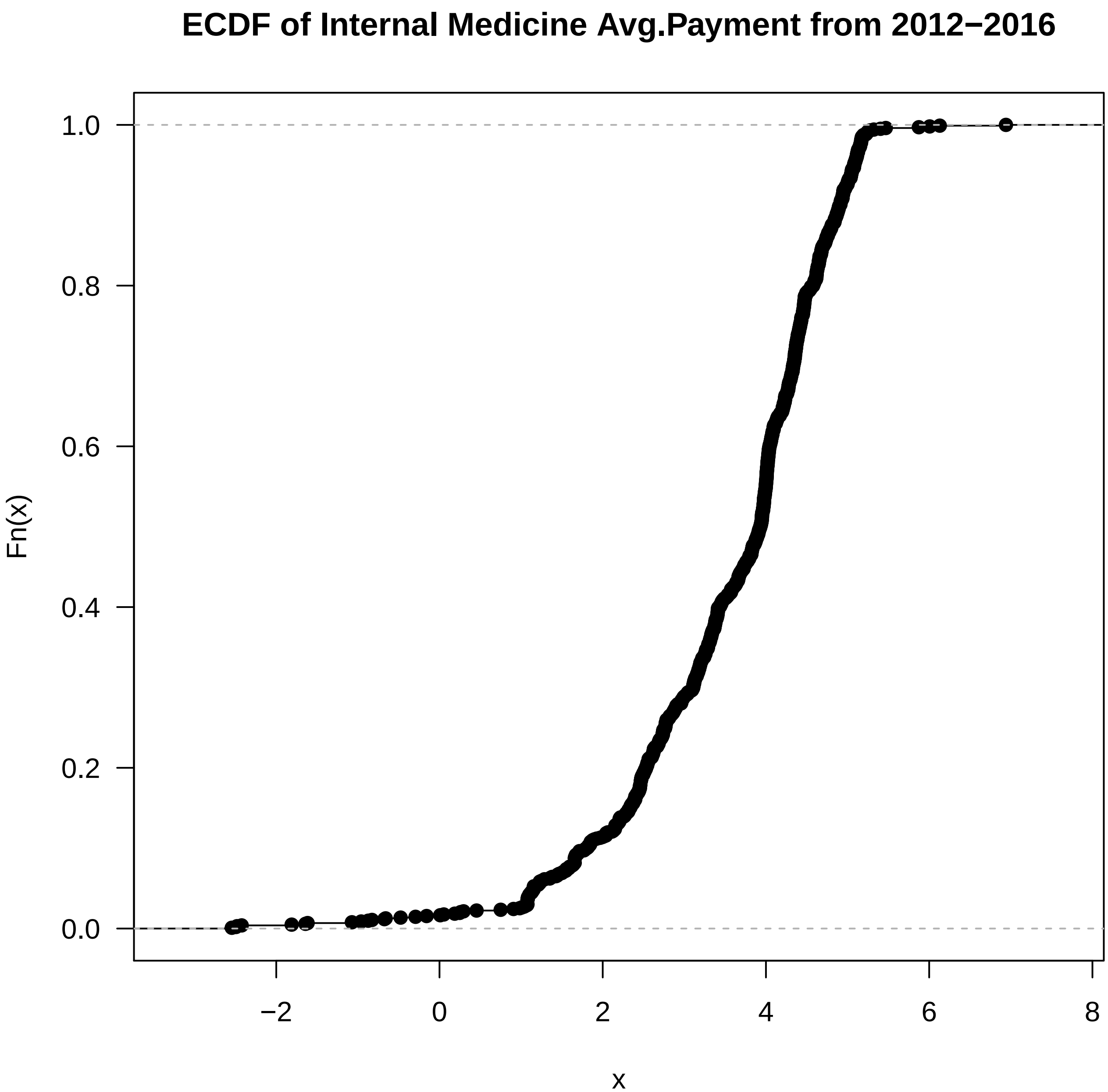
With concern to prov.types of Internal Medicine, we can express the relationship in the following way:

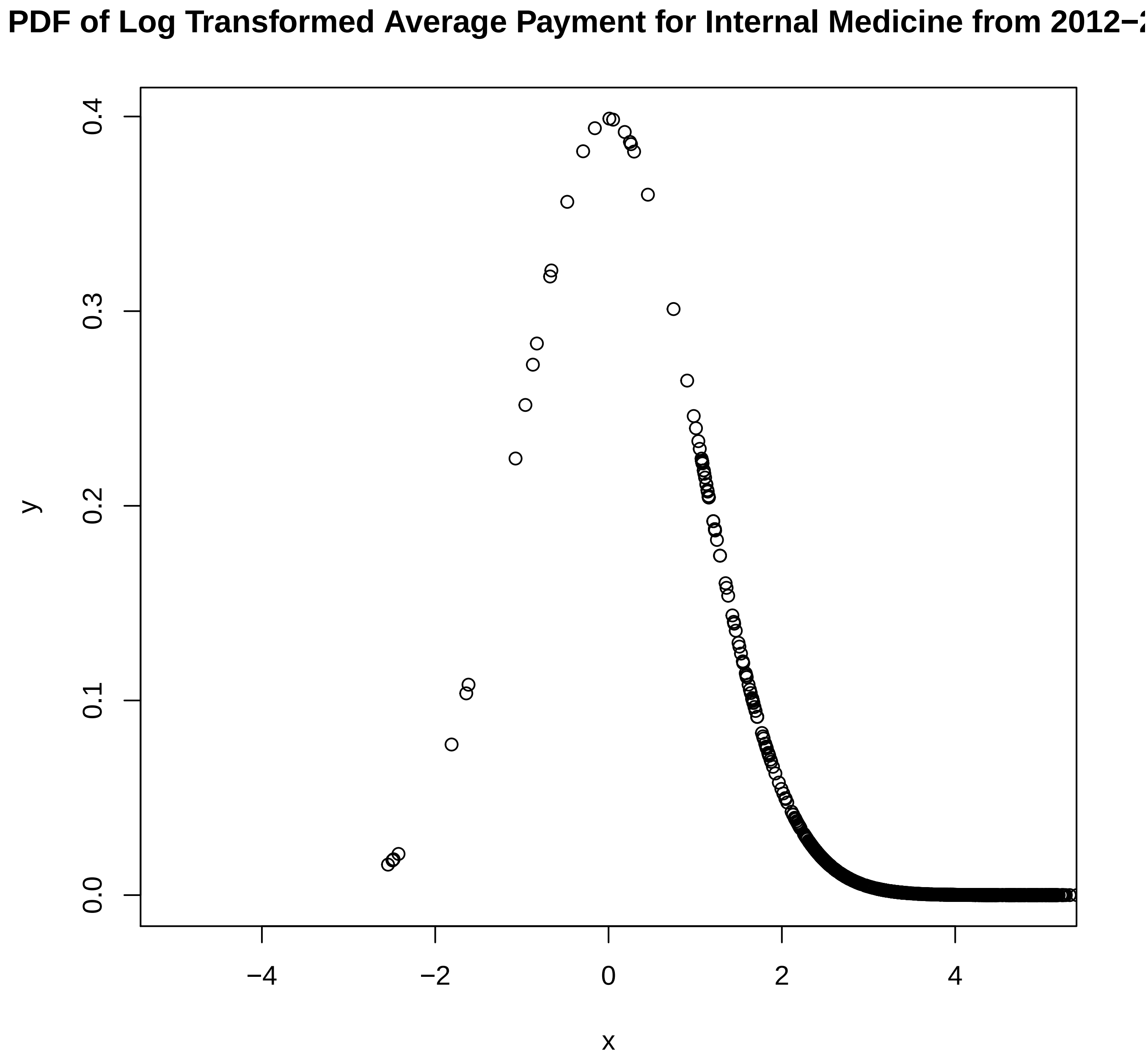
$$avg.payment = 59.183 + 81.42 * entity.code : O + 13.882 * prov.type : NotInternalMedicine$$

We find an R-squard value of .007, meaning that .7% of avg.payment can be attributed to our model relationship which is not a high correlation. There are other values, for instance other prov. types that were coerced to "Not Internal Medicine", that better relate to our avg.price.

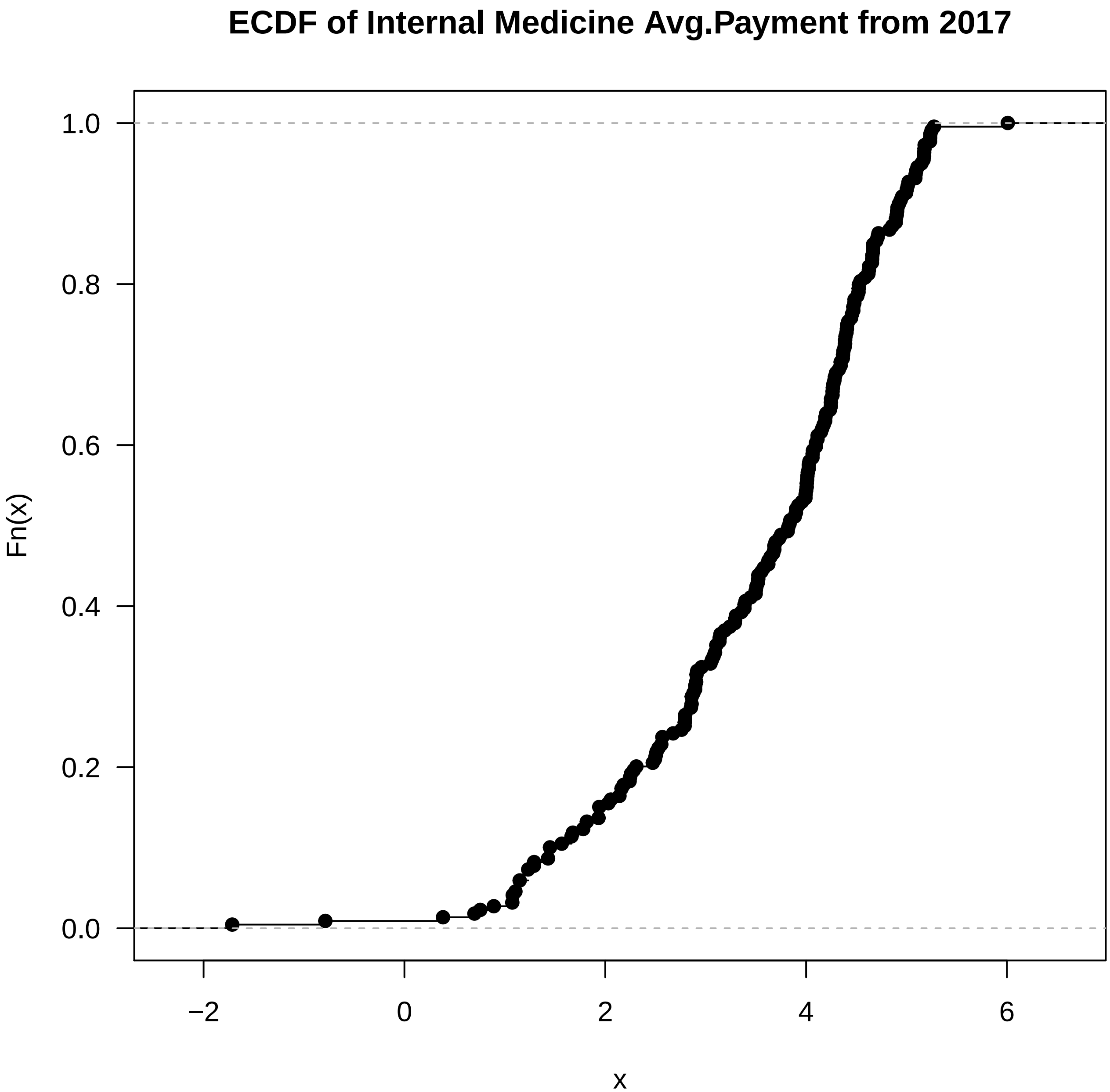
```
[1] 0.006679384
```

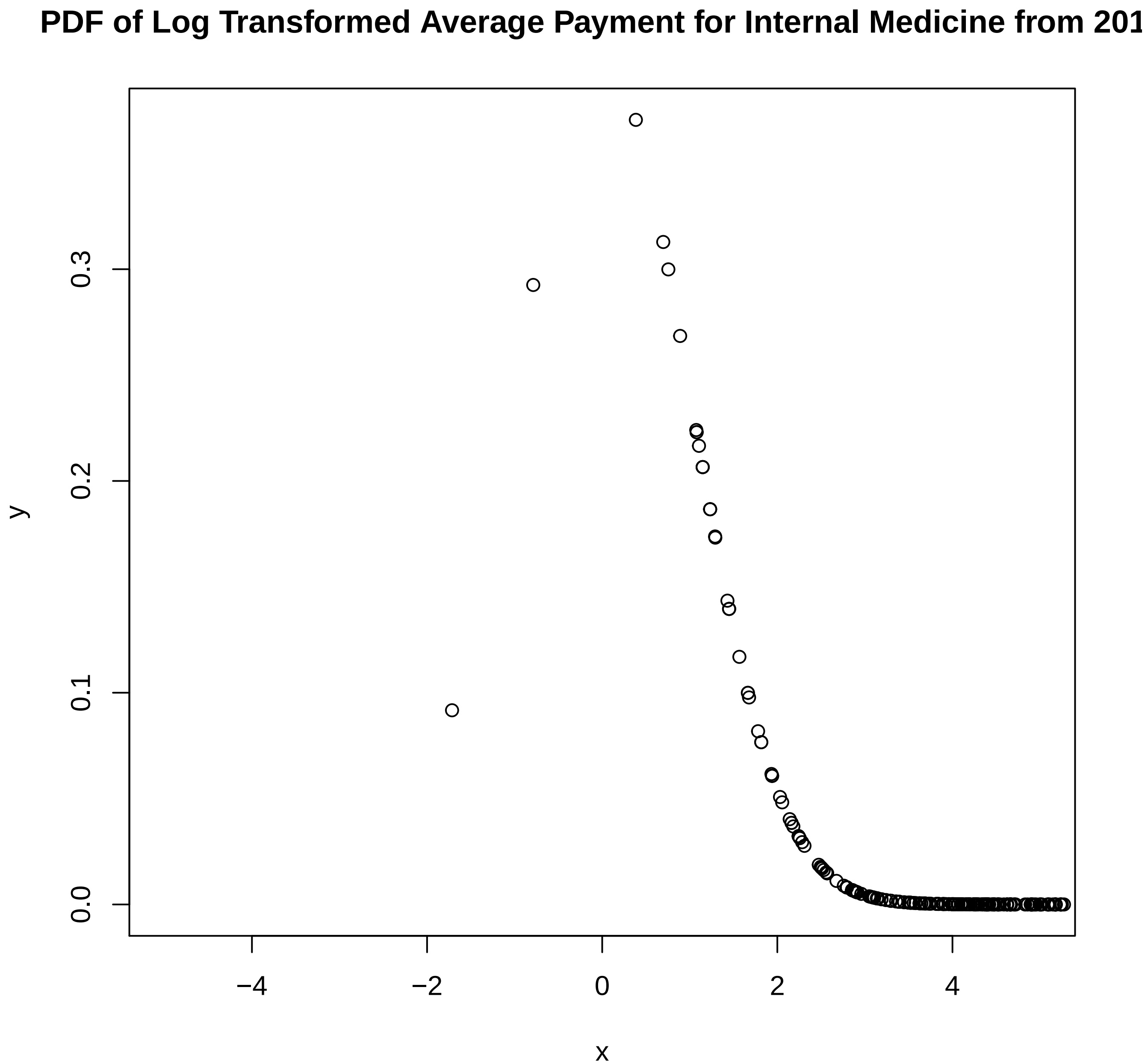
Below is the ECDF and PDF of the Internal Medicine prov.type from 2012-2016 and from 2017. Both distirbutions appears centered around zero and symmetric, following a normal distribution around each mean. A long tail exists to the outer end showing extreme avg.payments for some procedures. A log distribution was applied to scale down the extreme values. We proceed with a prediction method.





PDF of Log Transformed Average Payment for Internal Medicine from 2012−2
y
0.0
0.1
0.2
0.3
0.4
x
−4
−2
0
2
4


ECDF of Internal Medicine Avg.Payment from 2017
Fn(x)
1.0
0.8
0.6
0.4
0.2
0.0
−2
0
2
4
6
x



Prediction of Internal Medicine Avg. Payment

Based on our linear model above, we would expect 2017 average payments to Internal Medicine providers with entity.code = "I" to equal

$$avg.payment = 59.183 + 81.42 * 0 + 13.882 * 0 = 59.183$$

From Figure 4 below we see the expected value of Internal Medicine average payments in 2017 is \$59.1833, and the actual value is \$58.8.

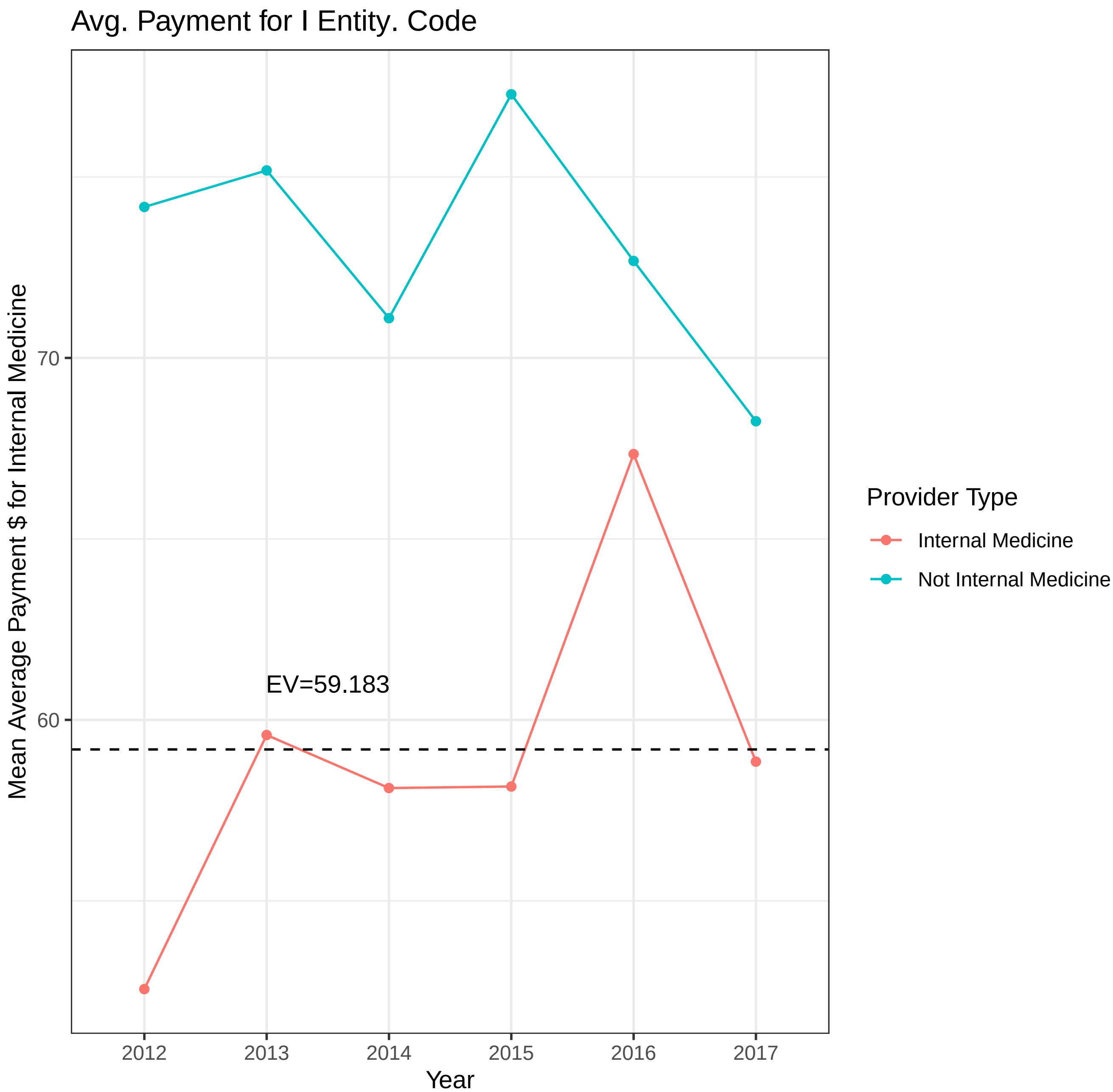


Figure 4

Conclusions and Future Work

In the above analysis, summary statistics were reviewed and discussed, the preferred credential type was discussed, and a model was fit to the data. We find that total Medicare reimbursement is relatively stable, however there exist fluctuations within the prov.types over the years. We estbalish that the MD credential receives the highest reimbursement from Medicare, however "MSN, FNP" credential minimizes the amount "lost" based on Medicare reimbursement - Amount Charged.

We then went to conduct an SLR on entity.code and prov.type, where both variables were binary (Entity.Code = O or I, prov.type = Internal Medicine or Not Internal Medicine). We found that the expected value in 2017 was less than 1 difference from the actual mean from 2017.

In the future, more analysis would be done on the cost of providing the care (as it is not clear what is reimbursed as profit or what portion is being reimbursed to "cover losses").

Also included in future research is a full coefficient analysis of all prov.types, going further than a basic compare of "Internal Medicine" or "Not Internal Medicine" and entity.code. It would be interesting to see the expected value of average payment for any claim, given the variables within the claim (entity.code and prov.type) to understand fraudlent claims.