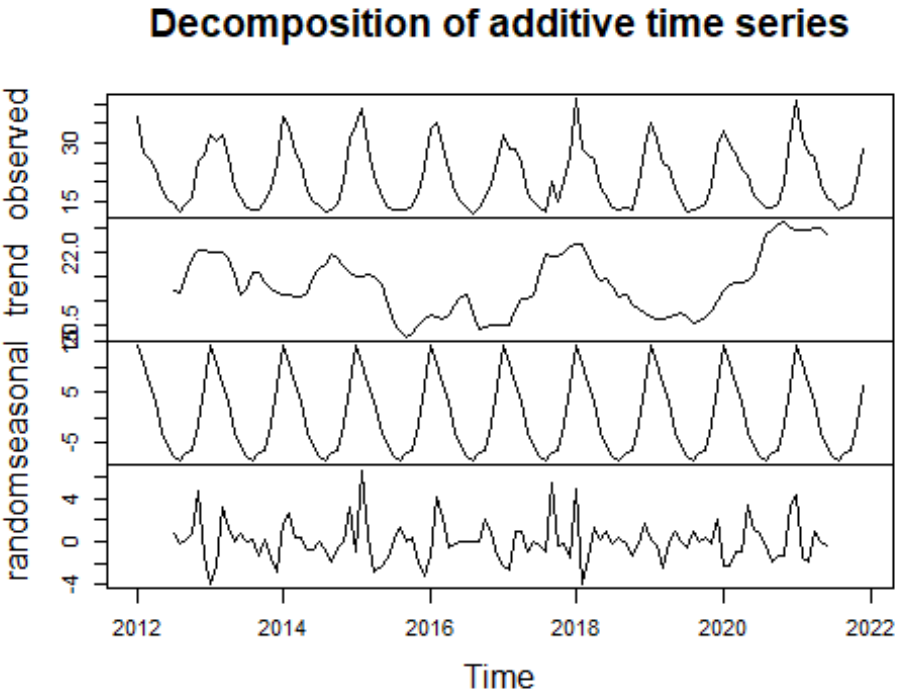
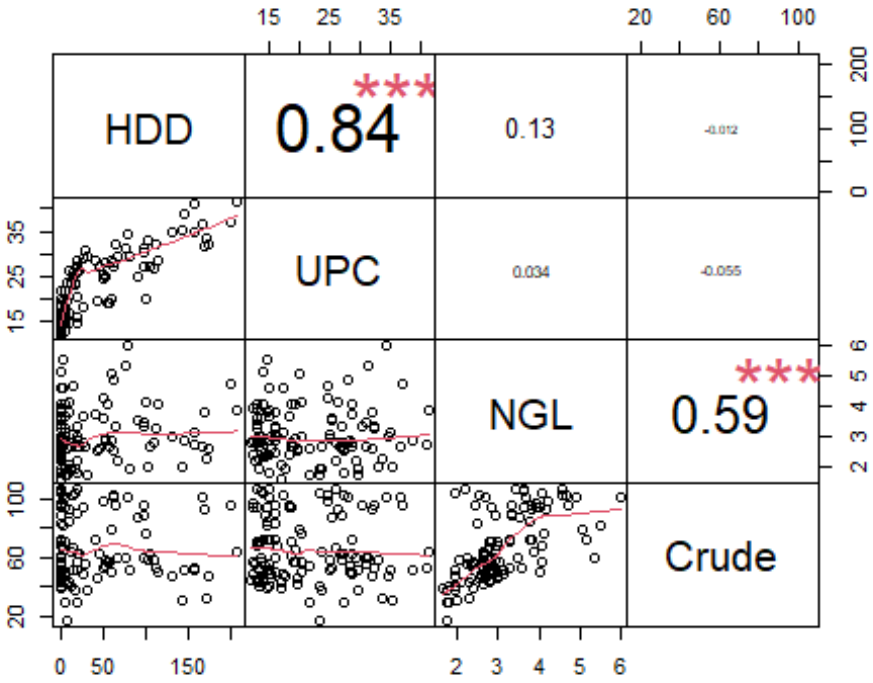
Residential Use per Customer Forecast

Overall Residential Use per Customer Forecast

In this section we will forecast Residential Use per Customer. Residential customers are filtered by Rate Class: 10,22,I0, and RS. We also include only customers that have greater than 0 reported volume usage for the month.

Customer Time-Series Decomposition



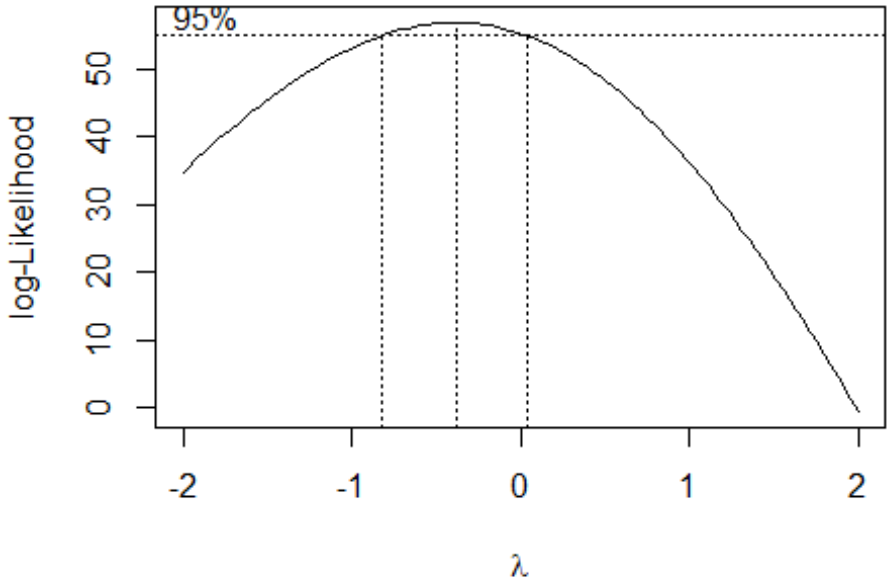


Multiple Linear Regression Model

In this section we will evaluate a Multiple Linear Regression Model using Month & Residential HDD to predict Residential Use per Customer.

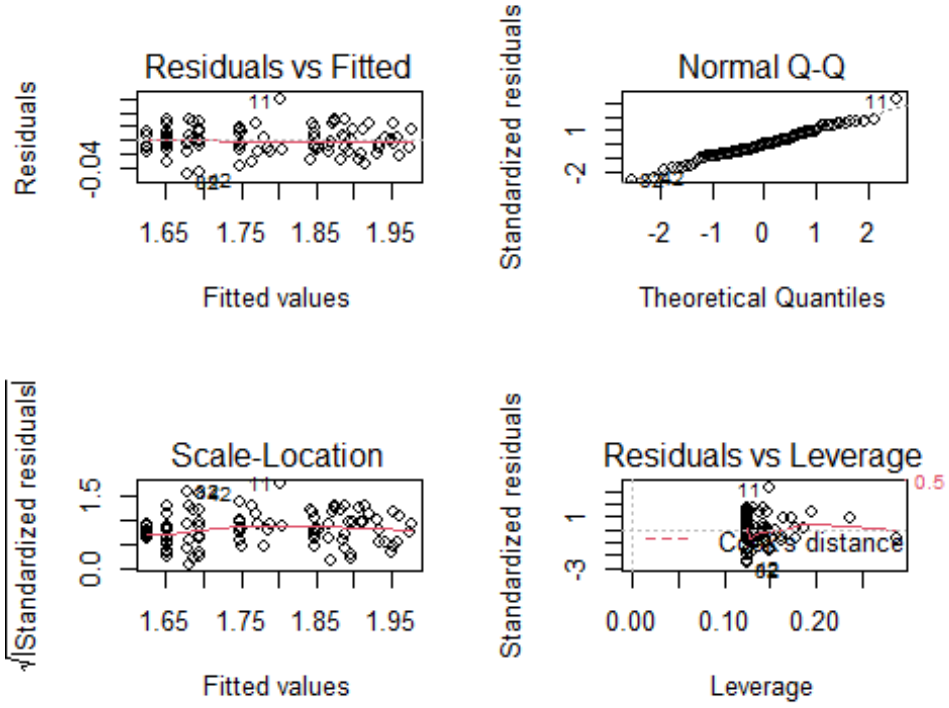
Diagnostics

Below we see that the model fits the data extremely well with an R-Squared value of .97. The model was created by removing one outlier and performing a Box-Cox transformation of -.3 to Use per Customer in order to fit the Linear Regression assumptions. In addition, we see that both regressors are significant in the model except for the month of February (month2).




```
[1] "Box-Cox Transformation Value (lambda): -0.38"


Call:
lm(formula = ((UPC^lambda - 1)/lambda) ~ HDD + month, data = train_data)

Coefficients:
(Intercept)          HDD       month2       month3       month4       month5
  1.8837678    0.0004452    0.0028588   -0.0238037   -0.0416953   -0.1364465
     month6       month7       month8       month9      month10      month11
 -0.1883721   -0.2329752   -0.2587697   -0.2322929   -0.2063773   -0.1221306
    month12
 -0.0387835
```

```
Call:
lm(formula = ((UPC^lambda - 1)/lambda) ~ HDD + month, data = train_data)

Residuals:
      Min        1Q    Median        3Q       Max
-0.046165 -0.011219 -0.000316  0.012652  0.057387

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.8837678  0.0134550 140.005  < 2e-16 ***
HDD          0.0004452  0.0000761   5.851 9.67e-08 ***
month2       0.0028588  0.0112215   0.255 0.799548
month3      -0.0238037  0.0117187  -2.031 0.045468 *
month4      -0.0416953  0.0144356  -2.888 0.004951 **
month5      -0.1364464  0.0149523  -9.125 4.05e-14 ***
month6      -0.1883721  0.0151144 -12.463  < 2e-16 ***
month7      -0.2329752  0.0151151 -15.413  < 2e-16 ***
month8      -0.2587697  0.0151151 -17.120  < 2e-16 ***
month9      -0.2322929  0.0153369 -15.146  < 2e-16 ***
month10     -0.2063773  0.0146370 -14.100  < 2e-16 ***
month11     -0.1221306  0.0124323  -9.824 1.67e-15 ***
month12     -0.0387835  0.0113377  -3.421 0.000976 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01948 on 82 degrees of freedom
Multiple R-squared:  0.9748, Adjusted R-squared:  0.9711
F-statistic: 264.3 on 12 and 82 DF,  p-value: < 2.2e-16
```



Back-Testing

In this section we will back test our Regression model to evaluate performance accuracy. We will trim our data from 2012 to December 2019 in order to predict the next 24 months and evaluate accuracy. To imitate a real-life scenario we will also use the 20 year average HDD as the HDD regressor values.

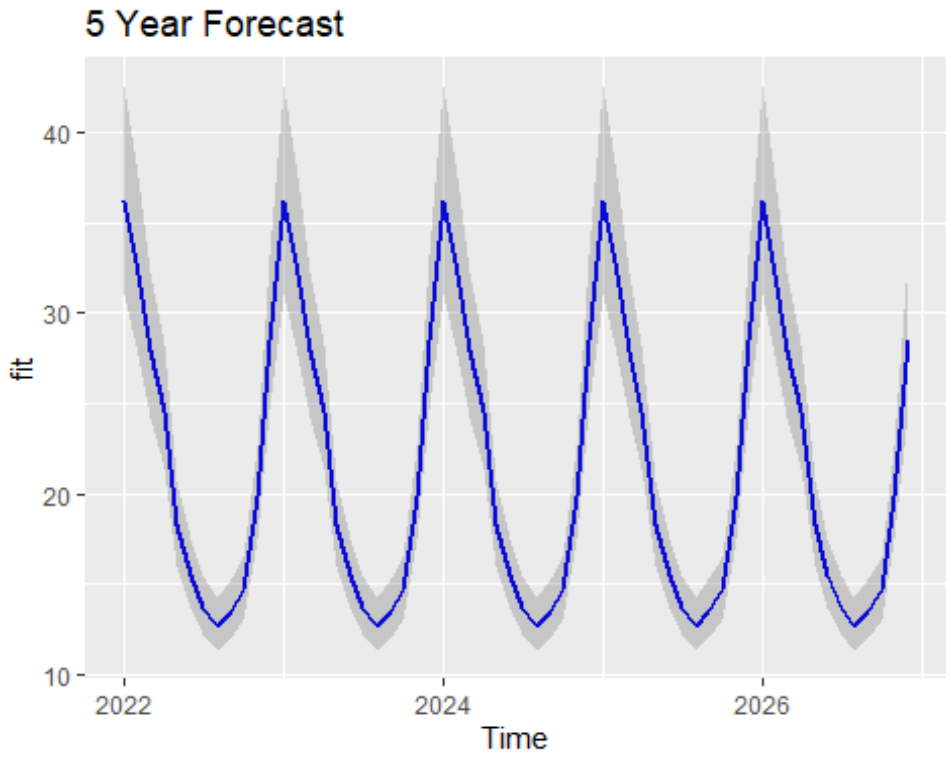
In the table below we see that regression model performs extremely well with an overall MAE of 1.47 and overall accuracy of 94% by predicting January 2020 to December 2021.

```
[1] 24 Month Mean Absolute Error (MAE): 1.41

[1] 24 Month Accuracy: 0.94
```

	Date	Actual.UPC	Predicted.UPC	Absolute.Error	Accuracy
97	Jan 2020	33.11	35.71	2.61	0.92
98	Feb 2020	29.59	32.41	2.82	0.90
99	Mar 2020	27.16	27.72	0.55	0.98
100	Apr 2020	23.43	24.42	0.99	0.96
101	May 2020	21.73	17.70	4.03	0.81
102	Jun 2020	16.79	15.18	1.61	0.90
103	Jul 2020	14.99	13.43	1.56	0.90
104	Aug 2020	13.36	12.54	0.82	0.94
105	Sep 2020	13.37	13.45	0.08	0.99
106	Oct 2020	14.49	14.60	0.11	0.99
107	Nov 2020	19.74	19.71	0.03	1.00
108	Dec 2020	32.00	28.18	3.82	0.88

109	Jan 2021	40.90	35.71	5.19	0.87
110	Feb 2021	31.38	32.41	1.03	0.97
111	Mar 2021	27.21	27.72	0.51	0.98
112	Apr 2021	26.47	24.42	2.05	0.92
113	May 2021	19.43	17.70	1.74	0.91
114	Jun 2021	16.01	15.18	0.83	0.95
115	Jul 2021	15.04	13.43	1.61	0.89
116	Aug 2021	13.00	12.54	0.46	0.96
117	Sep 2021	13.94	13.45	0.49	0.96
118	Oct 2021	14.63	14.60	0.03	1.00
119	Nov 2021	20.04	19.71	0.33	0.98
120	Dec 2021	28.68	28.18	0.50	0.98

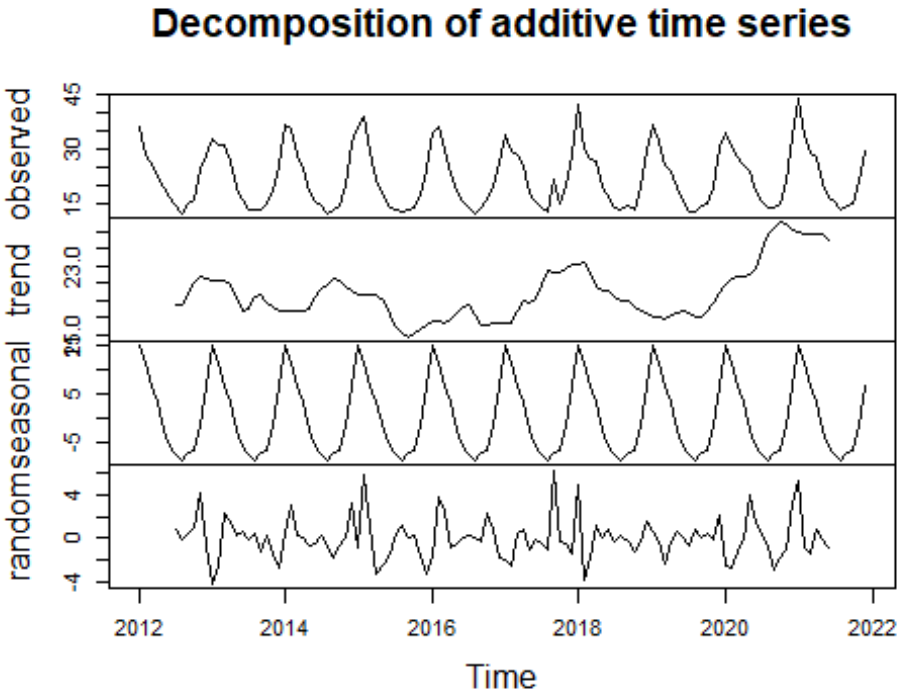
5 Year Forecast

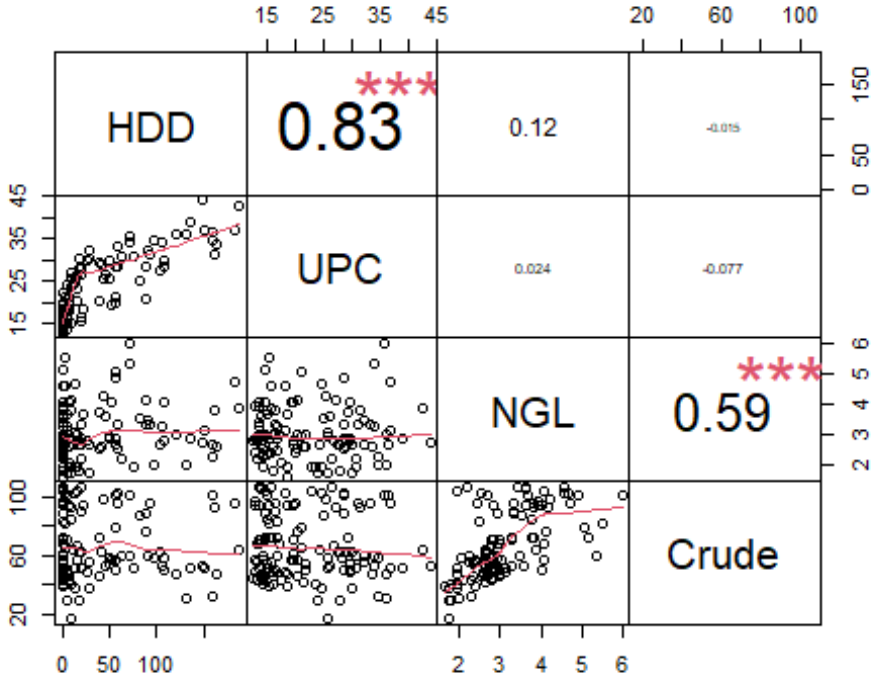


FPUC Residential Service (FPU-RS)

In this section we will forecast monthly client counts for FPU-RS. From the data given, these numbers are calculated by filtering for Tariff Schedule 'RS' and excluding Rate Class 22 which appears to be the Fort Meade residential clients.

Customer Time-Series Decomposition



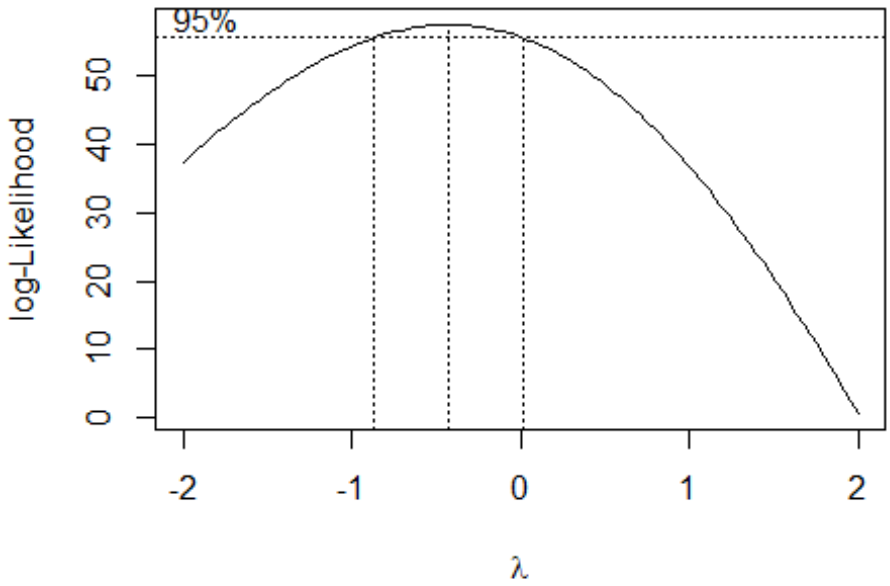


Multiple Linear Regression Model

In this section we will evaluate a Multiple Linear Regression Model using Month & Residential HDD to predict Residential Use per Customer.

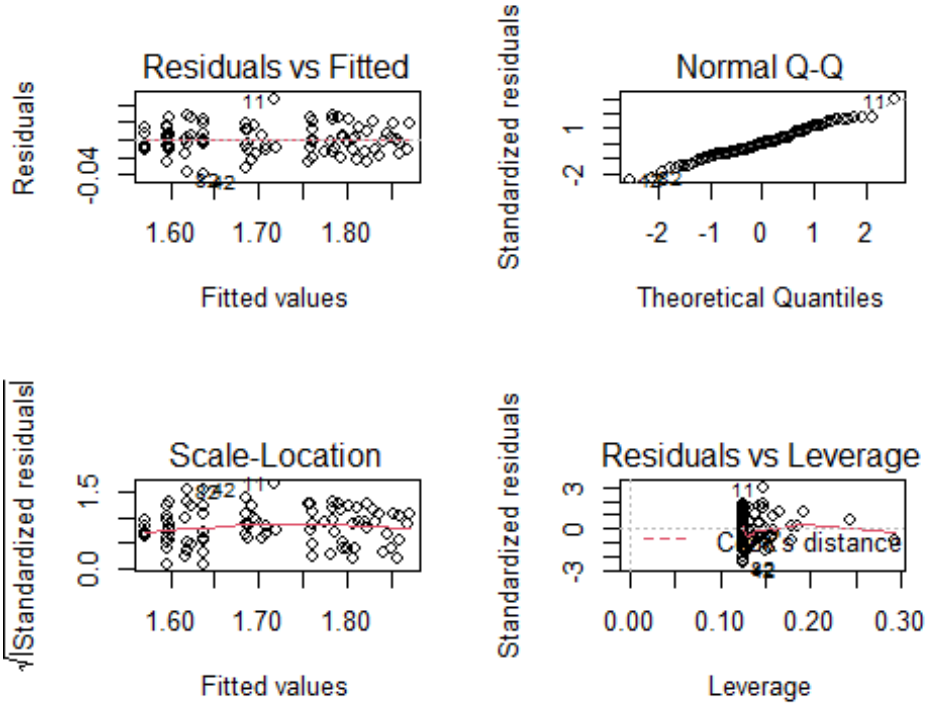
Diagnostics

Below we see that the model fits the data extremely well with an R-Squared value of .97. The model was created by removing one outlier and performing a Box-Cox transformation of -.3 to Use per Customer in order to fit the Linear Regression assumptions. In addition, we see that both regressors are significant in the model except for the month of February (month2).




```
[1] -0.42


Call:
lm(formula = ((UPC^lambda - 1)/lambda) ~ HDD + month, data = train_data)

Coefficients:
(Intercept)          HDD       month2       month3       month4       month5
  1.8031943    0.0003597    0.0001268   -0.0285917   -0.0468589   -0.1181726
     month6       month7       month8       month9      month10      month11
 -0.1655807   -0.2079561   -0.2320829   -0.2036373   -0.1866787   -0.1144724
    month12
 -0.0340199
```

```
Call:
lm(formula = ((UPC^lambda - 1)/lambda) ~ HDD + month, data = train_data)

Residuals:
      Min        1Q    Median        3Q       Max
-0.038172 -0.010642 -0.000098  0.012127  0.045216

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.803e+00  1.159e-02 155.543  < 2e-16 ***
HDD          3.597e-04  7.059e-05   5.096 2.19e-06 ***
month2       1.268e-04  9.693e-03   0.013 0.989598
month3      -2.859e-02  1.008e-02  -2.838 0.005728 **
month4      -4.686e-02  1.244e-02  -3.768 0.000309 ***
month5      -1.182e-01  1.289e-02  -9.168 3.34e-14 ***
month6      -1.656e-01  1.305e-02 -12.692  < 2e-16 ***
month7      -2.080e-01  1.305e-02 -15.940  < 2e-16 ***
month8      -2.321e-01  1.305e-02 -17.789  < 2e-16 ***
month9      -2.036e-01  1.324e-02 -15.380  < 2e-16 ***
month10     -1.867e-01  1.261e-02 -14.805  < 2e-16 ***
month11     -1.145e-01  1.080e-02 -10.601  < 2e-16 ***
month12     -3.402e-02  9.817e-03  -3.465 0.000845 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01693 on 82 degrees of freedom
Multiple R-squared:  0.974,  Adjusted R-squared:  0.9702
F-statistic: 256.3 on 12 and 82 DF,  p-value: < 2.2e-16
```



Back-Testing

In this section we will back test our Regression model to evaluate performance accuracy. We will trim our data from 2012 to December 2019 in order to predict the next 24 months and evaluate accuracy. To imitate a real-life scenario we will also use the 20 year average HDD as the HDD regressor values.

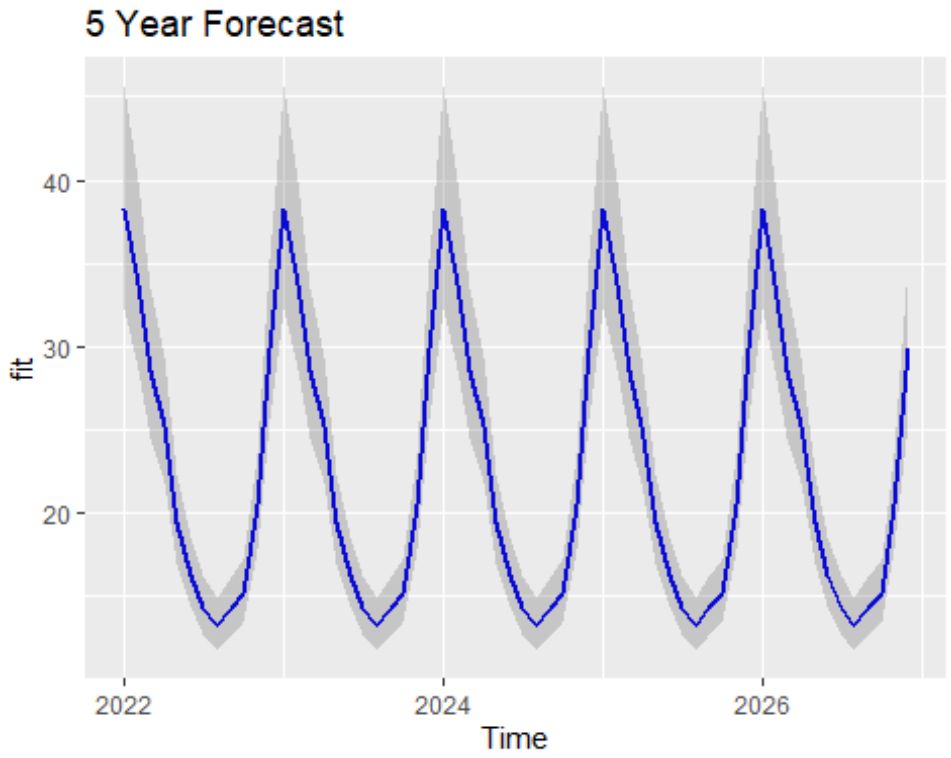
In the table below we see that regression model performs extremely well with an overall MAE of 1.47 and overall accuracy of 93% by predicting January 2020 to December 2021.

```
[1] "24 Month Mean Absolute Error (MAE): 1.72"

[1] "24 Month Accuracy: 0.93"
```

	Date	Actual.UPC	Predicted.UPC	Absolute.Error	Accuracy
97	Jan 2020	34.78	37.49	2.72	0.92
98	Feb 2020	30.91	33.77	2.86	0.91
99	Mar 2020	27.97	28.30	0.32	0.99
100	Apr 2020	25.65	24.82	0.83	0.97
101	May 2020	24.02	18.76	5.25	0.78
102	Jun 2020	18.42	15.99	2.44	0.87
103	Jul 2020	15.95	14.01	1.94	0.88
104	Aug 2020	14.31	13.04	1.27	0.91
105	Sep 2020	14.19	14.19	0.00	1.00
106	Oct 2020	15.36	15.12	0.24	0.98
107	Nov 2020	21.34	20.13	1.22	0.94
108	Dec 2020	33.79	29.54	4.24	0.87

109	Jan 2021	44.04	37.49	6.55	0.85
110	Feb 2021	34.11	33.77	0.34	0.99
111	Mar 2021	29.02	28.30	0.73	0.98
112	Apr 2021	27.81	24.82	2.99	0.89
113	May 2021	21.03	18.76	2.26	0.89
114	Jun 2021	16.90	15.99	0.92	0.95
115	Jul 2021	15.71	14.01	1.70	0.89
116	Aug 2021	13.60	13.04	0.56	0.96
117	Sep 2021	14.71	14.19	0.51	0.97
118	Oct 2021	15.31	15.12	0.19	0.99
119	Nov 2021	21.20	20.13	1.07	0.95
120	Dec 2021	29.41	29.54	0.13	1.00

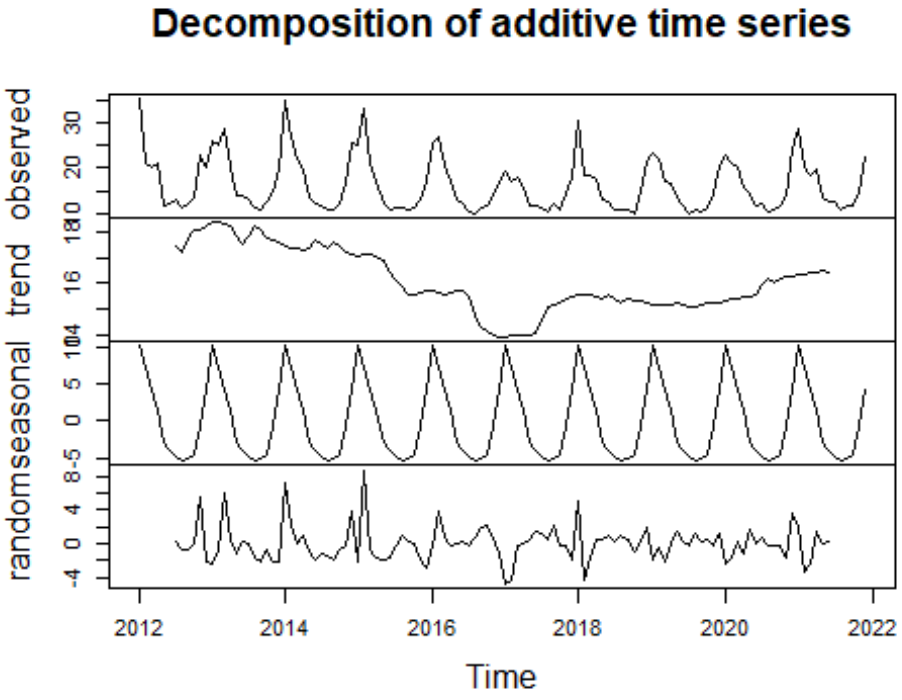
5 Year Forecast

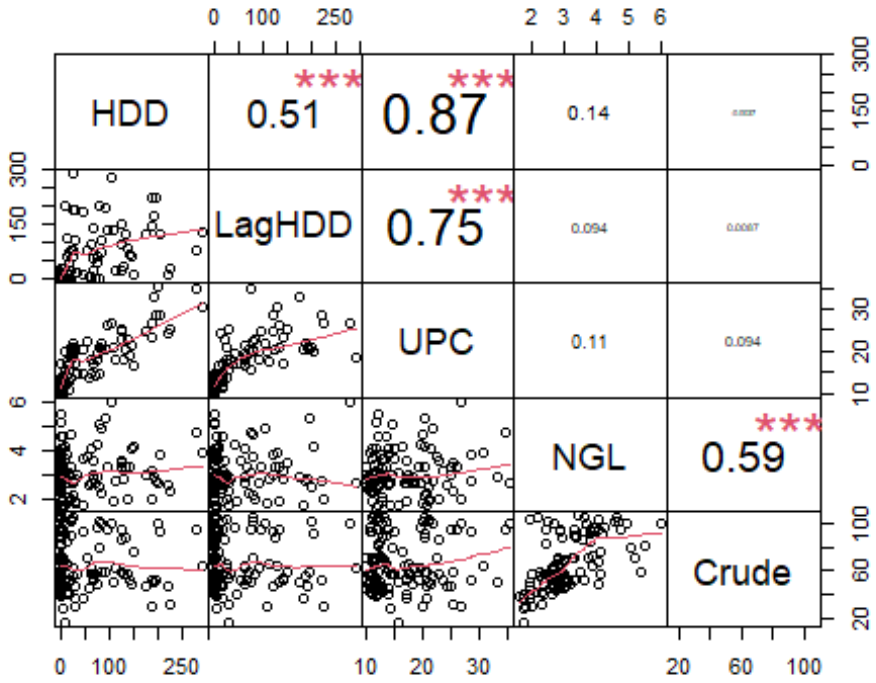


CFG Firm Transportation Service (FTS-1)

In this section we will forecast monthly client counts for FTS-1. From the data given, these numbers are calculated by filtering for Tariff Schedule ‘FTS-1’.

Customer Time-Series Decomposition

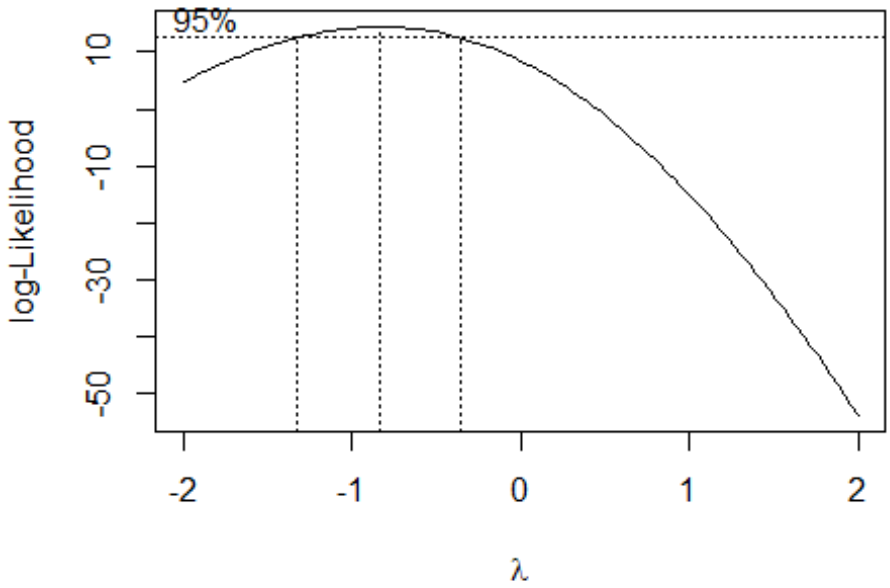




Multiple Linear Regression Model

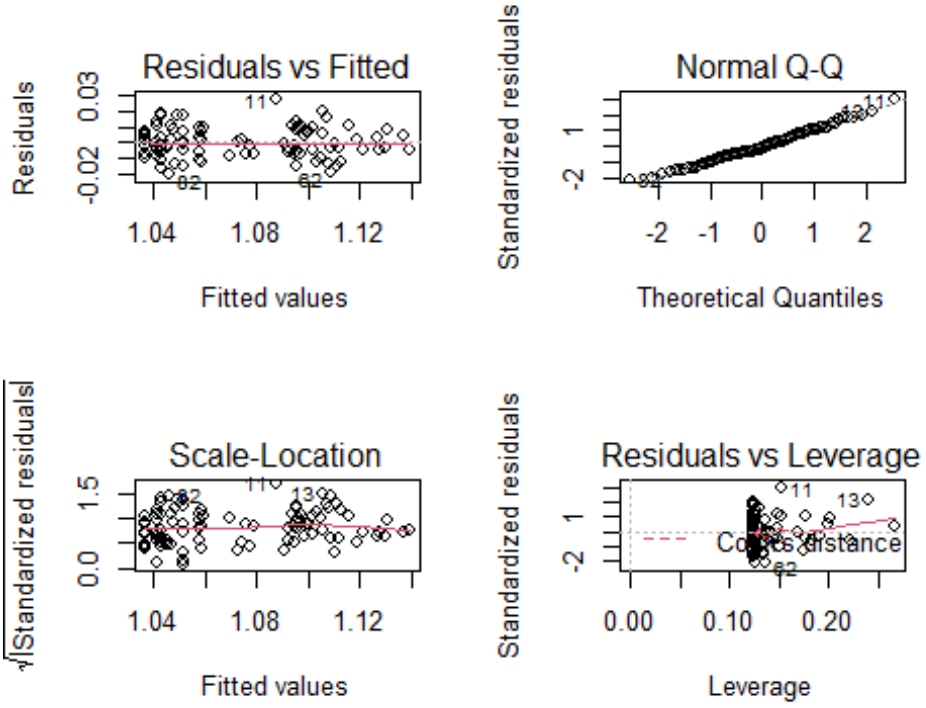
In this section we will evaluate a Multiple Linear Regression Model using Month & Residential HDD to predict Residential Use per Customer.

Diagnostics




```
[1] "Box-Cox Transformation Lambda: -0.83"


Call:
lm(formula = ((UPC^lambda - 1)/lambda) ~ HDD + month, data = train_data)

Coefficients:
(Intercept)          HDD       month2       month3       month4       month5
  1.0886216    0.0001738    0.0089987    0.0031591    0.0037874   -0.0304178
     month6       month7       month8       month9      month10      month11
 -0.0370627   -0.0451036   -0.0516192   -0.0466752   -0.0431134   -0.0221289
    month12
 -0.0030732


Call:
lm(formula = ((UPC^lambda - 1)/lambda) ~ HDD + month, data = train_data)

Residuals:
       Min         1Q     Median         3Q        Max
```

```
-0.0194158 -0.0059214 -0.0009875  0.0070425  0.0267816

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.089e+00  7.068e-03 154.024  < 2e-16 ***
HDD          1.738e-04  3.024e-05   5.747 1.50e-07 ***
month2       8.999e-03  5.936e-03   1.516 0.133407
month3       3.159e-03  6.262e-03   0.505 0.615245
month4       3.787e-03  7.581e-03   0.500 0.618691
month5      -3.042e-02  7.832e-03  -3.884 0.000207 ***
month6      -3.706e-02  7.890e-03  -4.697 1.05e-05 ***
month7      -4.510e-02  7.890e-03  -5.716 1.71e-07 ***
month8      -5.162e-02  7.890e-03  -6.542 4.88e-09 ***
month9      -4.668e-02  7.890e-03  -5.916 7.34e-08 ***
month10     -4.311e-02  7.687e-03  -5.608 2.68e-07 ***
month11     -2.213e-02  6.438e-03  -3.437 0.000925 ***
month12     -3.073e-03  5.955e-03  -0.516 0.607225
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.009919 on 82 degrees of freedom
Multiple R-squared:  0.9177, Adjusted R-squared:  0.9057
F-statistic: 76.21 on 12 and 82 DF,  p-value: < 2.2e-16

          GVIF Df GVIF^(1/(2*Df))
HDD   3.997689  1        1.999422
month 3.997689 11        1.065013
```



Back-Testing

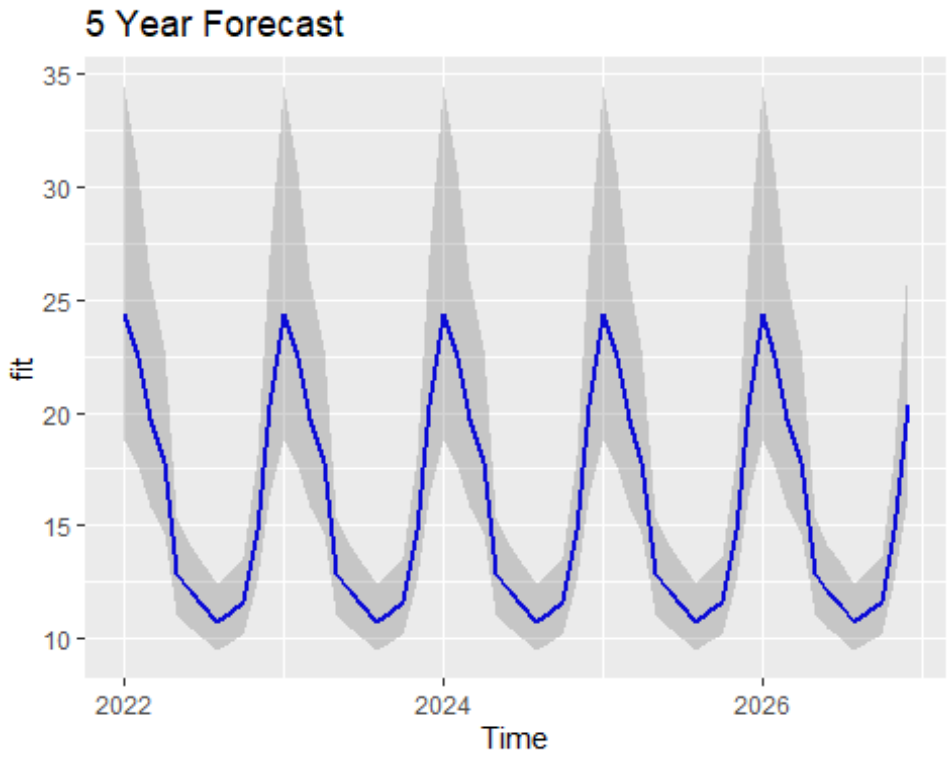
In this section we will back test our Regression model to evaluate performance accuracy. We will trim our data from 2012 to December 2019 in order to predict the next 24 months and evaluate accuracy. To imitate a real-life scenario we will also use the 20 year average HDD as the HDD regressor values.

In the table below we see that regression model performs extremely well with an overall MAE of 1.25 and overall accuracy of 93% by predicting January 2020 to December 2021.

```
[1] "24 Month Mean Absolute Error (MAE): 1.25"

[1] "24 Month Accuracy: 0.93"

        Date Actual.UPC Predicted.UPC Absolute.Error Accuracy
97   Jan 2020      22.88         23.33           0.44     0.98
98   Feb 2020      20.91         22.98           2.07     0.90
99   Mar 2020      20.05         19.65           0.41     0.98
100 Apr 2020       15.76         17.97           2.21     0.86
101 May 2020       14.30         12.69           1.61     0.89
102 Jun 2020       11.76         11.99           0.23     0.98
103 Jul 2020       11.99         11.28           0.72     0.94
104 Aug 2020       10.54         10.75           0.21     0.98
105 Sep 2020       10.97         11.14           0.18     0.98
106 Oct 2020       11.60         11.58           0.02     1.00
107 Nov 2020       13.96         14.84           0.88     0.94
108 Dec 2020       24.15         20.00           4.15     0.83
```

109	Jan 2021	28.39	23.33	5.07	0.82
110	Feb 2021	20.31	22.98	2.67	0.87
111	Mar 2021	18.26	19.65	1.38	0.92
112	Apr 2021	19.53	17.97	1.56	0.92
113	May 2021	13.61	12.69	0.93	0.93
114	Jun 2021	12.70	11.99	0.71	0.94
115	Jul 2021	12.42	11.28	1.15	0.91
116	Aug 2021	10.90	10.75	0.15	0.99
117	Sep 2021	11.55	11.14	0.40	0.96
118	Oct 2021	11.85	11.58	0.27	0.98
119	Nov 2021	14.52	14.84	0.31	0.98
120	Dec 2021	22.34	20.00	2.34	0.90

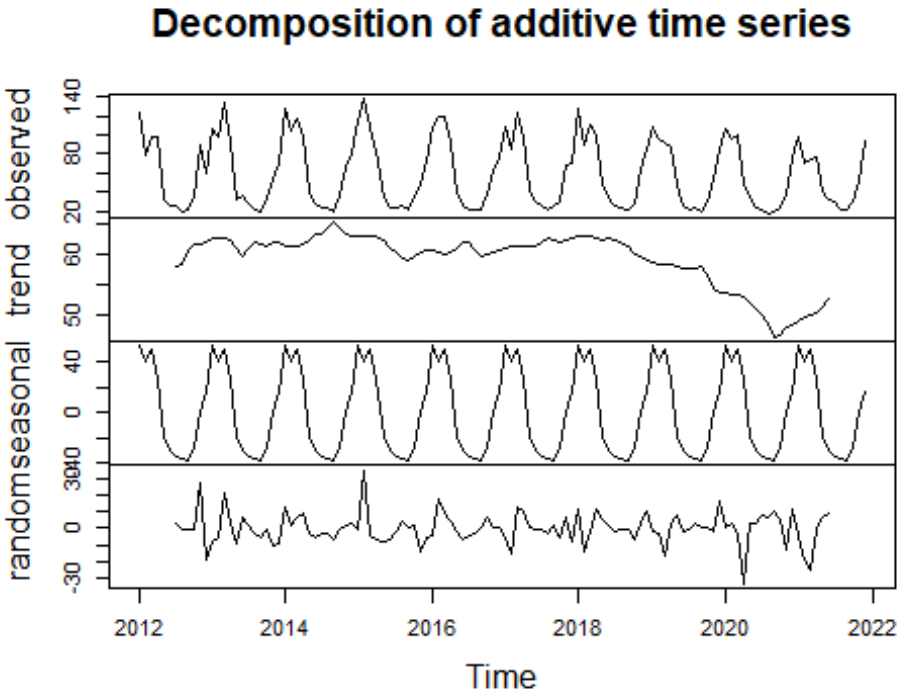
5 Year Forecast

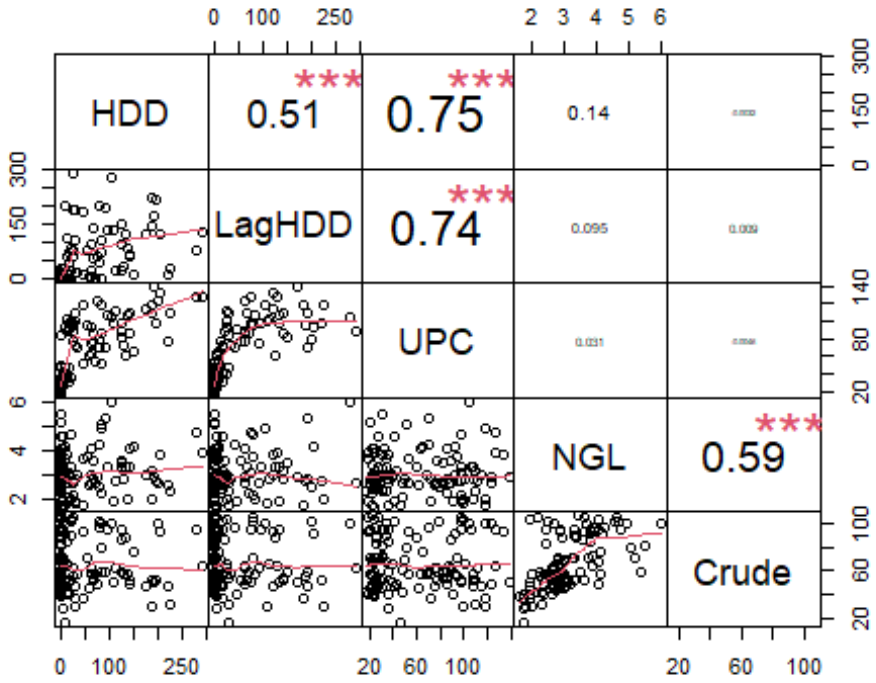


CFG Firm Transportation Service (FTS-2)

In this section we will forecast monthly client counts for FTS-2. From the data given, these numbers are calculated by filtering for Tariff Schedule'FTS-2'.

Customer Time-Series Decomposition

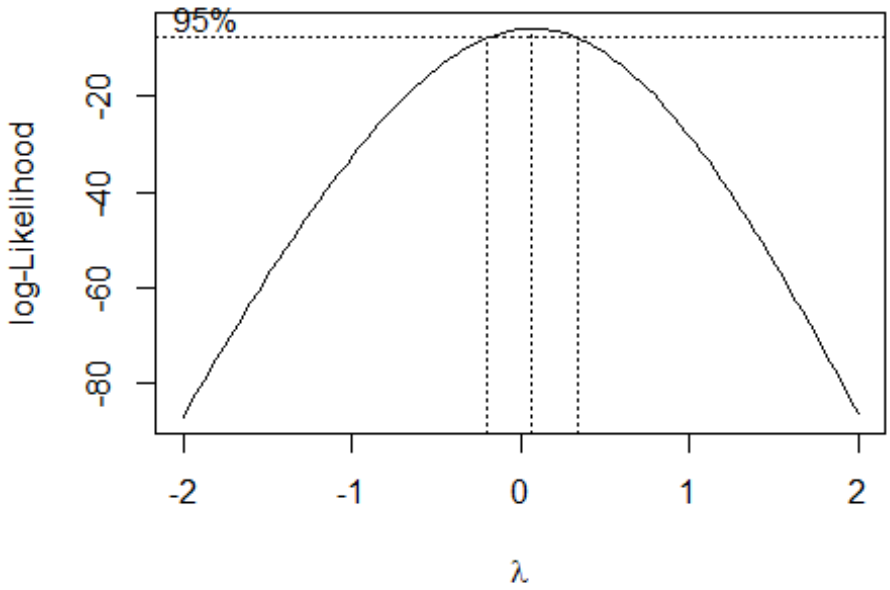




Multiple Linear Regression Model

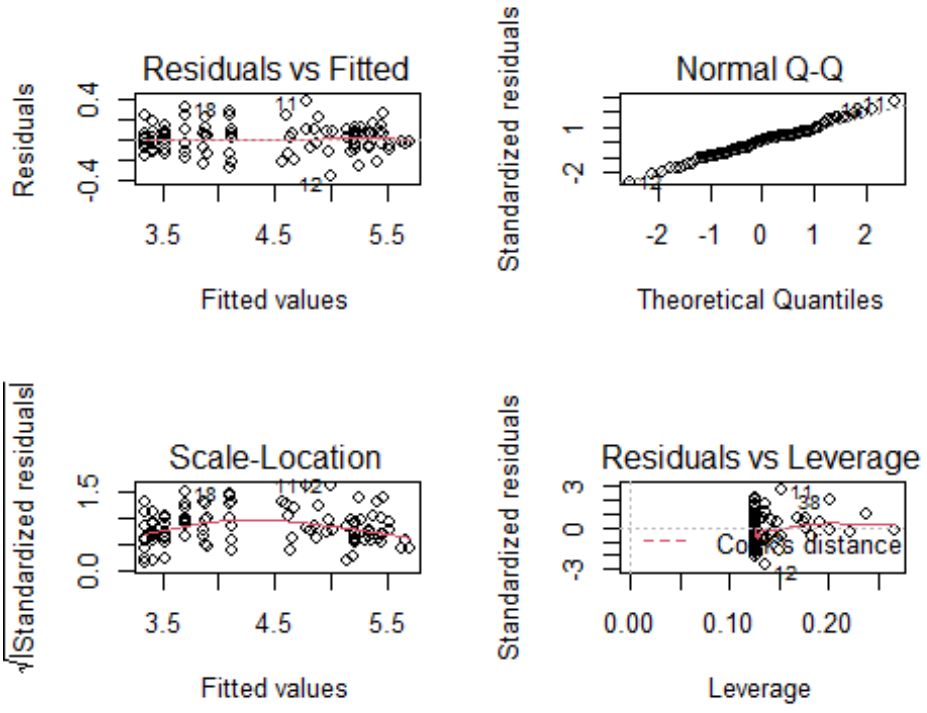
In this section we will evaluate a Multiple Linear Regression Model using Month & Residential HDD to predict Residential Use per Customer.

Diagnostics




```
[1] "Box-Cox Transformation Lambda: 0.06"


Call:
lm(formula = ((UPC^lambda - 1)/lambda) ~ HDD + month, data = train_data)

Coefficients:
(Intercept)          HDD       month2       month3       month4       month5
   5.081846     0.002024     0.012560     0.211257     0.104770    -0.977470
     month6       month7       month8       month9      month10      month11
  -1.374132    -1.556838    -1.664231    -1.738546    -1.218165    -0.552880
    month12
  -0.355884


Call:
lm(formula = ((UPC^lambda - 1)/lambda) ~ HDD + month, data = train_data)

Residuals:
     Min       1Q   Median       3Q      Max
```

```
-0.35945 -0.10032  0.00562  0.07030  0.36325

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.0818460  0.1056582  48.097  < 2e-16 ***
HDD          0.0020239  0.0004519   4.479 2.40e-05 ***
month2       0.0125598  0.0888582   0.141 0.887943
month3       0.2112565  0.0937016   2.255 0.026827 *
month4       0.1047704  0.1133862   0.924 0.358190
month5      -0.9774698  0.1171364  -8.345 1.45e-12 ***
month6      -1.3741316  0.1179907 -11.646  < 2e-16 ***
month7      -1.5568381  0.1179908 -13.195  < 2e-16 ***
month8      -1.6642308  0.1179908 -14.105  < 2e-16 ***
month9      -1.7385455  0.1179903 -14.735  < 2e-16 ***
month10     -1.2181651  0.1149769 -10.595  < 2e-16 ***
month11     -0.5528801  0.0962631  -5.743 1.52e-07 ***
month12     -0.3558838  0.0891035  -3.994 0.000141 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1485 on 82 degrees of freedom
Multiple R-squared:  0.9714, Adjusted R-squared:  0.9672
F-statistic: 232.2 on 12 and 82 DF,  p-value: < 2.2e-16
```



Back-Testing

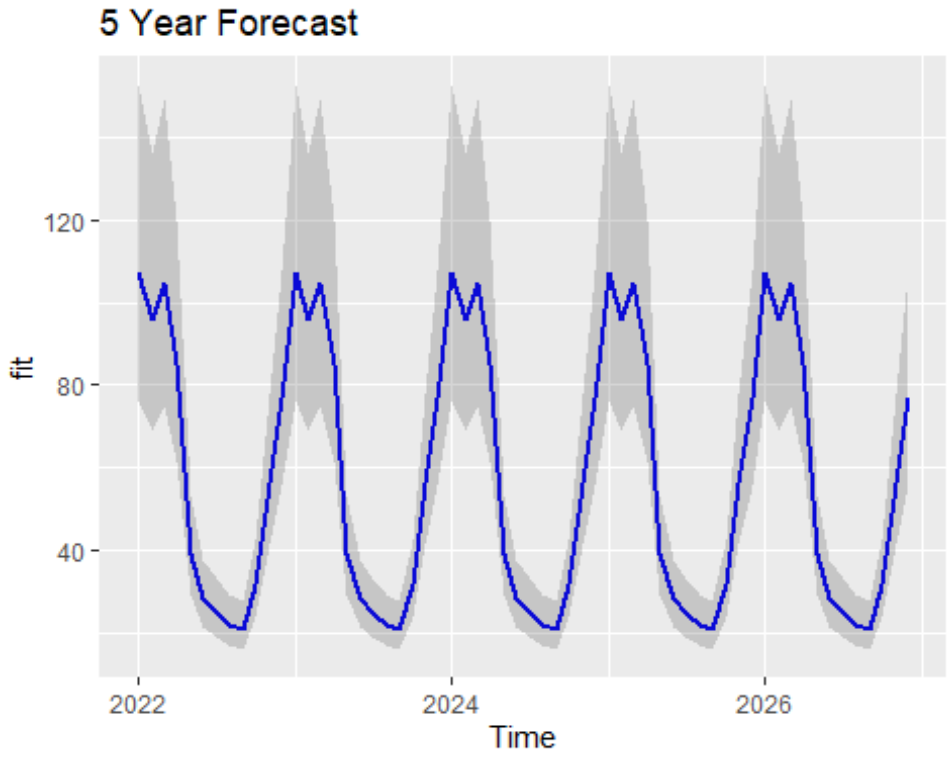
In this section we will back test our Regression model to evaluate performance accuracy. We will trim our data from 2012 to December 2019 in order to predict the next 24 months and evaluate accuracy. To imitate a real-life scenario we will also use the 20 year average HDD as the HDD regressor values.

In the table below we see that regression model performs well for most months except for a few outliers within the data that brings down the overall accuracy. In particular, April 2020 and November 2020 being the most significant of the outliers as they greatly differ from any years prior.

```
[1] "24 Month Mean Absolute Error (MAE): 10.07"

[1] "24 Month Accuracy: 0.79"
```

	Date	Actual.UPC	Predicted.UPC	Absolute.Error	Accuracy
97	Jan 2020	107.17	107.75	0.59	0.99
98	Feb 2020	96.83	99.70	2.86	0.97
99	Mar 2020	100.06	109.42	9.37	0.91
100	Apr 2020	45.93	93.73	47.80	-0.04
101	May 2020	36.56	39.34	2.78	0.92
102	Jun 2020	22.96	28.44	5.48	0.76
103	Jul 2020	21.84	24.48	2.64	0.88
104	Aug 2020	17.69	22.39	4.70	0.73
105	Sep 2020	18.64	21.05	2.42	0.87
106	Oct 2020	24.21	32.76	8.55	0.65
107	Nov 2020	36.04	60.10	24.06	0.33
108	Dec 2020	76.08	75.96	0.13	1.00
109	Jan 2021	97.01	107.75	10.75	0.89
110	Feb 2021	70.95	99.70	28.75	0.59
111	Mar 2021	74.56	109.42	34.87	0.53
112	Apr 2021	76.76	93.73	16.97	0.78
113	May 2021	39.54	39.34	0.20	1.00
114	Jun 2021	31.35	28.44	2.91	0.91
115	Jul 2021	30.12	24.48	5.65	0.81
116	Aug 2021	22.04	22.39	0.36	0.98
117	Sep 2021	20.81	21.05	0.24	0.99
118	Oct 2021	29.88	32.76	2.88	0.90
119	Nov 2021	51.94	60.10	8.16	0.84
120	Dec 2021	94.64	75.96	18.68	0.80

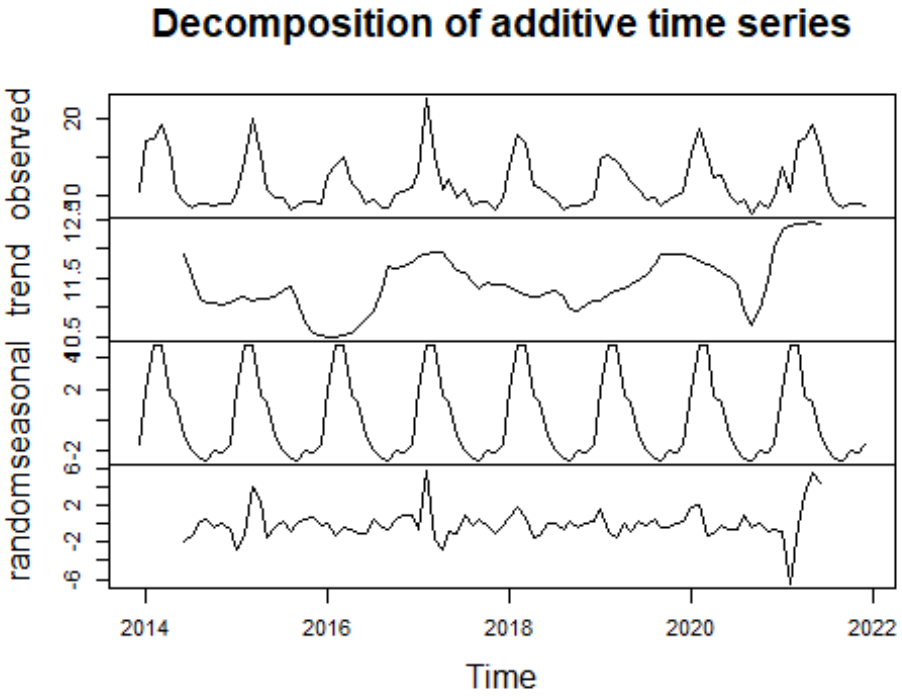
5 Year Forecast

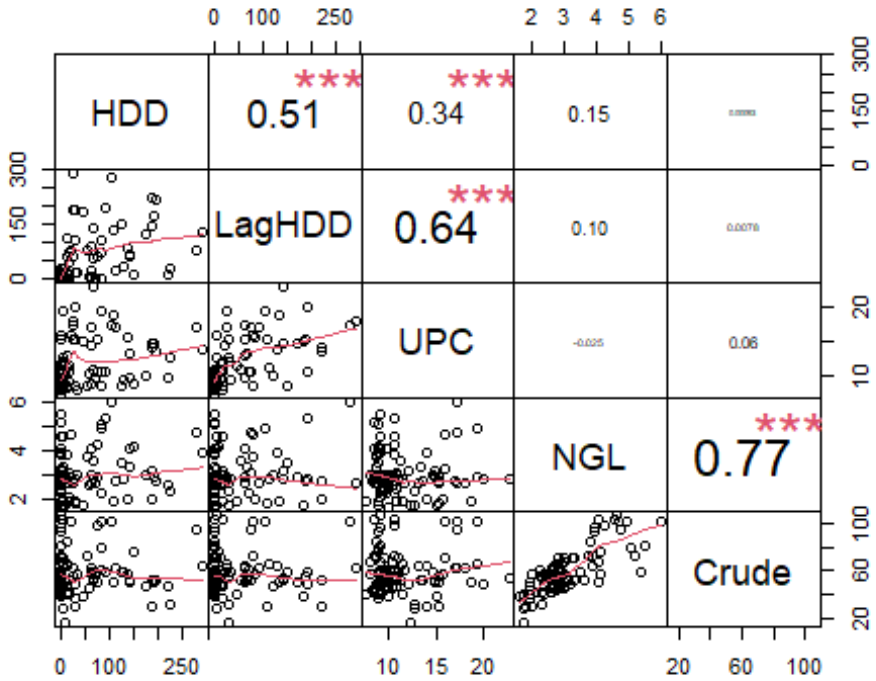


Ft. Meade Residential Service (FT-RS)

In this section we will forecast monthly client counts for FT-RS. From the data given, these numbers are calculated by filtering for Rate Class 22.

Customer Time-Series Decomposition

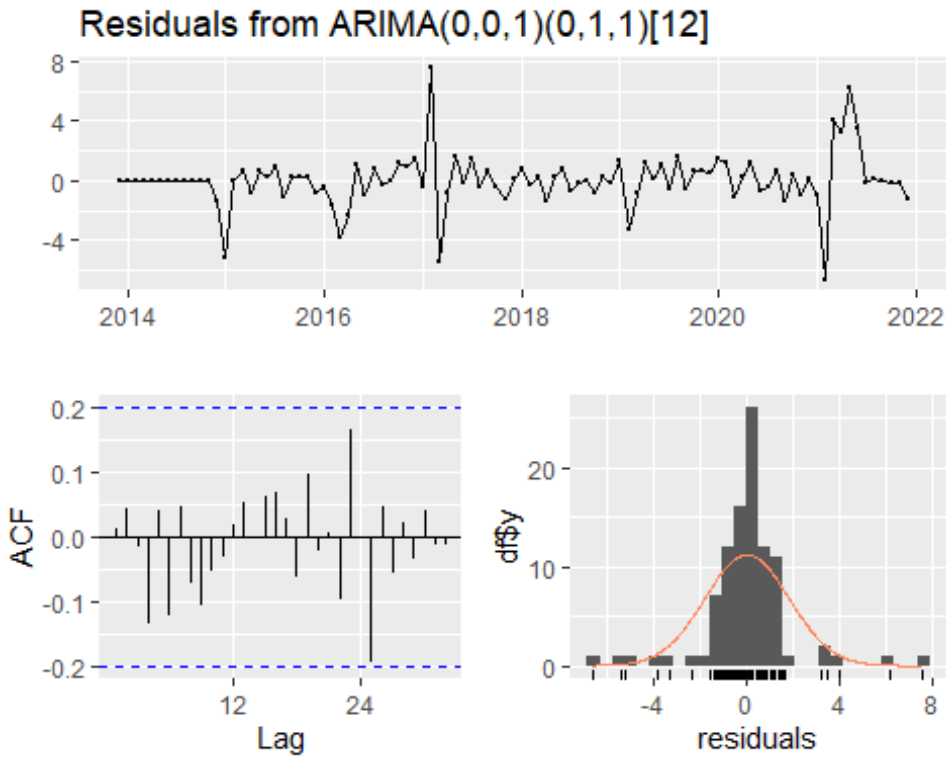




ARIMA Model

Diagnostics

```
Series: resi_upc
ARIMA(0,0,1)(0,1,1)[12]

Coefficients:
          ma1     sma1
      0.3967  -0.6482
s.e.  0.1050   0.1963

sigma^2 estimated as 3.813:  log likelihood=-179.83
AIC=365.65   AICc=365.95   BIC=372.98

Training set error measures:
                     ME     RMSE      MAE         MPE    MAPE      MASE
Training set 0.009258739 1.806218 1.05502 -0.9152387 8.53301 0.7068592
                    ACF1
Training set 0.01086437
```




```
 Ljung-Box test

data:  Residuals from ARIMA(0,0,1)(0,1,1)[12]
Q* = 9.3821, df = 17, p-value = 0.9275
```

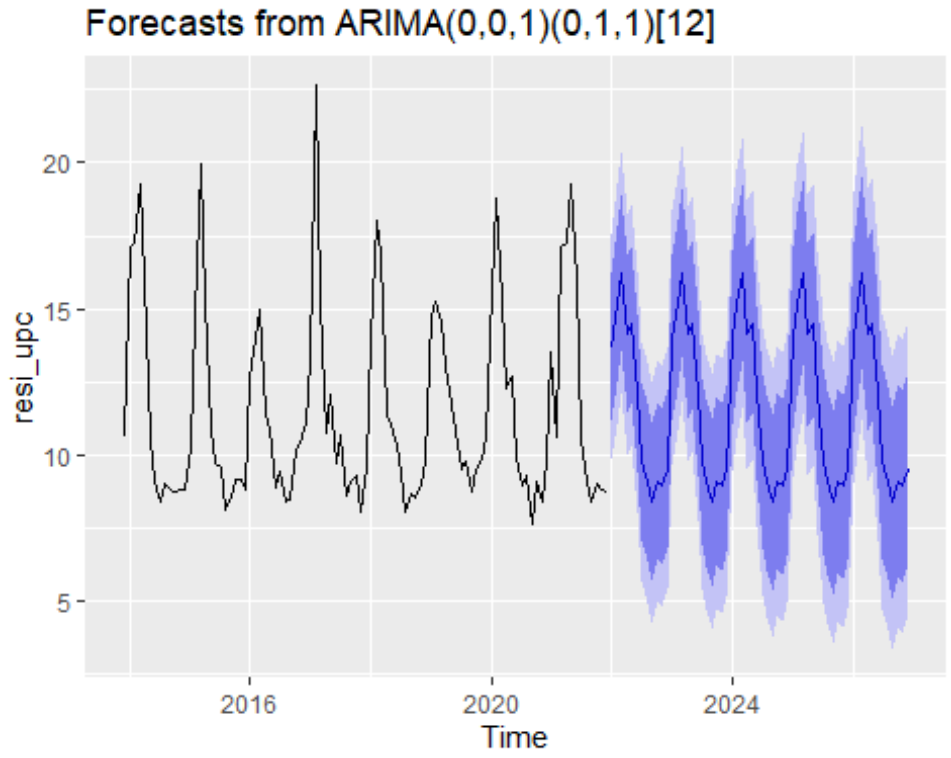
```
Model df: 2.   Total lags used: 19
```

Back-Testing

```
[1] "24 Month Mean Absolute Error (MAE): 1.64"

[1] "24 Month Accuracy: 0.88"
```

	Date	Actual.UPC	Predicted.UPC	Absolute.Error	Accuracy
1	Jan 2020	15.63	13.83	1.80	0.88
2	Feb 2020	18.78	16.93	1.85	0.90
3	Mar 2020	15.34	16.32	0.98	0.94
4	Apr 2020	12.25	12.62	0.37	0.97
5	May 2020	12.68	11.16	1.51	0.88
6	Jun 2020	9.91	9.82	0.08	0.99
7	Jul 2020	8.93	9.44	0.51	0.94
8	Aug 2020	9.33	8.64	0.68	0.93
9	Sep 2020	7.64	8.62	0.98	0.87
10	Oct 2020	9.11	9.12	0.01	1.00
11	Nov 2020	8.43	9.14	0.71	0.92
12	Dec 2020	9.75	9.88	0.13	0.99
13	Jan 2021	13.55	13.66	0.11	0.99
14	Feb 2021	10.59	16.90	6.31	0.40
15	Mar 2021	17.16	16.29	0.87	0.95
16	Apr 2021	17.26	12.59	4.67	0.73
17	May 2021	19.29	11.13	8.16	0.58
18	Jun 2021	15.98	9.79	6.20	0.61
19	Jul 2021	10.57	9.40	1.17	0.89
20	Aug 2021	9.12	8.61	0.51	0.94
21	Sep 2021	8.42	8.59	0.16	0.98
22	Oct 2021	8.99	9.09	0.10	0.99
23	Nov 2021	8.80	9.10	0.31	0.97
24	Dec 2021	8.71	9.85	1.14	0.87

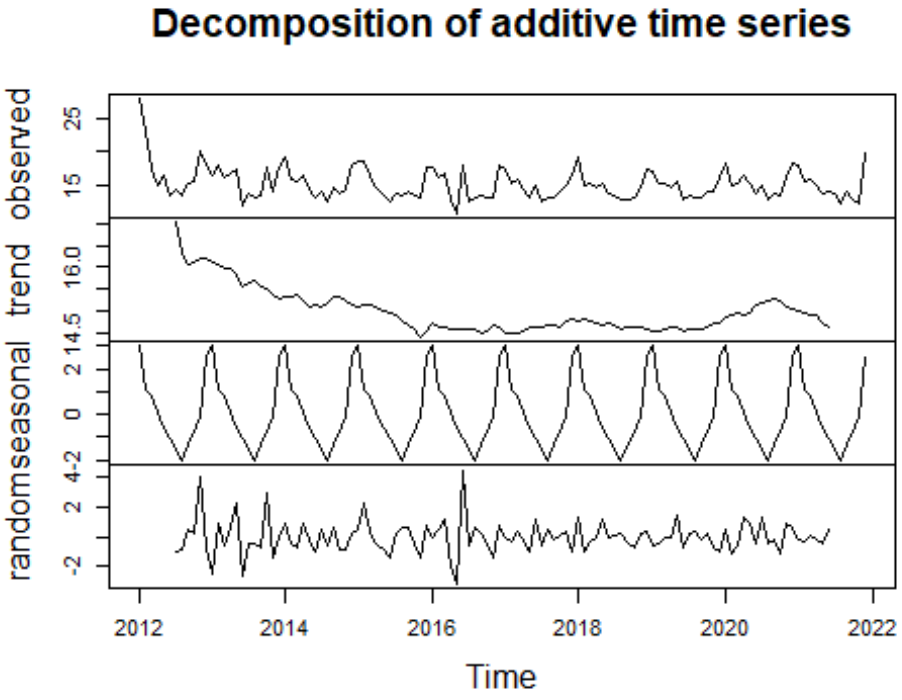
5 Year Forecast

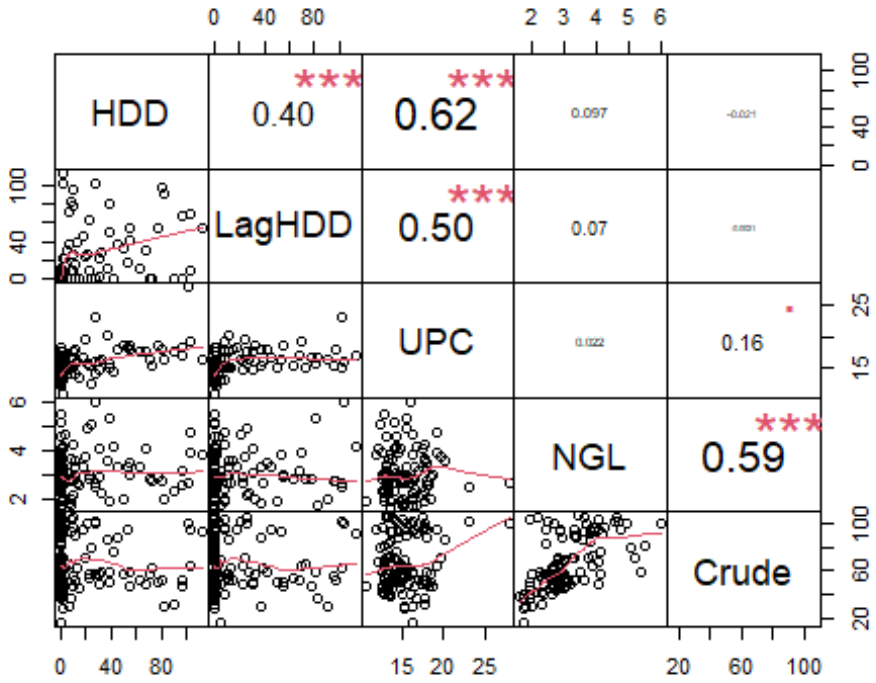


Indiantown Transportation Service 1 (IGC-TS1)

In this section we will forecast monthly client counts for IGC-TS1. From the data given, these numbers are calculated by filtering for Tariff Schedule 'TS-1' or Rate Class I0.

Customer Time-Series Decomposition





ARIMA Model

For this rate class, multiple linear regression did not fit the data so we have decided to use a Time-Series model instead.

Diagnostics

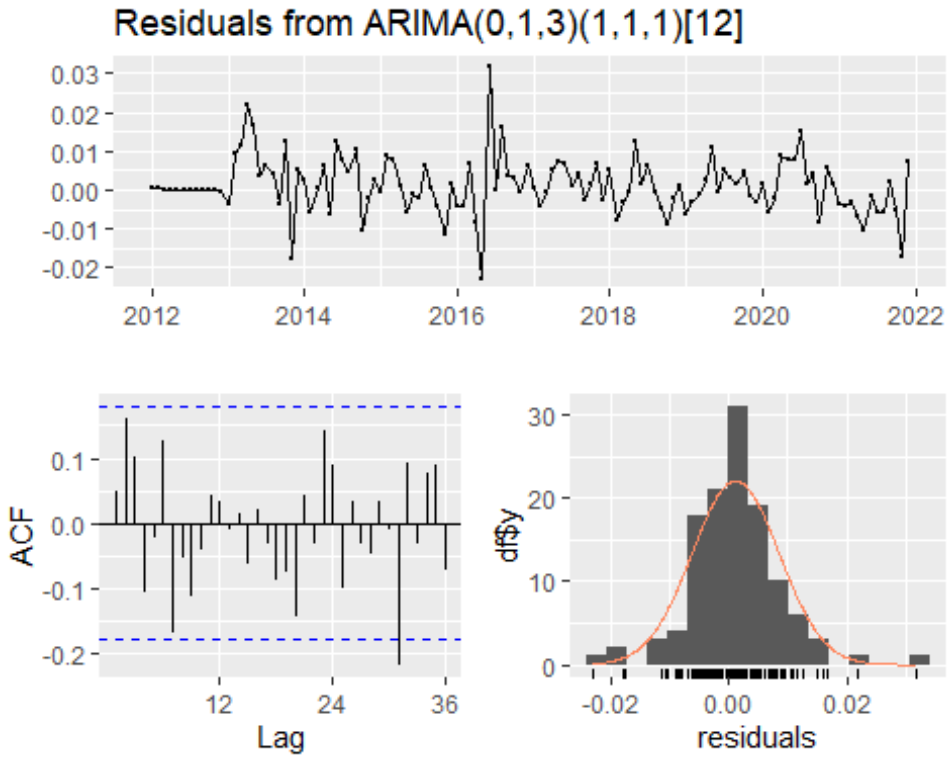
Below are the cross-validated model results to gather expected accuracy.

```
                       ME RMSE  MAE
Forecast Horizon  1   0.10 1.03 0.75
Forecast Horizon  2   0.00 1.13 0.89
Forecast Horizon  3   0.01 1.22 0.93
Forecast Horizon  4   0.07 1.28 0.91
Forecast Horizon  5  -0.01 1.41 0.99
Forecast Horizon  6   0.11 1.20 0.87
Forecast Horizon  7   0.03 1.32 0.87
Forecast Horizon  8   0.08 1.17 0.78
Forecast Horizon  9   0.09 1.18 0.78
Forecast Horizon  10  0.07 1.17 0.84
Forecast Horizon  11  0.12 1.12 0.87
Forecast Horizon  12  0.25 1.22 0.94
Forecast Horizon  13  0.24 1.46 1.02
Forecast Horizon  14  0.16 1.36 1.01
Forecast Horizon  15  0.28 1.41 1.10
Forecast Horizon  16  0.33 1.52 1.01
Forecast Horizon  17  0.20 1.32 0.96
Forecast Horizon  18  0.27 1.39 1.05
Forecast Horizon  19  0.15 1.25 0.90
Forecast Horizon  20  0.16 1.22 0.82
Forecast Horizon  21  0.19 1.24 0.83
Forecast Horizon  22  0.18 1.25 0.90
Forecast Horizon  23  0.12 1.15 0.92
Forecast Horizon  24  0.27 1.30 0.95
```

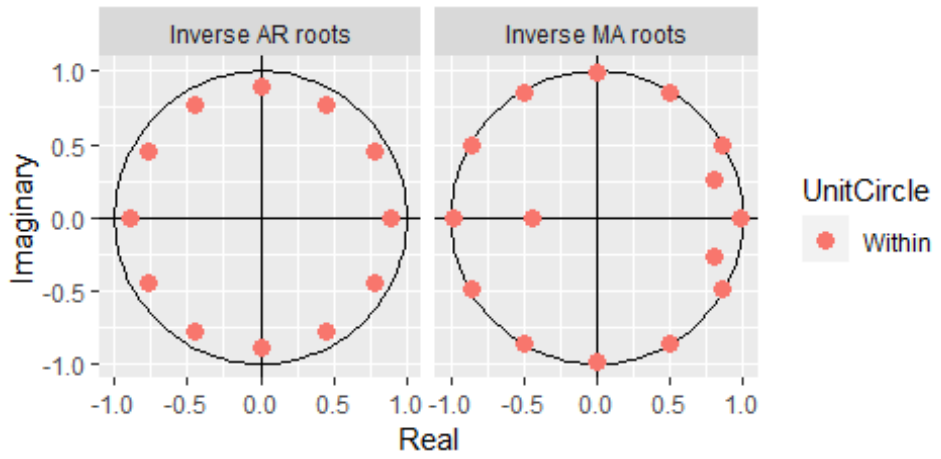
Below are the ARIMA(0,1,3)(1,1,1) model diagnostics.

```
Series: resi_upc
ARIMA(0,1,3)(1,1,1)[12]
Box Cox transformation: lambda= -0.8999268

Coefficients:
          ma1     ma2     ma3    sar1     sma1
      -1.1695  0.0082  0.3142  0.2627  -0.8432
s.e.   0.1025  0.1588  0.1125  0.1698   0.2019

sigma^2 estimated as 6.491e-05:  log likelihood=360.86
AIC=-709.72   AICc=-708.88   BIC=-693.69

Training set error measures:
```

```
                     ME      RMSE       MAE        MPE     MAPE      MASE
Training set 0.1995961 1.223077 0.8722494 0.9721375 5.785917 0.7040556
                    ACF1
Training set 0.02039349
```




```
 Ljung-Box test

data:  Residuals from ARIMA(0,1,3)(1,1,1)[12]
Q* = 25.349, df = 19, p-value = 0.1494

Model df: 5.   Total lags used: 24
```



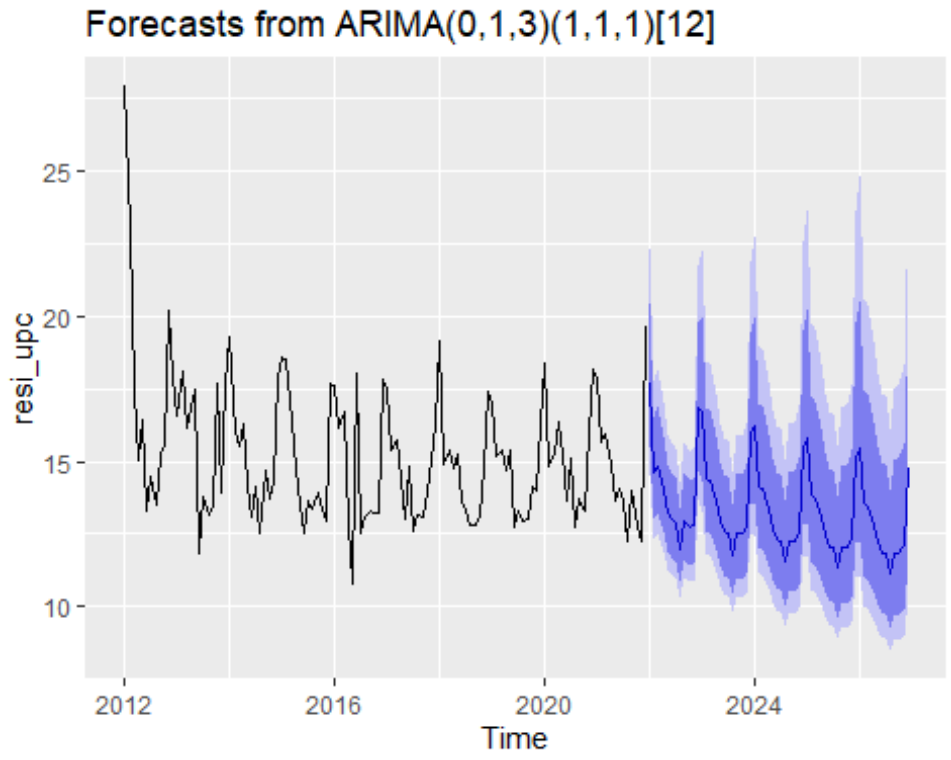
Back-Testing

We see that by Back-Testing the previous 24 Months that we have an overall accuracy level of 94% and a MAE of .87.

```
[1] "24 Month Mean Absolute Error (MAE): 0.87"

[1] "24 Month Accuracy: 0.94"

        Date Actual.UPC Predicted.UPC Absolute.Error Accuracy
1   Jan 2020      18.38         18.11           0.28     0.98
2   Feb 2020      14.83         15.67           0.84     0.94
3   Mar 2020      15.24         15.60           0.36     0.98
4   Apr 2020      16.36         14.58           1.78     0.89
5   May 2020      15.27         14.37           0.89     0.94
6   Jun 2020      13.67         13.55           0.12     0.99
7   Jul 2020      15.08         13.19           1.89     0.87
8   Aug 2020      12.71         12.97           0.27     0.98
9   Sep 2020      13.72         13.10           0.62     0.96
10 Oct 2020      13.28         13.80           0.52     0.96
11 Nov 2020      16.00         14.22           1.77     0.89
12 Dec 2020      18.20         16.79           1.42     0.92
13 Jan 2021      17.89         17.89           0.01     1.00
14 Feb 2021      15.67         15.53           0.14     0.99
15 Mar 2021      15.91         15.57           0.34     0.98
16 Apr 2021      14.92         14.55           0.37     0.98
17 May 2021      13.62         14.35           0.73     0.95
```

18 Jun 2021	14.06	13.53	0.53	0.96
19 Jul 2021	13.63	13.17	0.46	0.97
20 Aug 2021	12.22	12.95	0.73	0.94
21 Sep 2021	13.98	13.08	0.89	0.94
22 Oct 2021	12.82	13.78	0.96	0.92
23 Nov 2021	12.26	14.20	1.94	0.84
24 Dec 2021	19.67	16.75	2.91	0.85

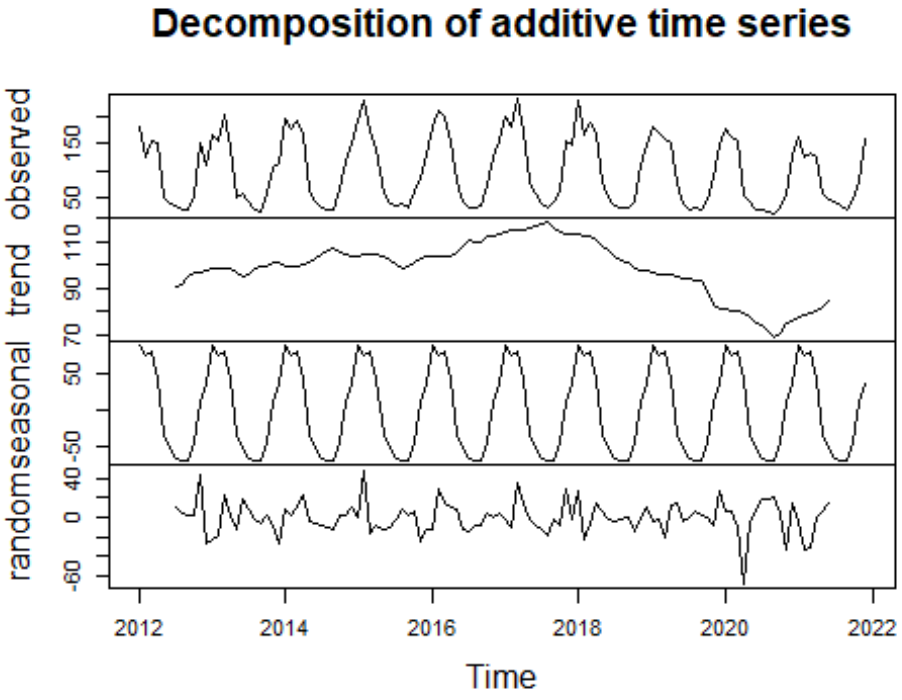
5 Year Forecast

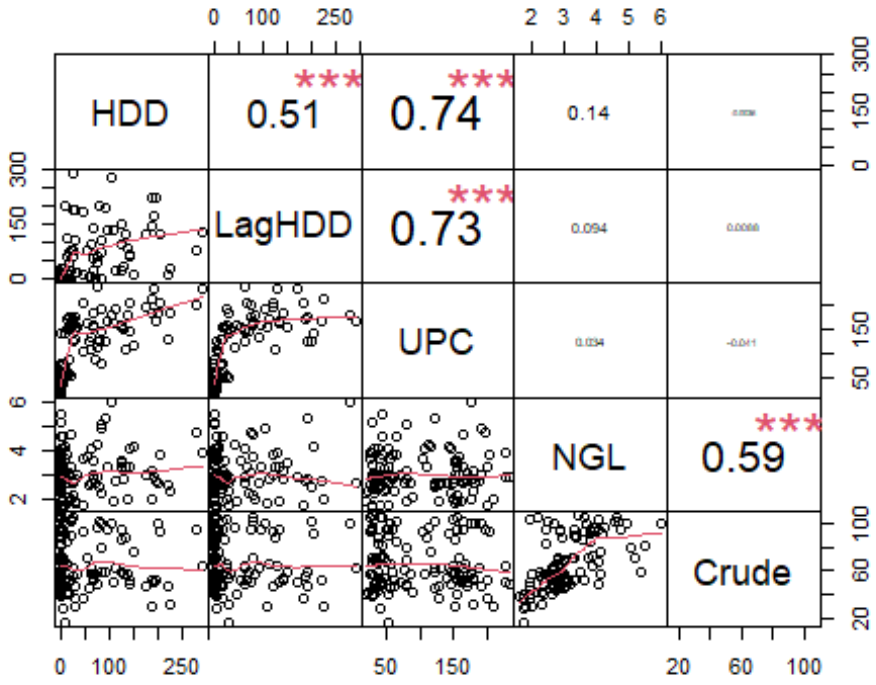


CFG Firm Transportation Service 2.1 (FTS-2.1)

In this section we will forecast monthly client counts for FTS-2.1. From the data given, these numbers are calculated by filtering for Tariff Schedule 'FTS21'.

Customer Time-Series Decomposition



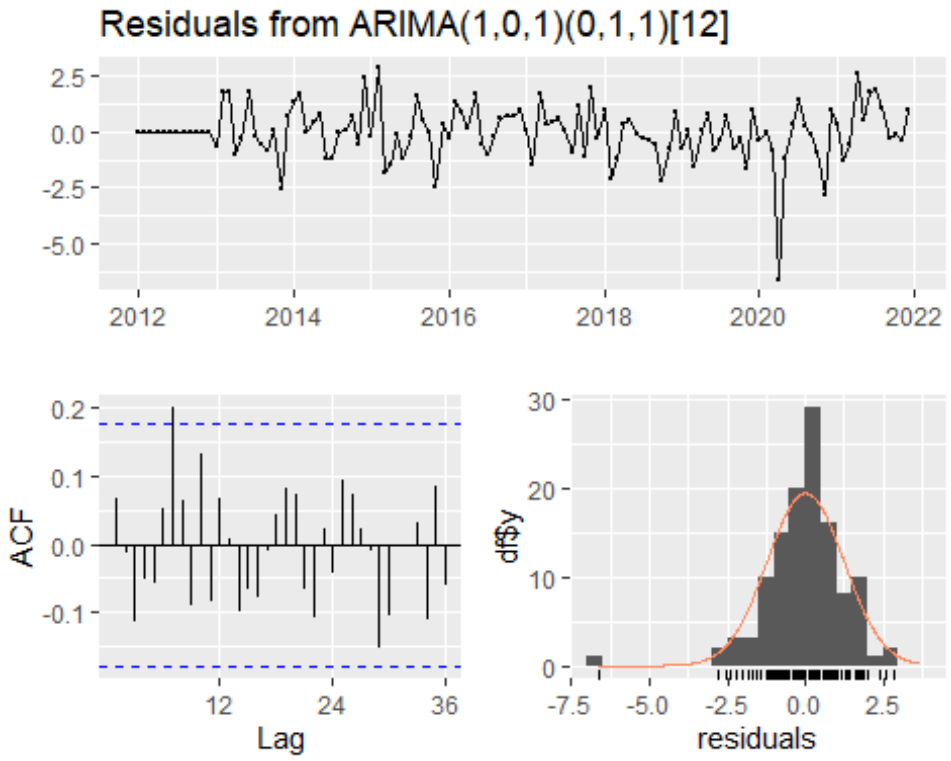


ARIMA Model

In this section we will evaluate a Multiple Linear Regression Model using Month & Residential HDD to predict Residential Use per Customer.

Diagnostics

```
Series: resi_upc
 ARIMA(1,0,1)(0,1,1)[12]
 Box Cox transformation: lambda= 0.4217038

 Coefficients:
          ar1      ma1     sma1
       0.9250  -0.6995  -0.5932
 s.e.  0.0613   0.1117   0.1126

 sigma^2 estimated as 1.717:  log likelihood=-183.52
 AIC=375.04   AICc=375.43   BIC=385.77

 Training set error measures:
                     ME     RMSE      MAE       MPE     MAPE      MASE
 Training set 0.3061915 18.64201 12.11199 -2.193503 14.0626 0.7289964
                     ACF1
 Training set -0.01828996
```




```
 Ljung-Box test
```

```
data:  Residuals from ARIMA(1,0,1)(0,1,1)[12]
Q* = 21.808, df = 21, p-value = 0.4107

Model df: 3.   Total lags used: 24
```

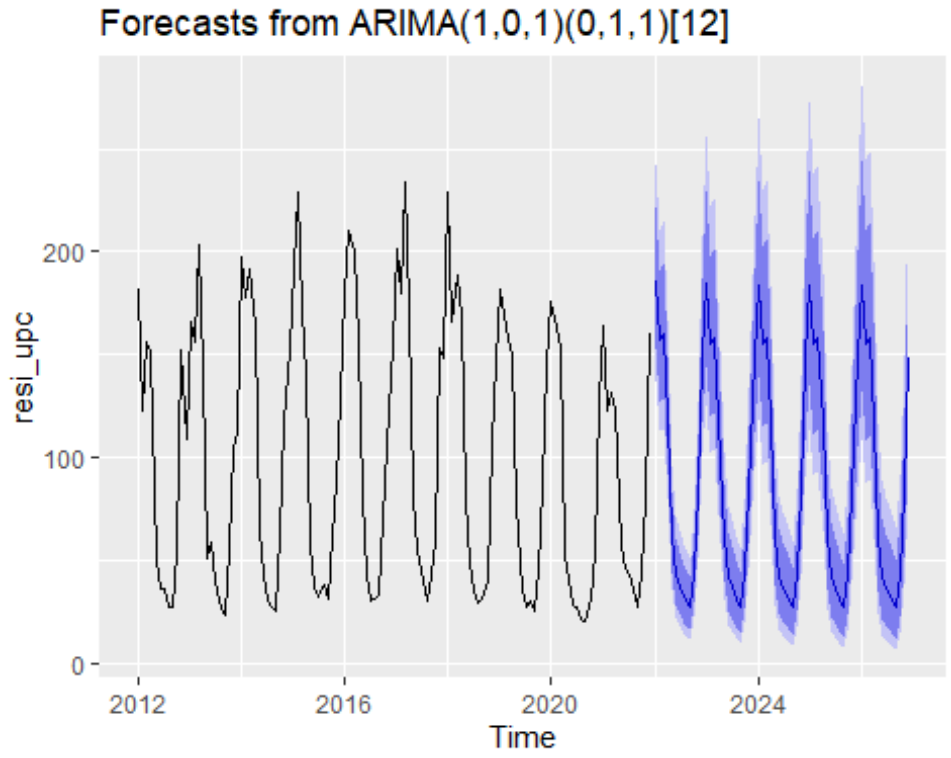
Back-Testing

In this section we will back test our Regression model to evaluate performance accuracy. We will trim our data from 2012 to December 2019 in order to predict the next 24 months and evaluate accuracy. To imitate a real-life scenario we will also use the 20 year average HDD as the HDD regressor values.

```
[1] "24 Month Mean Absolute Error (MAE): 18.67"

[1] "24 Month Accuracy: 0.69"
```

	Date	Actual.UPC	Predicted.UPC	Absolute.Error	Accuracy
1	Jan 2020	176.13	182.78	6.65	0.96
2	Feb 2020	162.59	163.52	0.93	0.99
3	Mar 2020	153.19	169.08	15.89	0.90
4	Apr 2020	53.20	147.52	94.32	-0.77
5	May 2020	40.61	69.87	29.26	0.28
6	Jun 2020	27.90	38.50	10.60	0.62
7	Jul 2020	27.39	28.17	0.78	0.97
8	Aug 2020	21.69	27.04	5.35	0.75
9	Sep 2020	20.05	26.47	6.42	0.68
10	Oct 2020	30.62	44.97	14.36	0.53
11	Nov 2020	53.37	98.86	45.49	0.15
12	Dec 2020	127.49	138.19	10.70	0.92
13	Jan 2021	163.62	182.57	18.96	0.88
14	Feb 2021	122.90	163.34	40.43	0.67
15	Mar 2021	131.01	168.90	37.89	0.71
16	Apr 2021	123.37	147.36	23.99	0.81
17	May 2021	56.47	69.78	13.30	0.76
18	Jun 2021	46.04	38.43	7.60	0.83
19	Jul 2021	43.24	28.12	15.12	0.65
20	Aug 2021	35.85	26.99	8.86	0.75
21	Sep 2021	26.73	26.42	0.30	0.99
22	Oct 2021	43.68	44.91	1.23	0.97
23	Nov 2021	80.61	98.77	18.16	0.77
24	Dec 2021	159.57	138.09	21.49	0.87

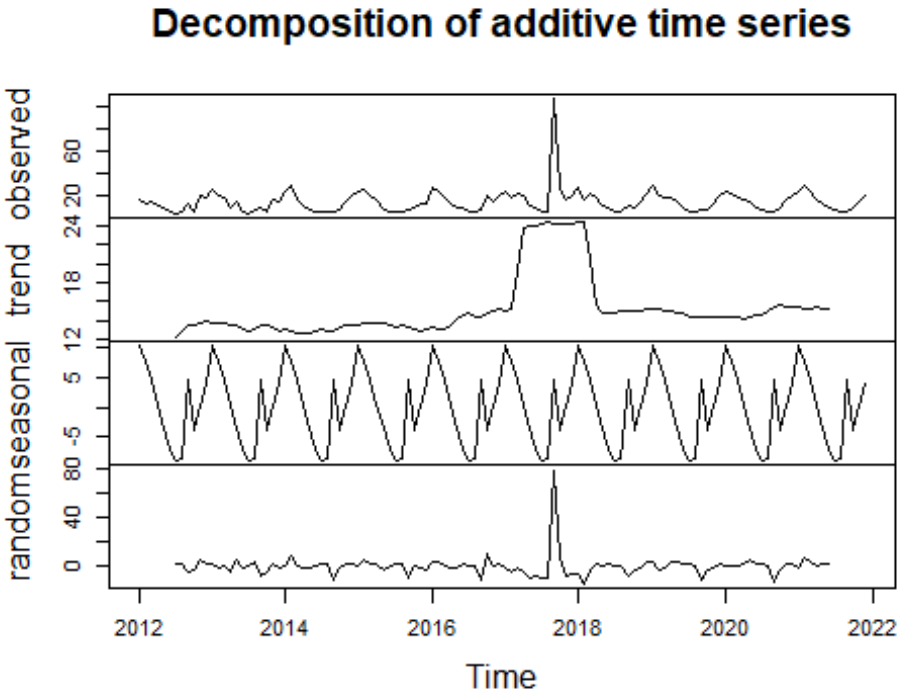
5 Year Forecast

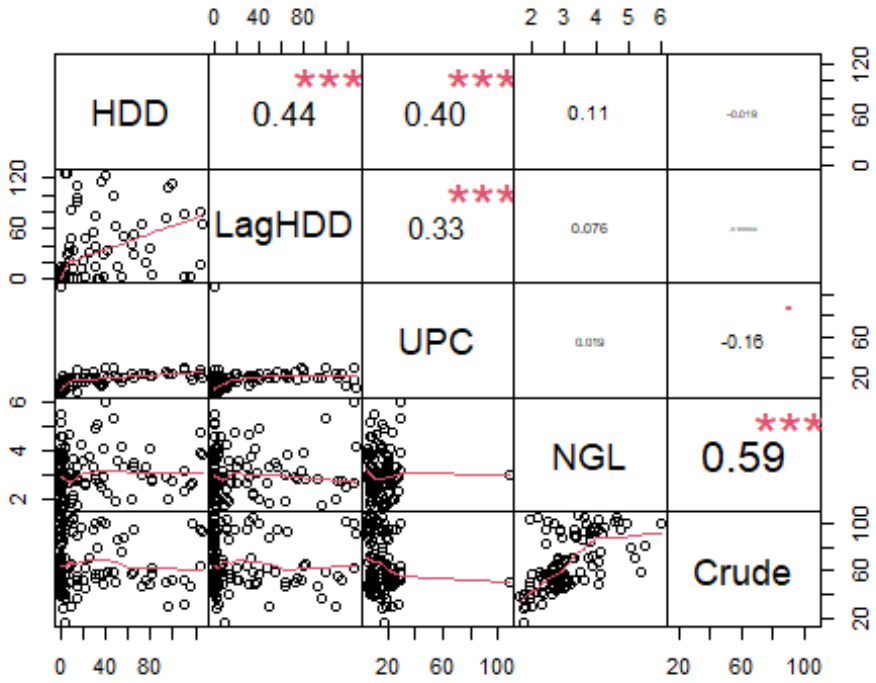


FPUC Residential Standby Generator Service (FPU-RSGS)

In this section we will forecast monthly client counts for FPU-RSGS. From the data given, these numbers are calculated by filtering for Tariff Schedule 'RS-GS'.

Customer Time-Series Decomposition





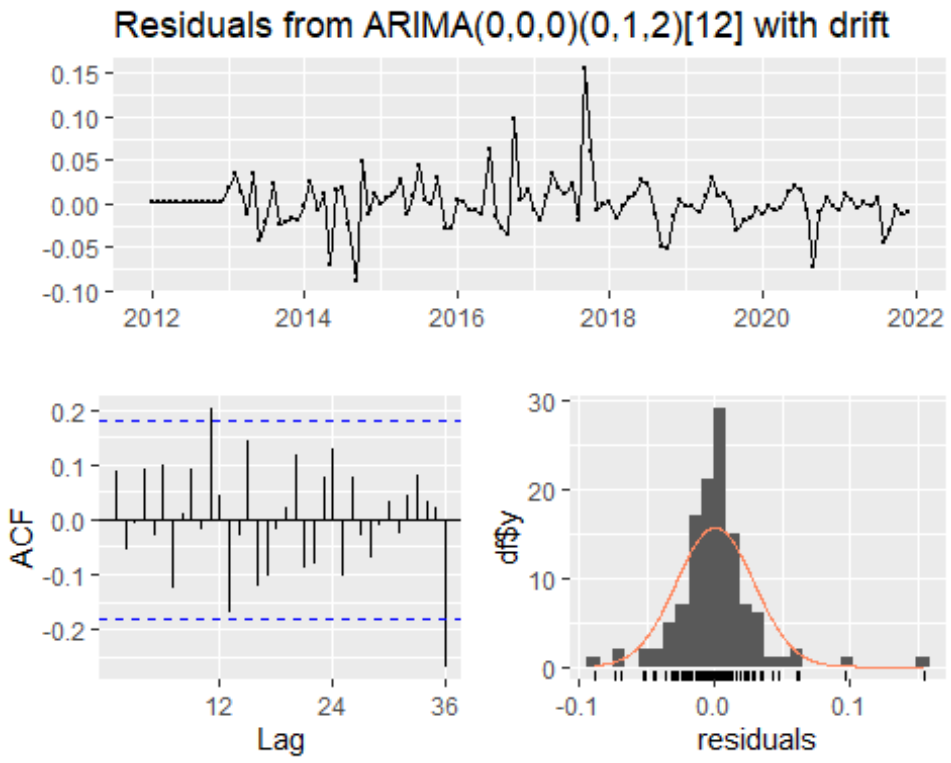
ARIMA Model

In this section we will evaluate a Multiple Linear Regression Model using Month & Residential HDD to predict Residential Use per Customer.

Diagnostics

Below we see the diagnostics of the ARIMA(0,0,1)(0,1,2) model. This ARIMA model is purely a seasonal model that looks at the previous year's moving averages to forecast forward.

```
 Series: resi_upc
 ARIMA(0,0,0)(0,1,2)[12] with drift
 Box Cox transformation: lambda= -0.8999268

 Coefficients:
          sma1     sma2  drift
       -0.4547  -0.2632  2e-04
 s.e.   0.1227   0.1067  1e-04

 sigma^2 estimated as 0.0009282:  log likelihood=221.62
 AIC=-435.25   AICc=-434.86   BIC=-424.52

 Training set error measures:
                     ME     RMSE     MAE       MPE     MAPE      MASE     ACF
1
 Training set 0.8940233 9.713584 3.00866 -1.769399 16.16157 0.7152101 0.13460
9
```




```
 Ljung-Box test

data:  Residuals from ARIMA(0,0,0)(0,1,2)[12] with drift
Q* = 30.995, df = 21, p-value = 0.07373

Model df: 3.   Total lags used: 24
```

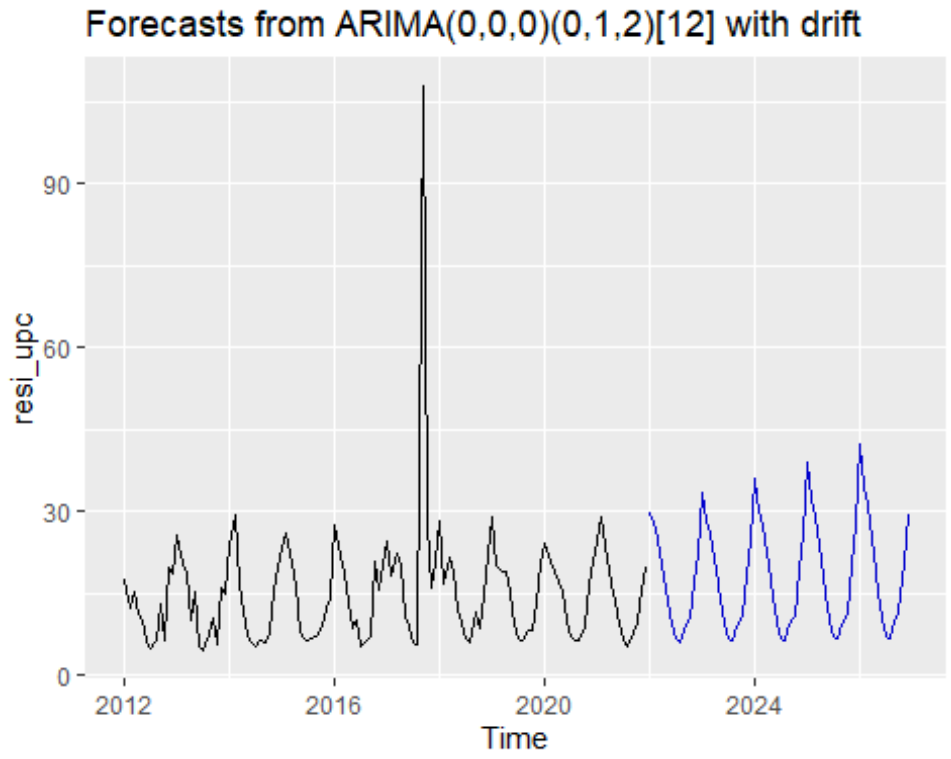
Back-Testing

```
[1] "24 Month Mean Absolute Error (MAE): 1.8"

[1] "24 Month Accuracy: 0.87"
```

	Date	Actual.UPC	Predicted.UPC	Absolute.Error	Accuracy
1	Jan 2020	24.12	26.83	2.71	0.89
2	Feb 2020	22.13	20.04	2.09	0.91
3	Mar 2020	19.07	19.22	0.16	0.99
4	Apr 2020	17.11	17.04	0.06	1.00
5	May 2020	15.37	12.45	2.92	0.81
6	Jun 2020	10.30	8.59	1.70	0.83
7	Jul 2020	7.01	6.25	0.76	0.89
8	Aug 2020	6.43	6.36	0.07	0.99
9	Sep 2020	6.32	9.51	3.19	0.50
10	Oct 2020	9.13	9.35	0.22	0.98
11	Nov 2020	16.98	14.27	2.72	0.84
12	Dec 2020	20.95	19.67	1.29	0.94
13	Jan 2021	24.53	26.21	1.69	0.93

14 Feb 2021	28.96	19.64	9.32	0.68
15 Mar 2021	23.71	19.47	4.24	0.82
16 Apr 2021	16.67	16.37	0.30	0.98
17 May 2021	13.71	11.52	2.20	0.84
18 Jun 2021	9.50	8.42	1.08	0.89
19 Jul 2021	6.96	6.13	0.83	0.88
20 Aug 2021	5.30	6.28	0.98	0.82
21 Sep 2021	6.76	9.96	3.20	0.53
22 Oct 2021	9.92	9.62	0.30	0.97
23 Nov 2021	15.60	14.61	0.99	0.94
24 Dec 2021	19.82	19.61	0.21	0.99

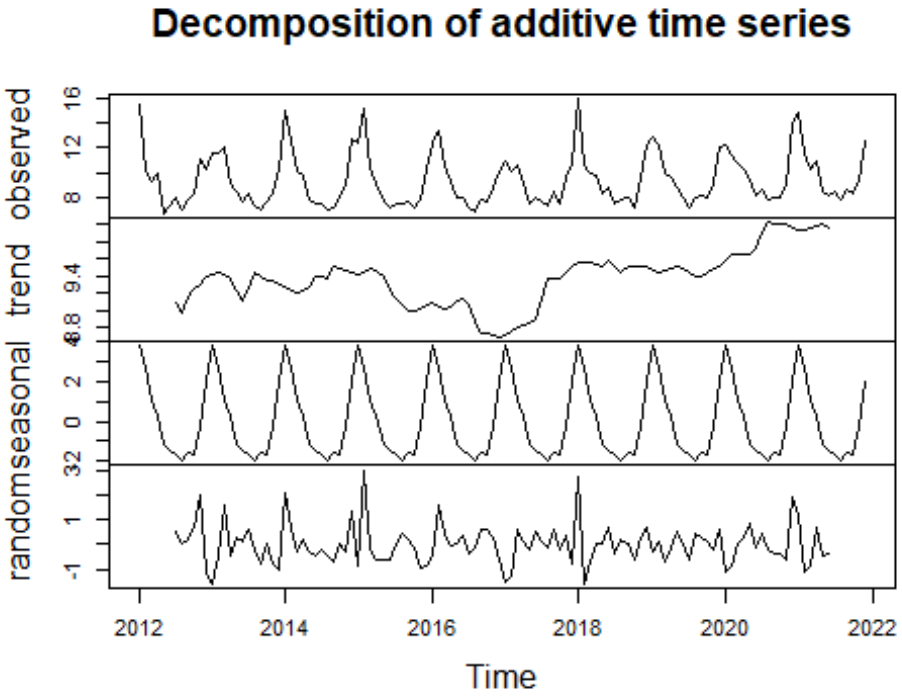
5 Year Forecast

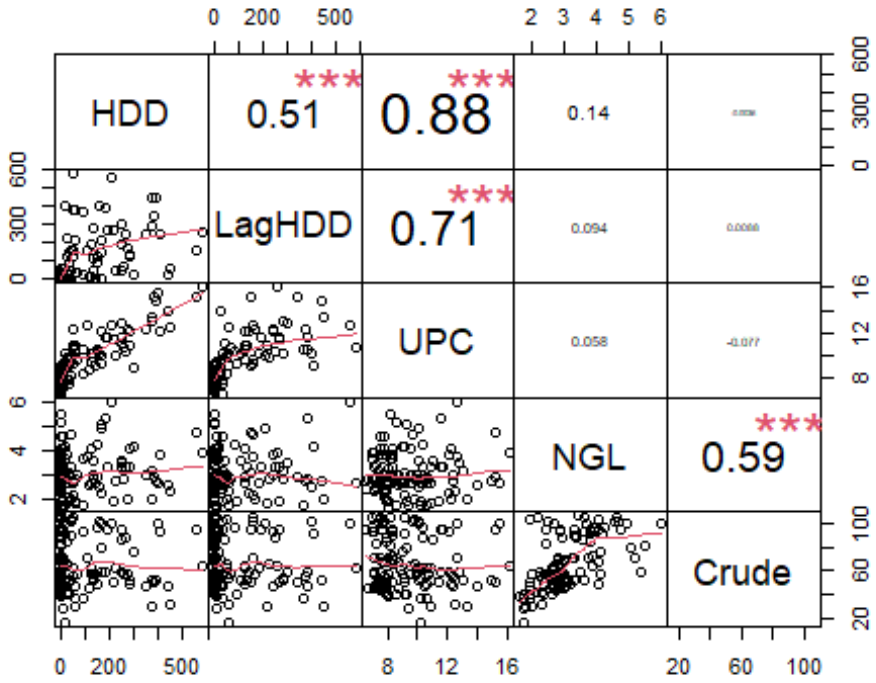


CFG Firm Transportation Service A & B (FTS-A & FTS-B)

In this section we will forecast monthly client counts for FTS-A & FTS-B from the data given, these numbers are calculated by filtering for Tariff Schedule 'FTS-A & FTS-B'.

Customer Time-Series Decomposition

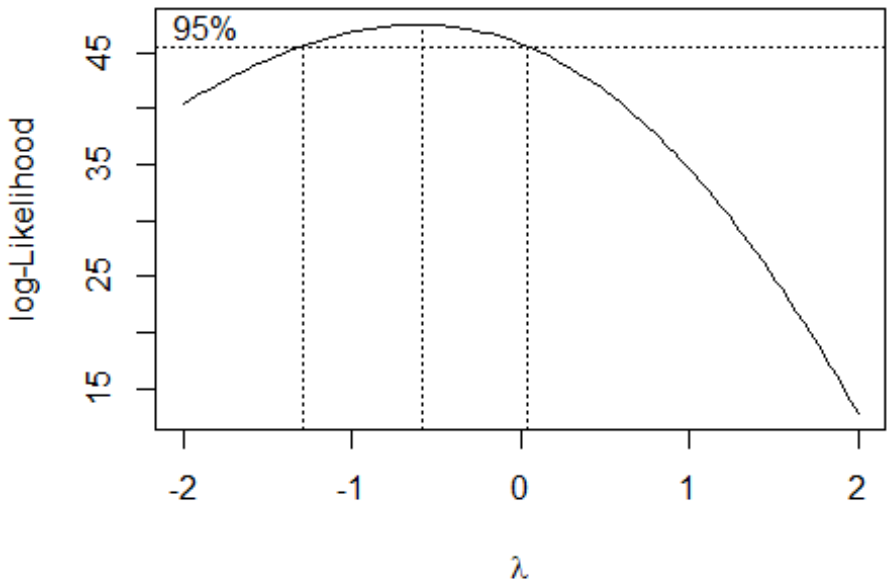




Multiple Linear Regression Model

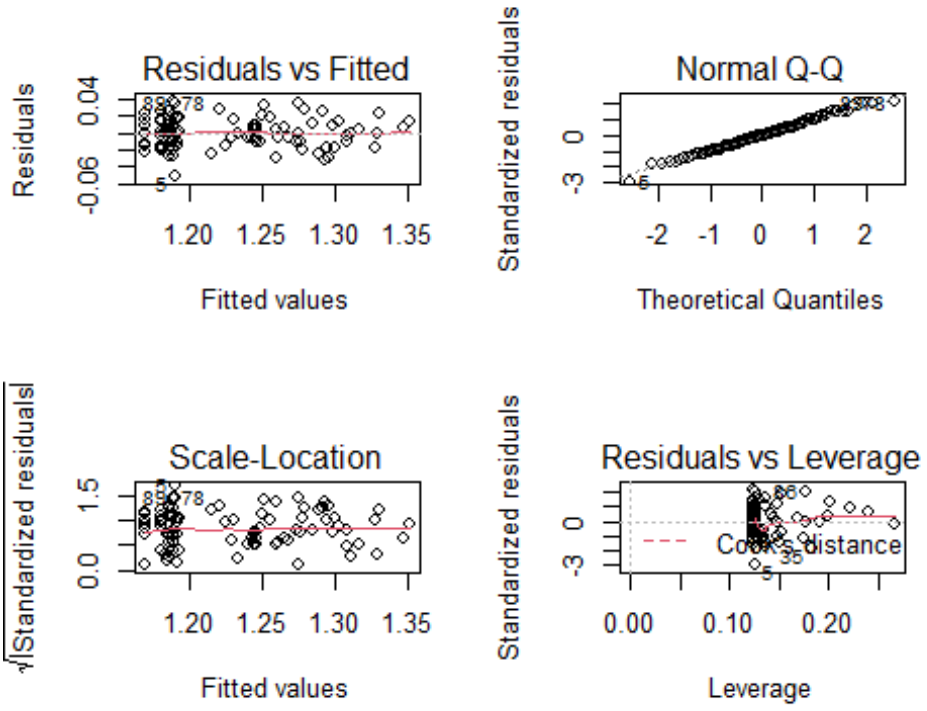
In this section we will evaluate a Multiple Linear Regression Model using Month & Residential HDD to predict Residential Use per Customer.

Diagnostics




```
[1] -0.59


Call:
lm(formula = ((UPC^lambda - 1)/lambda) ~ HDD + month, data = train_data)

Coefficients:
(Intercept)          HDD       month2       month3       month4       month5
  1.2525485    0.0001697    0.0140000   -0.0126208   -0.0129634   -0.0617005
     month6       month7       month8       month9      month10      month11
 -0.0642101   -0.0721308   -0.0830194   -0.0671218   -0.0714032   -0.0431529
    month12
 -0.0024249


Call:
lm(formula = ((UPC^lambda - 1)/lambda) ~ HDD + month, data = train_data)

Residuals:
      Min        1Q    Median        3Q       Max
```

```
-0.050145 -0.011368 -0.000190  0.009969  0.036417

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.253e+00  1.305e-02  95.980  < 2e-16 ***
HDD          1.697e-04  2.792e-05   6.078 3.66e-08 ***
month2       1.400e-02  1.097e-02   1.277 0.205320
month3      -1.262e-02  1.157e-02  -1.091 0.278367
month4      -1.296e-02  1.400e-02  -0.926 0.357189
month5      -6.170e-02  1.446e-02  -4.266 5.30e-05 ***
month6      -6.421e-02  1.457e-02  -4.407 3.15e-05 ***
month7      -7.213e-02  1.457e-02  -4.951 3.89e-06 ***
month8      -8.302e-02  1.457e-02  -5.698 1.84e-07 ***
month9      -6.712e-02  1.457e-02  -4.607 1.48e-05 ***
month10     -7.140e-02  1.420e-02  -5.030 2.85e-06 ***
month11     -4.315e-02  1.189e-02  -3.630 0.000492 ***
month12     -2.425e-03  1.100e-02  -0.220 0.826077
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01833 on 82 degrees of freedom
Multiple R-squared:  0.901,  Adjusted R-squared:  0.8866
F-statistic: 62.22 on 12 and 82 DF,  p-value: < 2.2e-16
```



Back-Testing

In this section we will back test our Regression model to evaluate performance accuracy. We will trim our data from 2012 to December 2019 in order to predict the next 24 months

and evaluate accuracy. To imitate a real-life scenario we will also use the 20 year average HDD as the HDD regressor values.

In the table below we see that regression model performs extremely well with an overall MAE of .8 and overall accuracy of 92% by predicting January 2020 to December 2021.

```
[1] "24 Month Mean Absolute Error (MAE): 0.8"

[1] "24 Month Accuracy: 0.92"
```

	Date	Actual.UPC	Predicted.UPC	Absolute.Error	Accuracy
97	Jan 2020	12.24	12.17	0.07	0.99
98	Feb 2020	11.45	11.88	0.42	0.96
99	Mar 2020	10.81	10.10	0.71	0.93
100	Apr 2020	10.19	9.47	0.72	0.93
101	May 2020	9.35	7.83	1.52	0.84
102	Jun 2020	8.08	7.74	0.33	0.96
103	Jul 2020	8.66	7.54	1.11	0.87
104	Aug 2020	7.80	7.28	0.52	0.93
105	Sep 2020	8.01	7.67	0.34	0.96
106	Oct 2020	7.96	7.64	0.32	0.96
107	Nov 2020	9.10	8.91	0.18	0.98
108	Dec 2020	13.88	11.19	2.70	0.81
109	Jan 2021	14.91	12.17	2.74	0.82
110	Feb 2021	11.54	11.88	0.34	0.97
111	Mar 2021	10.24	10.10	0.13	0.99
112	Apr 2021	10.89	9.47	1.41	0.87
113	May 2021	8.37	7.83	0.55	0.93
114	Jun 2021	8.10	7.74	0.36	0.96
115	Jul 2021	8.38	7.54	0.83	0.90
116	Aug 2021	7.79	7.28	0.51	0.93
117	Sep 2021	8.59	7.67	0.92	0.89
118	Oct 2021	8.23	7.64	0.60	0.93
119	Nov 2021	9.26	8.91	0.35	0.96
120	Dec 2021	12.62	11.19	1.43	0.89

5 Year Forecast