Non Linear Models

Dan N

2024-02-28

Piecewise Regression Model

As an alternative to linear models, we apply a piecewise regression method to non-essential visit trends in Duval County Florida to gain insight to the impact of local policy measures to mitigate the spread of COVID-19. Model fit is achieved through non-linear least squares regression methods and a non-constant rate of change. We generated parameter estimates along with 95% confidence limits. Percent decrease from baseline values are provided by Unacast, where the baseline is the average from trends before "COVID period" which begins on March 09 ,2020. The predefined knot value (*C*), or breakpoint, is the day the mandatory stay at home order was instated in Duval County on April 03, 2020. The slopes before and after the knot value represent the daily change in non-essential visits, and end on May 04, 2020 when social distancing measures were relaxed.

This is what the data looks like:

```
PROC IMPORT OUT = b
DATAFILE = 'C:/Social Distancing Models.xlsx'
DBMS = xlsx replace;
SHEET = 'Sheet2';
DATAROW = 2;
GETNAMES = yes;
RUN;
```

pan_day	Date	NEA
1	2020-03-09	-0.1016540
2	2020-03-10	-0.0747824
3	2020-03-11	-0.2456490
4	2020-03-12	-0.0557131
5	2020-03-13	-0.1103487

LOESS

First we can use the LOESS procedure to get smoothed nonparametric fit of data. At this step, we can estimate where the potential breakpoint will be.

```
PROC LOESS DATA = b plots(maxpoints = none);
MODEL NEA = pan_day;
ODS OUTPUT OUTPUTSTATISTICS = LOESSFIT;
RUN;QUIT;
```

Linear Regression

Next, we apply linear regression model to the data for later comparison to non-linear model

```
PROC REG DATA = b plots(maxpoints=none);
MODEL NEA = pan_day;
OUTPUT OUT = LINEARFIT P = PRED;
TITLE 'LINEAR REG';
RUN;QUIT;
```

LINEAR REG

The REG Procedure
Model: MODEL1
Dependent Variable: NEA NEA

Number of Observations Read	64
Number of Observations Used	57
Number of Observations with Missing Values	7

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	7262.63894	7262.63894	39.55	<.0001
Error	55	10101	183.65285		
Corrected Total	56	17364			

Root MSE	13.55186	R-Square	0.4183
Dependent Mean	-51.99326	Adj R-Sq	0.4077
Coeff Var	-26.06464		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-32.09613	3.63774	-8.82	<.0001
pan_day	pan day	1	-0.68611	0.10910	-6.29	<.0001

Note the mean square error of 183.7. We will reference this following non-linear modeling.

Below Breakpoint Model

Now we can apply a linear regression model to fit data **below** our estimated breakpoint from our LOESS model (pan_day $\leq$ 25). The slope and intercept of this model will be used as parameters in non-linear model.

```
PROC REG DATA = b plots(maxpoints = none);
MODEL NEA = pan_day;
OUTPUT OUT = FITBELOW P = PREDBELOW;
WHERE pan_day <= 25;
TITLE 'FIT BELOW';
RUN;QUIT;
```

FIT BELOW

The REG Procedure
Model: MODEL1
Dependent Variable: NEA NEA

Number of Observations Read	32
Number of Observations Used	25
Number of Observations with Missing Values	7

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	6788.23111	6788.23111	62.96	<.0001
Error	23	2479.94827	107.82384		
Corrected Total	24	9268.17939			

Root MSE	10.38383	R-Square	0.7324
Dependent Mean	-39.59842	Adj R-Sq	0.7208
Coeff Var	-26.22283		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-9.89203	4.28136	-2.31	0.0302
pan_day	pan day	1	-2.28511	0.28800	-7.93	<.0001

Above Breakpoint Model

Next, we apply a linear regression model to fit data **above** breakpoint (pan_day $>$ 25). Only the slope of this model will be used as a parameter in non-linear model.

```
PROC REG DATA = b plots(maxpoints = none);
MODEL NEA = pan_day;
OUTPUT OUT = FITABOVE P = PREDABOVE;
WHERE pan_day > 25;
TITLE 'FIT ABOVE';
RUN;QUIT;
```

FIT ABOVE

The REG Procedure
Model: MODEL1
Dependent Variable: NEA NEA

Number of Observations Read	32
Number of Observations Used	32

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	542.64511	542.64511	22.89	<.0001
Error	30	711.29406	23.70980		
Corrected Total	31	1253.93917			

Root MSE	4.86927	R-Square	0.4328
Dependent Mean	-61.67673	Adj R-Sq	0.4138
Coeff Var	-7.89483		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-80.18577	3.96352	-20.23	<.0001
pan_day	pan day	1	0.44600	0.09323	4.78	<.0001

Non-Linear Model

Now we can model our piece-wise model using the non-linear procedure. Note the parameters come from respective fit-above and fit-below linear models above.

```
*a1 = intercept of 'fit below' ;
*b1 = slope of 'fit below' ;
*b2 = slope of 'fit above' ;
*c  = est breakpoint ;

PROC NLIN DATA = b MAXITER = 1000 METHOD = MARQUARDT;
PARMS a1 = -9.89203  b1 = -2.28511  b2 = 0.44600  c = 25;
Xpart = a1 + b1*pan_day;
IF (pan_day > c) THEN DO;
Xpart = a1 + c*(b1-b2) + b2*pan_day;
end;
MODEL NEA = Xpart;
OUTPUT OUT = PIECEFIT  R = RESID  P = PRED;
ods output ParameterEstimates = Nlin_est0;
TITLE 'FIT PIECEWISE MODEL USING NONLINEAR PROCEDURE';
RUN;QUIT;
SYMBOL1 f=marker v=U i=none c=black;
SYMBOL2 v=none i=join line=1 w=3 c=black;
AXIS2 label = (a=90 r=0);
AXIS1 reflabel=(j=c c=red h=8pt 'Mandatory Stay at Home' 'Relaxed Social Distancing');
PROC GPLOT DATA=PIECEFIT;
PLOT NEA*pan_day=1 PRED*pan_day=2 / OVERLAY FRAME VAXIS=AXIS2 haxis=axis1 href=26 57;
Label NEA='% Decrease In Non-Essential Activity';
Label pan_day='Pandemic Day Beginning March 09, 2020';
Title 'Non-Essential Activity Before and After Mandatory Stay-at-Home Order: Duval County Florid
a';
RUN;QUIT;
```

FIT PIECEWISE MODEL USING NONLINEAR PROCEDURE

The NLIN Procedure
Dependent Variable NEA
Method: Marquardt

Iterative Phase					
Iter	a1	b1	b2	c	Sum of Squares
0	-9.8920	-2.2851	0.4460	25.0000	3321.3
1	-9.8920	-2.2851	0.4460	25.7382	3191.2

NOTE: Convergence criterion met.

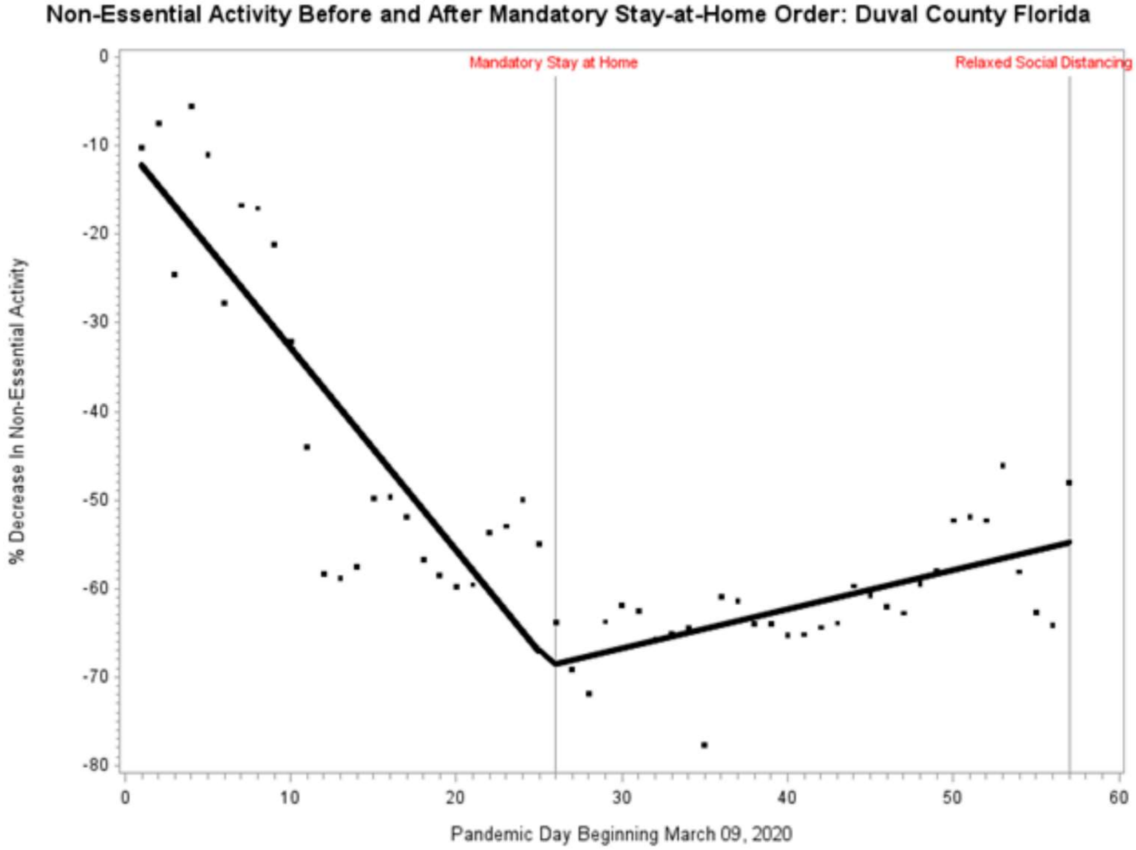
Estimation Summary	
Method	Marquardt
Iterations	1
R	1.168E-7
PPC(c)	1.659E-8
RPC(c)	0.029527
Object	0.039156
Objective	3191.242
Observations Read	64
Observations Used	57
Observations Missing	7

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Model	3	14172.3	4724.1	78.46	<.0001
Error	53	3191.2	60.2121		
Corrected Total	56	17363.5			

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
a1	-9.8920	3.1994	-16.3092	-3.4749
b1	-2.2851	0.2152	-2.7168	-1.8534
b2	0.4460	0.1486	0.1480	0.7440
c	25.7382	1.5225	22.6845	28.7918

The estimated knot value (25.7) is approximately that of the predefined mandatory stay at home order (26). The confidence limits suggest a knot value to be between 22.7 and 28.8 days after the beginning of the “COVID period”. Percent change from baseline in visits to non-essential places can be seen before the mandatory stay at home order is given at a rate of -2.29% per day (non-essential activity increases 2.29% per day). The r-squared value before *C* is 0.73, with a mean square error of 107.8. Conversely, after the order the percent change from baseline becomes smaller at a rate of .45% per day (non-essential activity decreases 0.43% per day). The r-squared value during this period is 0.43, with a mean square error of 23.71.

The non-linear model had a mean square error of 60.2, and for comparison, the linear model had a mean square error of 183.7. The non-linear model is a better fit to describe the percent decrease in non-essential activity.



Remember to keep in mind that the y-axis is a double negative; -% decreases, which are increases in non-essential activity. The next model addresses this.

Segmented Regression Model: Time Series Approach

This can be used as an alternative to the above methods. Note here non-essential activity (NEA) has been made positive for easier interpretation.

Evaluating percent difference in activity at non-essential locations in Duval County using segmented regression analysis of interrupted time series. This method can help evaluate the trends of activity before and after stay-at-home (MSaH) orders were instated. Daily changes in activity are compared to a baseline metric taken four weeks before the “COVID period” beginning March 09, 2020 and the difference is recorded. We then model the day-to-day differences in the pre-MSaH, the immediate difference following MSaH, and finally the day-to-day differences post-MSaH.

```
DATA b; SET b;
IF pan_day = . THEN delete;                    *sas may import some blank spaces;
_nea = NEA*-1;                                 *to make Non Essential Activity +;
IF pan_day <= 25 THEN time_period = 0;         *to get that indicator of before and after;
  ELSE time_period = 1;
time_after_msh = 0;                            *count the number of days after the stay at home
order;
IF pan_day > 25 THEN time_after_msh = pan_day-25;
RUN;
ODS RTF;
PROC AUTOREG DATA = b PLOTS(UNPACK) = FITPLOT;
MODEL _nea = pan_day time_period time_after_msh / METHOD=ML NLAG=10 BACKSTEP DWPROB LOGLIKL;
output out=p p=yhat lcl=lcl ucl=ucl;
RUN;
ODS RTF close;
```