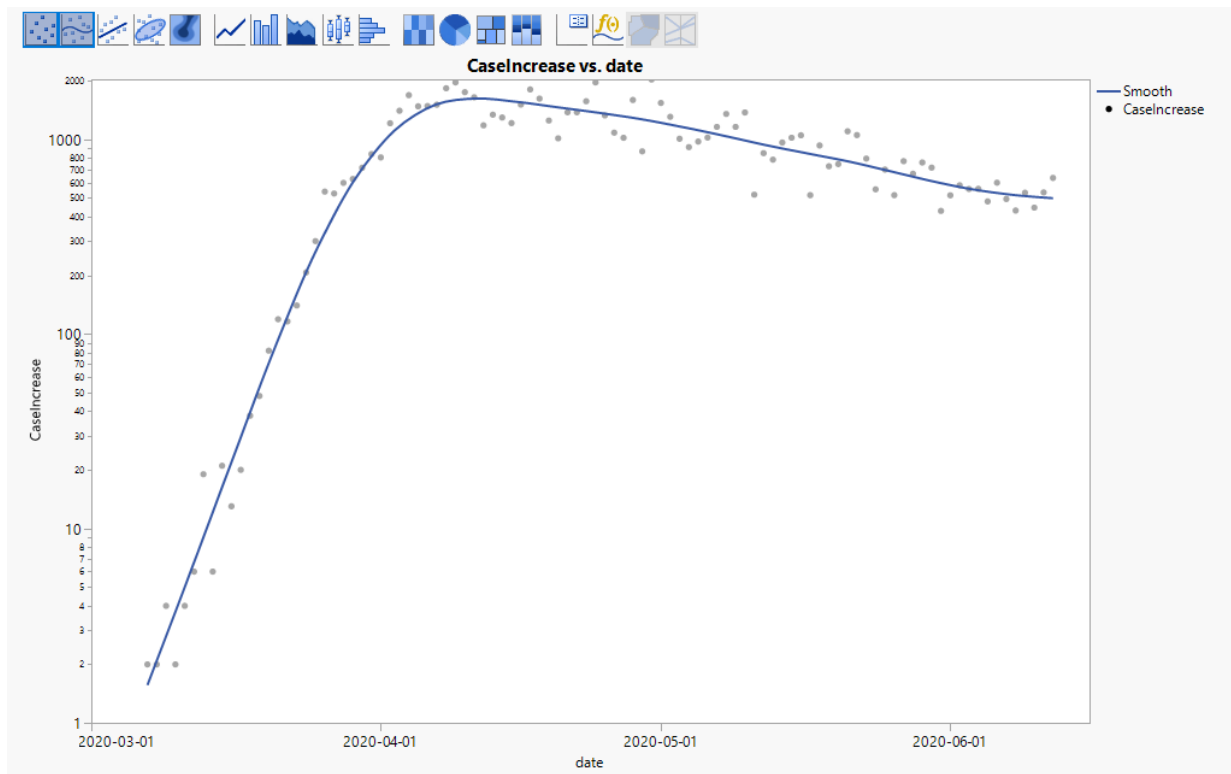Initial Scatter plot of Log(CaseIncrease) vs Date

First regression run.

Response CaseIncreaseLog

Singularity Details

Term	Details
GathRestrict10	= 0
GathRestrictAny	= 0
GathRecomAny	=OtherBusinessClose
PublicMask	= 0
Quarantine	= 0
RestaurantRestrict	=SchoolClose

Effect Summary

Source	LogWorth		PValue
NEBusinessClose	16.311		0.00000
EmergDec	12.452		0.00000
StayAtHome	3.269		0.00054
SchoolClose	.		.
RestaurantRestrict	.		.
Quarantine	.		.
PublicMask	.		.
OtherBusinessClose	.		.
GathRestrictAny	.		.
GathRestrict10	.		.
GathRecomAny	.		.

Summary of Fit

RSquare	0.906436
RSquare Adj	0.901351
Root Mean Square Error	0.542318
Mean of Response	6.104534
Observations (or Sum Wgts)	98

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	262.13516	52.4270	178.2570
Error	92	27.05805	0.2941	Prob > F
C. Total	97	289.19321		<.0001*

Parameter Estimates

Term		Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept		6.452049	0.542318	11.90	<.0001*	.
EmergDec		-4.849326	0.571654	-8.48	<.0001*	1.0997732
GathRecomAny	Biased	1.5090427	0.585771	2.58	0.0116*	11.392857
GathRestrict10	Zeroed	0	0	.	.	0
GathRestrictAny	Zeroed	0	0	.	.	0
NEBusinessClose		2.4727709	0.239627	10.32	<.0001*	2.7432945
OtherBusinessClose	Zeroed	0	0	.	.	0
PublicMask	Zeroed	0	0	.	.	0
Quarantine	Zeroed	0	0	.	.	0

RestaurantRestrict	Biased	0.962226	0.571654	1.68	0.0957	9.9773243
SchoolClose	Zeroed	0	0	.	.	0
StayAtHome		0.4363274	0.121642	3.59	0.0005*	1.2279883

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F	
EmergDec	1	1	21.164362	71.9609	<.0001*	
GathRecomAny	1	0	0.000000	.	.	LostDFs
GathRestrict10	1	0	0.000000	.	.	LostDFs
GathRestrictAny	1	0	0.000000	.	.	LostDFs
NEBusinessClose	1	1	31.318662	106.4865	<.0001*	
OtherBusinessClose	1	0	0.000000	.	.	LostDFs
PublicMask	1	0	0.000000	.	.	LostDFs
Quarantine	1	0	0.000000	.	.	LostDFs
RestaurantRestrict	1	0	0.000000	.	.	LostDFs
SchoolClose	1	0	0.000000	.	.	LostDFs
StayAtHome	1	1	3.784128	12.8664	0.0005*	

First regression run. Revealing features that were zeroed out, biased, insignificant, and correlated with eachother.

Final regression model

Response CaseIncreaseLog

Effect Summary

Source	LogWorth		PValue
NEBusinessClose	16.151		0.00000
GathRecomAny	12.305		0.00000
EmergDec	12.046		0.00000
StayAtHome	3.220		0.00060

Summary of Fit

RSquare	0.903555
RSquare Adj	0.899406
Root Mean Square Error	0.547638
Mean of Response	6.104534
Observations (or Sum Wgts)	98

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	261.30187	65.3255	217.8192
Error	93	27.89134	0.2999	Prob > F
C. Total	97	289.19321		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	6.452049	0.547638	11.78	<.0001*	.
EmergDec	-4.753103	0.574367	-8.28	<.0001*	1.0887755
GathRecomAny	2.3750461	0.282799	8.40	<.0001*	2.6040816
NEBusinessClose	2.4727709	0.241978	10.22	<.0001*	2.7432945
StayAtHome	0.4363274	0.122835	3.55	0.0006*	1.2279883

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
EmergDec	1	1	20.538171	68.4818	<.0001*
GathRecomAny	1	1	21.153164	70.5324	<.0001*
NEBusinessClose	1	1	31.318662	104.4280	<.0001*
StayAtHome	1	1	3.784128	12.6177	0.0006*

After running multiple regressions and eliminating features, this is the final regression model we came up with. Finding that EmergDec (Emergency Declaration), GathRecomAny (Gathering Recommendation Any Amount of People), NEBusinessClose (Non-Essential Business Close), and StayAtHome.

Time series.

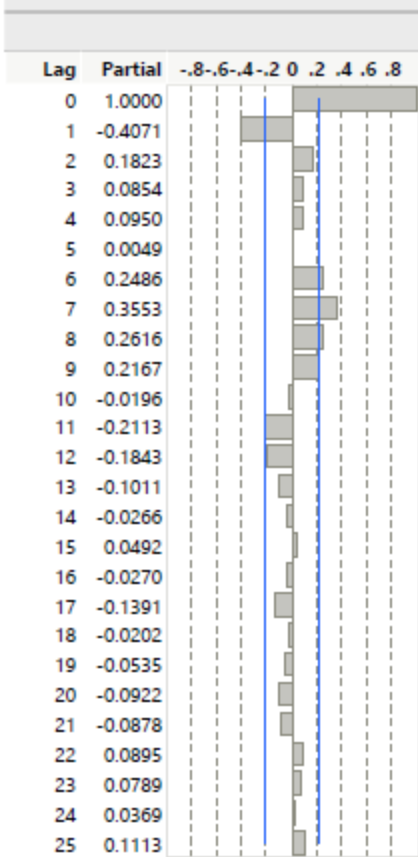## Time Series CaseIncreaseLog

Difference: (1-B)^1

Lag	AutoCorr	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value
0	1.0000		.	.
1	-0.4071		16.5773	<.0001*
2	0.3178		26.7876	<.0001*
3	-0.1112		28.0500	<.0001*
4	0.1470		30.2817	<.0001*
5	-0.0707		30.8038	<.0001*
6	0.2661		38.2787	<.0001*
7	0.0935		39.2110	<.0001*
8	0.1324		41.1014	<.0001*
9	0.0866		41.9195	<.0001*
10	-0.0428		42.1215	<.0001*
11	-0.0209		42.1701	<.0001*
12	0.0701		42.7247	<.0001*
13	0.0929		43.7116	<.0001*
14	0.1072		45.0409	<.0001*
15	0.1148		46.5846	<.0001*
16	-0.0367		46.7440	<.0001*
17	-0.0568		47.1312	0.0001*
18	0.0347		47.2779	0.0002*
19	-0.0630		47.7669	0.0003*
20	0.0715		48.4051	0.0004*
21	0.0587		48.8404	0.0005*
22	0.1098		50.3829	0.0005*
23	-0.1020		51.7325	0.0005*
24	-0.0020		51.7331	0.0008*
25	-0.0538		52.1186	0.0012*

Model: ARI(1, 1)

Model Summary

DF	95	Stable	Yes
Sum of Squared Errors	11.2942369	Invertible	Yes
Variance Estimate	0.1188867		
Standard Deviation	0.34479951		
Akaike's 'A' Information Criterion	70.8611273		
Schwarz's Bayesian Criterion	76.0105493		
RSquare	0.95649287		
RSquare Adj	0.9560349		
MAPE	6.19220365		
MAE	0.24566594		
-2LogLikelihood	66.8611273		

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate	Mu
AR1	1	-0.4034265	0.0921269	-4.38	<.0001*	0.08310023	0.05921238
Intercept	0	0.0592124	0.0247591	2.39	0.0187*		

Lag	Partial	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
0	1.0000	
1	-0.4071	
2	0.1823	
3	0.0854	
4	0.0950	
5	0.0049	
6	0.2486	
7	0.3553	
8	0.2616	
9	0.2167	
10	-0.0196	
11	-0.2113	
12	-0.1843	
13	-0.1011	
14	-0.0266	
15	0.0492	
16	-0.0270	
17	-0.1391	
18	-0.0202	
19	-0.0535	
20	-0.0922	
21	-0.0878	
22	0.0895	
23	0.0789	
24	0.0369	
25	0.1113	

Time Series CaseIncreaseLog

Model: ARI(1, 1)

Forecast
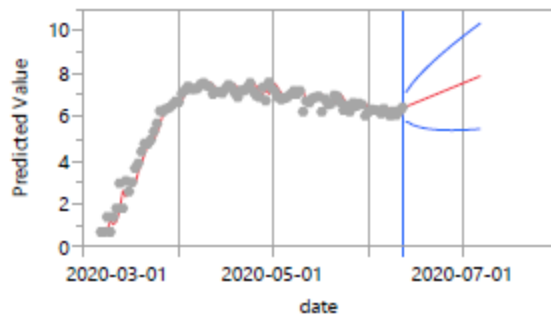
Model: ARI(5, 1)

Model Summary

DF	91	Stable	Yes
Sum of Squared Errors	10.7244452	Invertible	Yes
Variance Estimate	0.11785105		
Standard Deviation	0.3432944		
Akaike's 'A' Information Criterion	73.965844		
Schwarz's Bayesian Criterion	89.4141099		
RSquare	0.95859744		
RSquare Adj	0.95632258		
MAPE	6.10021108		
MAE	0.24357875		
-2LogLikelihood	61.965844		

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate	Mu
AR1	1	-0.3513983	0.1010664	-3.48	0.0008*	0.05781761	0.06132979
AR2	2	0.1889429	0.1066977	1.77	0.0799		
AR3	3	0.1092890	0.1090751	1.00	0.3190		
AR4	4	0.1008063	0.1067879	0.94	0.3477		
AR5	5	0.0096271	0.0917535	0.10	0.9167		
Intercept	0	0.0613298	0.0355274	1.73	0.0877		

Running time series.

Arima

Nonseasonal differencing order: 1
Differencing order: 1
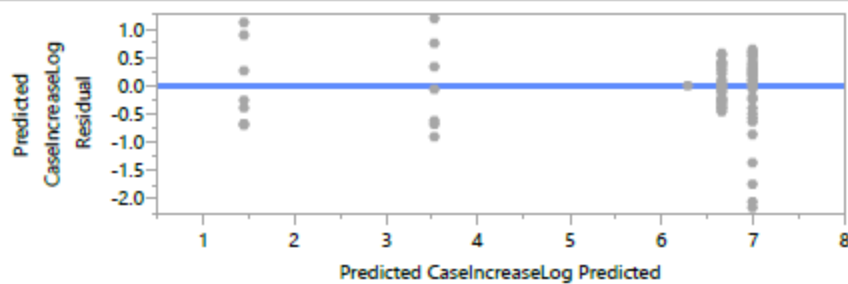Autoregression: 1

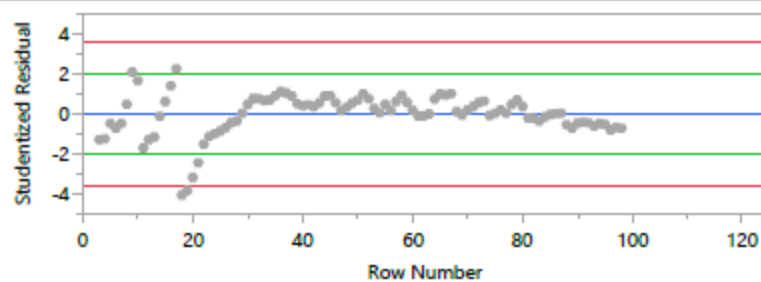## Response Predicted CaseIncreaseLog

Actual by Predicted Plot

RMSE=0.5746 RSq=0.90 PValue=<.0001
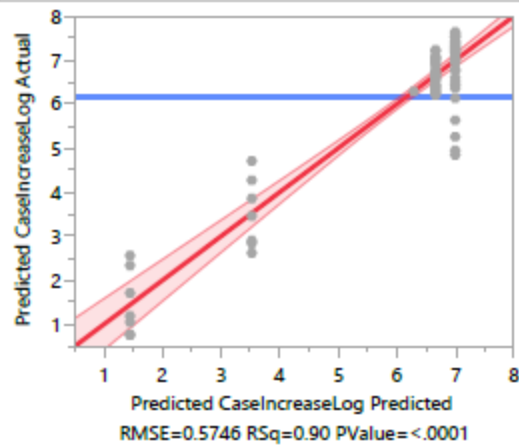
Effect Summary

Source	LogWorth		PValue
NEBusinessClose	22.185		0.00000
EmergDec	11.333		0.00000
SchoolClose	9.364		0.00000
StayAtHome	1.969		0.01073

Residual by Predicted Plot

Studentized Residuals

Response Predicted CaseIncreaseLog

Studentized Residuals

Externally studentized residuals with 95% simultaneous limits (Bonferroni) in red, individual limits in green.

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	6.2908483	0.574636	10.95	<.0001*
EmergDec	-4.843637	0.609494	-7.95	<.0001*
NEBusinessClose	3.1363926	0.237922	13.18	<.0001*
SchoolClose	2.0779818	0.297402	6.99	<.0001*
StayAtHome	0.3356891	0.128891	2.60	0.0107*

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
EmergDec	1	1	20.854059	63.1545	<.0001*
NEBusinessClose	1	1	57.382259	173.7767	<.0001*
SchoolClose	1	1	16.120564	48.8196	<.0001*
StayAtHome	1	1	2.239832	6.7831	0.0107*

Residual Normal Quantile Plot
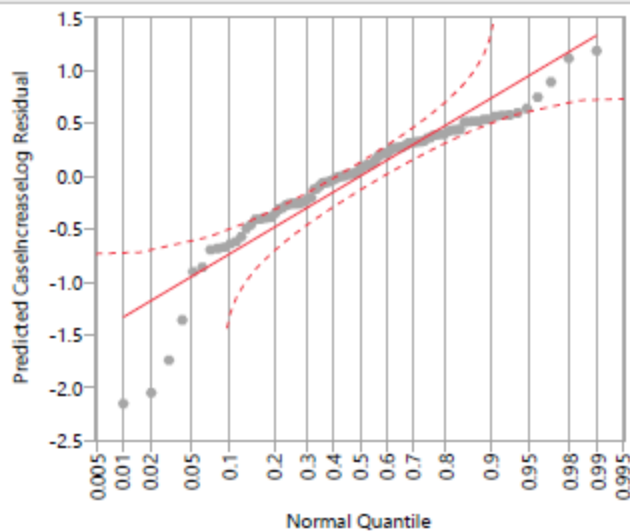
Prediction Profiler
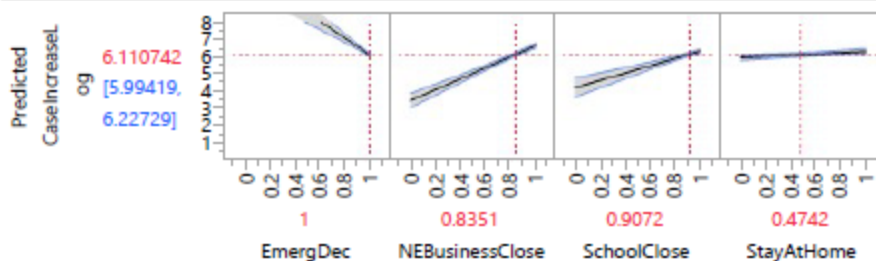
Final Arima Result. In the final Arima result we found that EmergDec, NEBusinessClose, SchoolClose, and StayAtHome policies impacted Pennsylvania's case count the most.