Intervention model to study the effect of the bank's decision by started accepting electronic checks.

Roxanne Salinas & Abeer Ali

STA 6253: Time Series Analysis

Abstract:

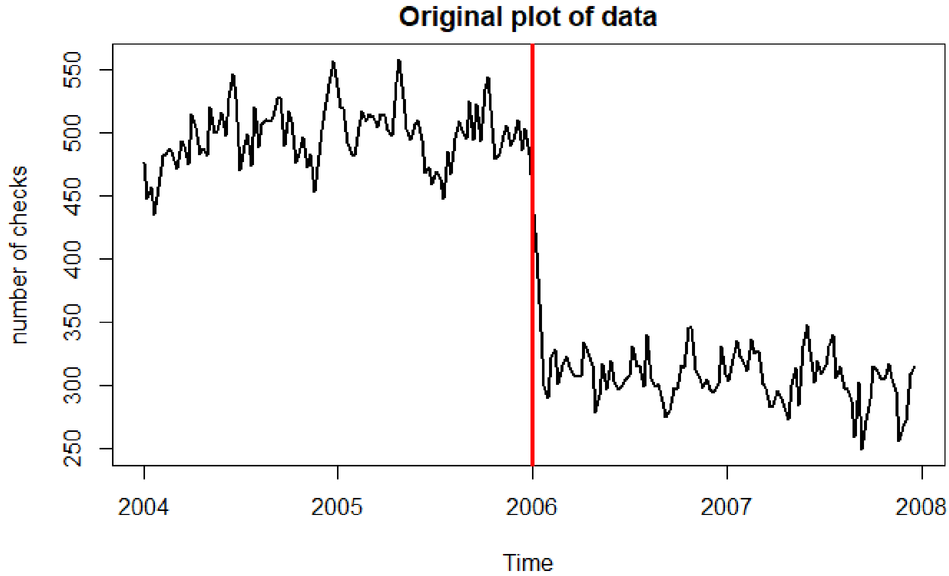
Since computers became accessible to everyone, tasks have come become increasingly more digital. It is no surprise that banks would want to use this technology to their advantage. In our problem, a bank implemented an electronic deposit system, thereby reducing the number of checks cashed at the bank each day. This project is aimed at developing an intervention model that accurately displays the data.

Introduction

Intervention analysis, introduced by Box and Tiao (1975), provides a framework for assessing the effect of an intervention on a time series under study. It is assumed that the intervention affects the process by changing the mean function or trend of a time series. Interventions can be natural or man-made, i.e., the series can be affected by a known event that happens at a specific time, such as

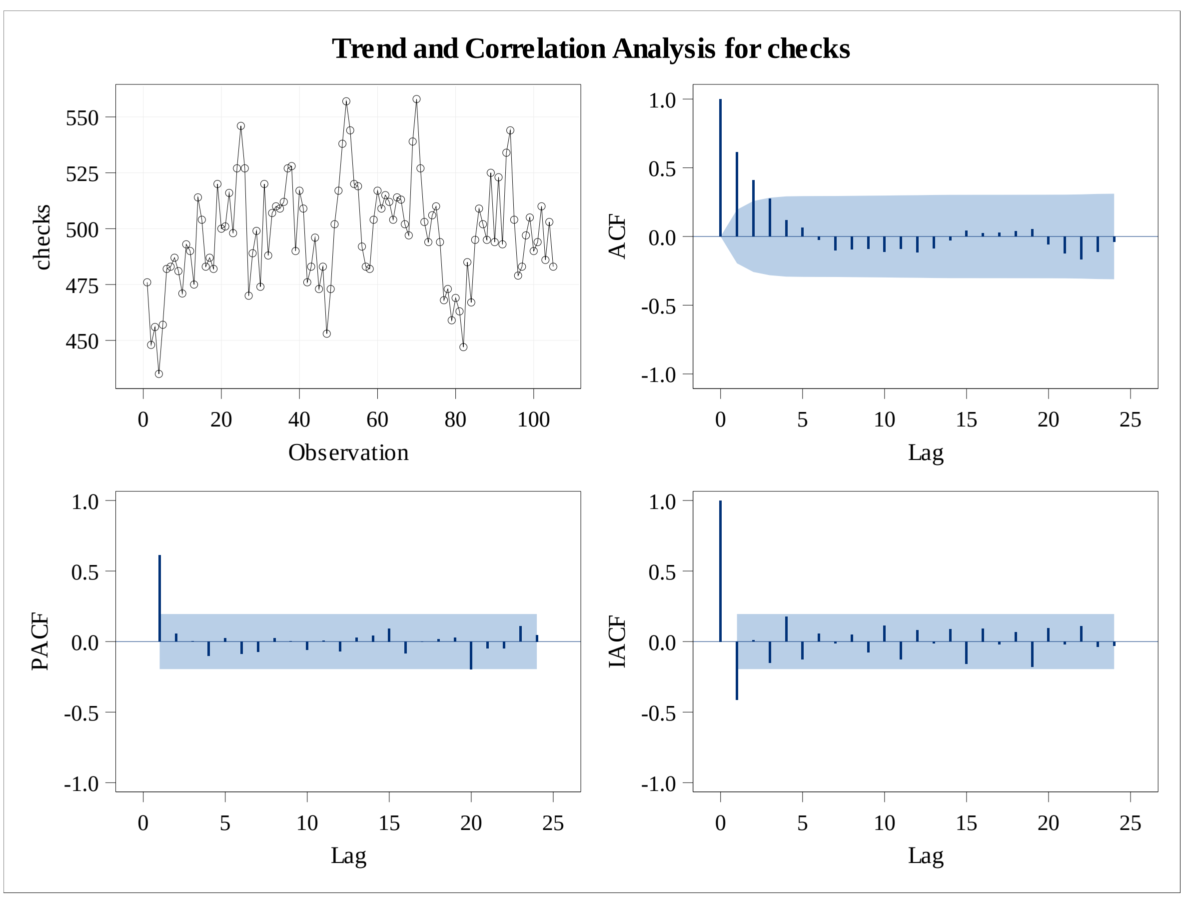
- Fiscal policy changes.
- Introduction of new regulatory laws.
- Switching suppliers.

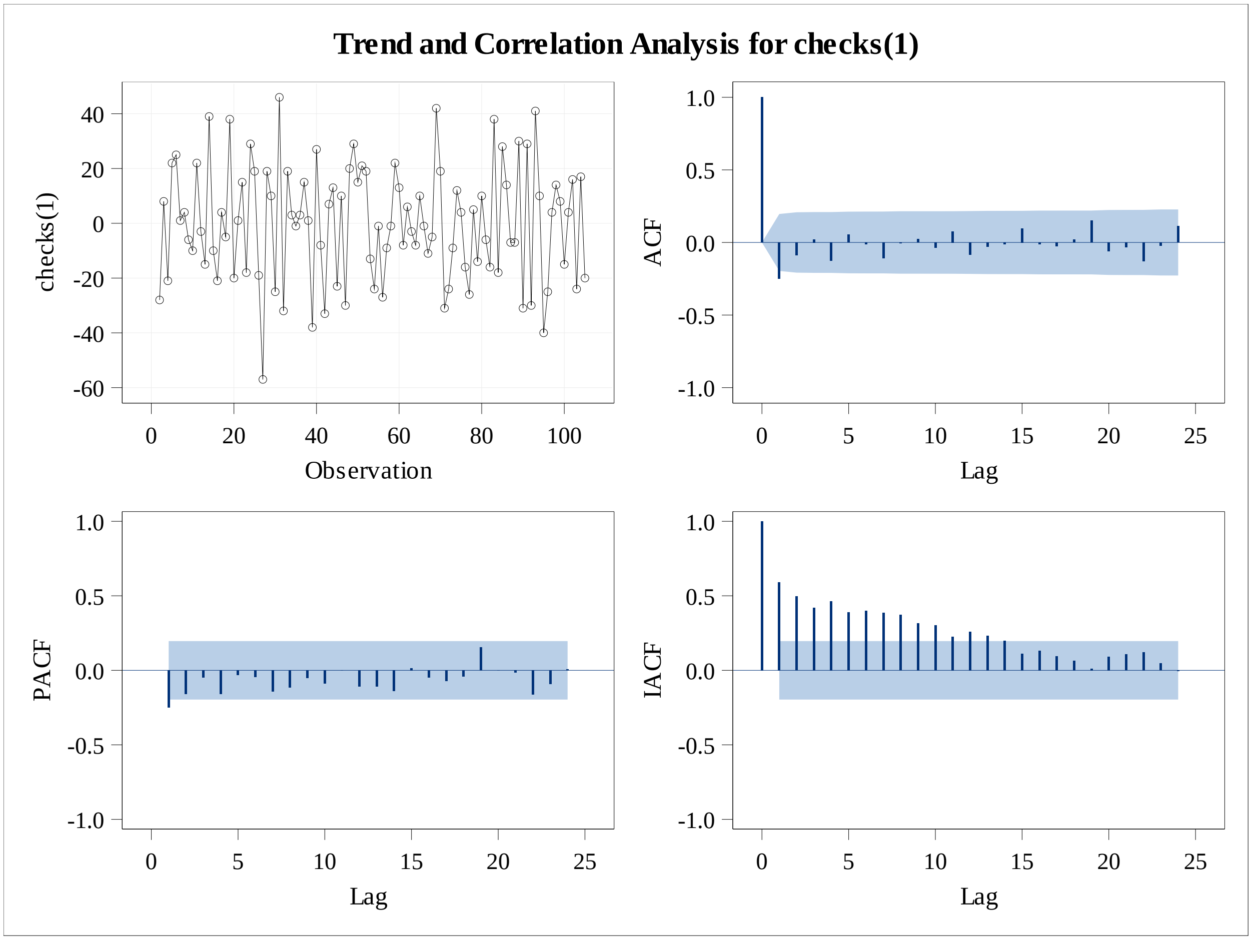
We start by visualizing the data in the plot below. We can see that there is a Sharp decrease in the use of paper checks after 2006, the date the bank began accepting electronic checks, and it does not appear to be a change that does not die off. Based on the plot we will us an intervention model. Furthermore, the checks will continue to be accepted at the bank so we will use a step function to model the data. Our data is from January 2004 to January 2008.



Developing the model:

First, we will to work with pre-intervention data. That is, data prior to year 2006. The plot below shows that the pre-intervention series is not stationary and the ACF is dying slowly. We will attempt to correct this by taking the first difference.





The plot above shows the results after taking the first difference. The auto correlation plot indicates that there may be a $MA(1)$ component. The partial auto-correlation function tells us that an $AR(1)$ component may be needed. Using this information, we compared three different models: $ARIMA(0,1,1)$, $ARIMA(0,1,2)$, and $ARIMA(1,1,1)$. A summary of the three models is shown below.

$ARIMA(0,1,1)$

```
Call:
arima(x = ts.pre.intervention, order = c(0, 1, 1))

Coefficients:
          ma1
      -0.3608
s.e.   0.1127

sigma^2 estimated as 397.2:  log likelihood = -458.83,  aic = 919.66

Training set error measures:
test elements must be within sample              ME RMSE MAE MPE MAPE
Training set NaN  NaN NaN NaN  NaN

z test of coefficients:

    Estimate Std. Error z value Pr(>|z|)
ma1 -0.36083    0.11270 -3.2016 0.001367 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$ARIMA(0,1,2)$

```
Call:
arima(x = ts.pre.intervention, order = c(0, 1, 2))

Coefficients:
          ma1      ma2
      -0.3491  -0.1978
s.e.   0.1013   0.1325

sigma^2 estimated as 387.4:  log likelihood = -457.61,  aic = 919.21

Training set error measures:
test elements must be within sample              ME RMSE MAE MPE MAPE
Training set NaN  NaN NaN NaN  NaN

z test of coefficients:

    Estimate Std. Error z value  Pr(>|z|)
ma1 -0.34913    0.10133 -3.4456 0.0005699 ***
ma2 -0.19780    0.13248 -1.4931 0.1354203
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$ARIMA(1,1,1)$

```
Call:
arima(x = ts.pre.intervention, order = c(1, 1, 1))

Coefficients:
         ar1      ma1
      0.6299  -1.0000
s.e.  0.0784   0.0605

sigma^2 estimated as 354.3:  log likelihood = -454.42,  aic = 912.84

Training set error measures:
test elements must be within sample              ME RMSE MAE MPE MAPE
Training set NaN  NaN NaN NaN  NaN

z test of coefficients:

     Estimate Std. Error  z value  Pr(>|z|)
ar1  0.629863   0.078369   8.0372 9.193e-16 ***
ma1 -0.999950   0.060493 -16.5300 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

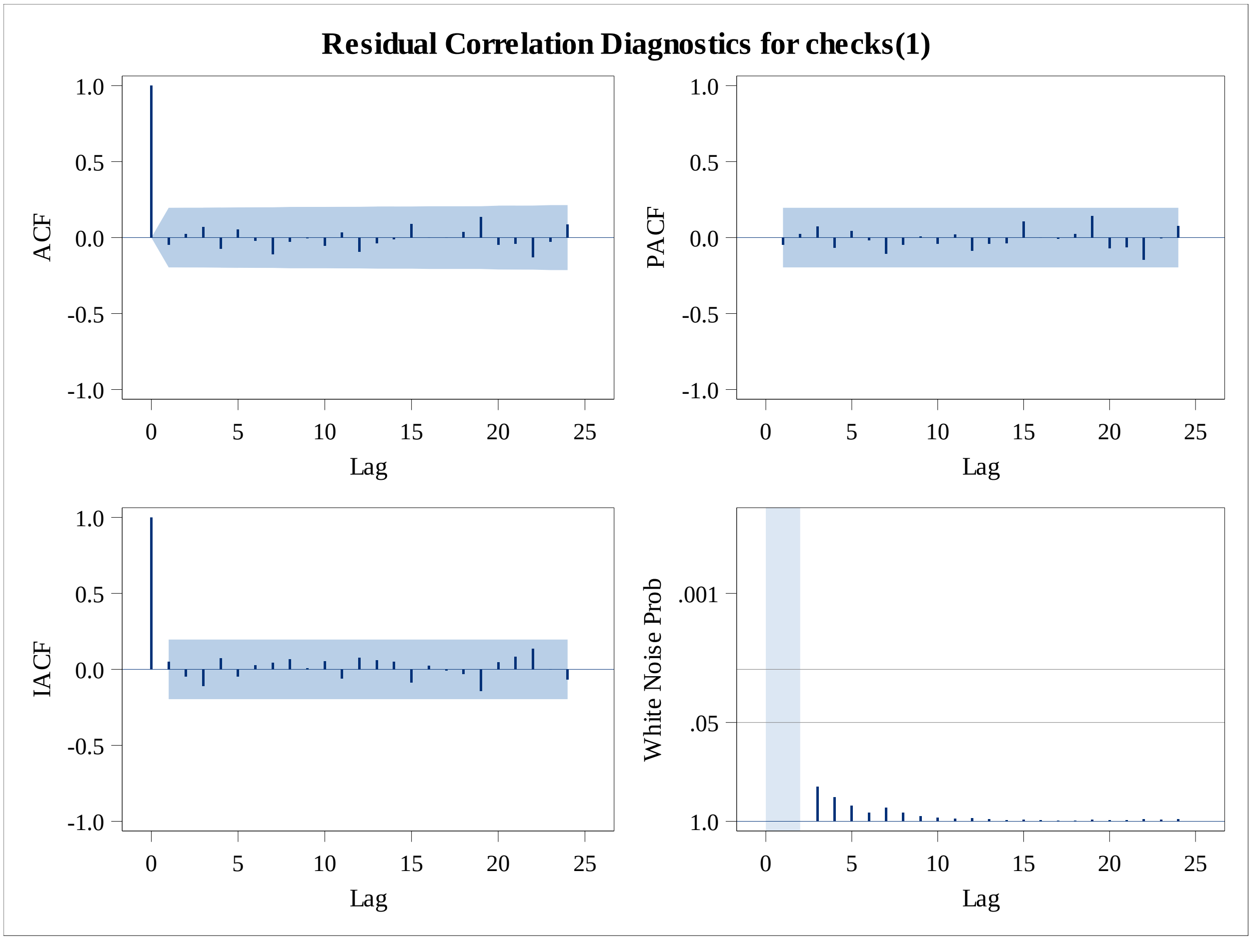
Based on this information, the second model was immediately rejected because the MA(2) coefficient was insignificant. In the first model, the MA component is significant at the .05 level and the $ARIMA(1,1,1)$ model has both the AR component significant and the MA component significant. We decided to use the $ARIMA(1,1,1)$ model for our data since it has a lower AIC than the MA(1) model.

The $ARIMA(1,1,1)$ model is shown below. We can see from the table above that the model parameters are significant.

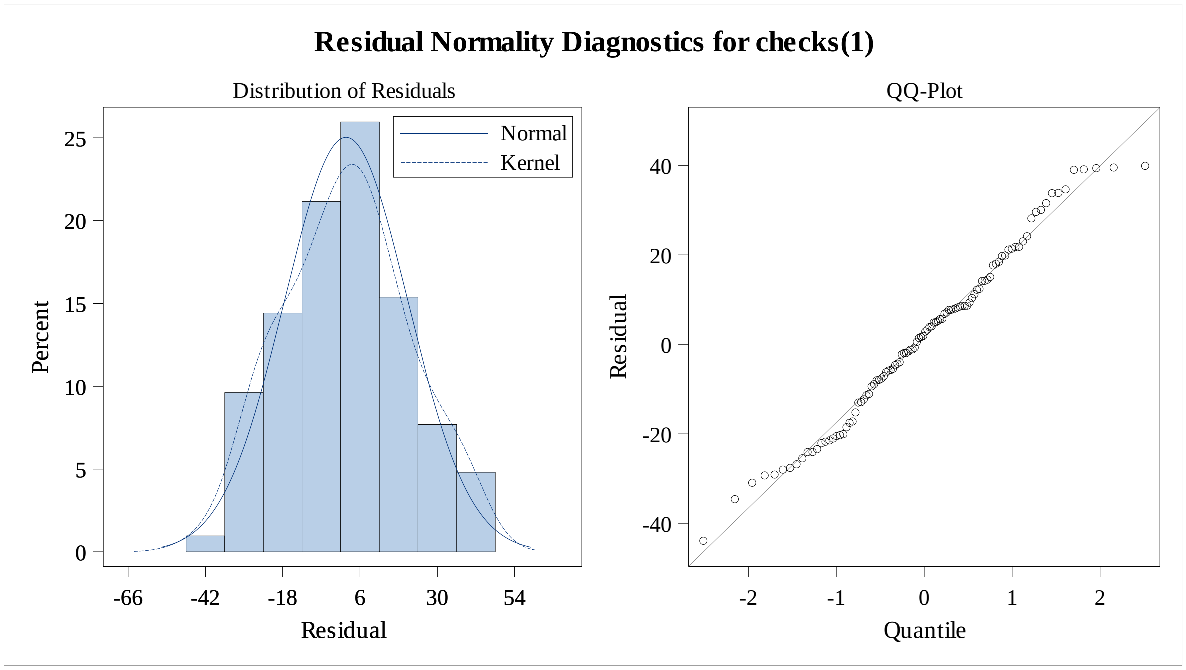
Conditional Least Squares Estimation					
Parameter	**Estimate**	**Standard Error**	**t Value**	**Approx Pr > |t|**	**Lag**
MA1,1	0.95088	0.03706	25.66	<.0001	1
AR1,1	0.59586	0.09702	6.14	<.0001	1

Autocorrelation Check of Residuals									
To Lag	**Chi-Square**	**DF**	**Pr > ChiSq**	**Autocorrelations**					
6	1.89	4	0.7555	-0.038	0.035	0.082	-0.062	0.063	-0.010
12	4.37	10	0.9293	-0.101	-0.018	0.006	-0.043	0.043	-0.084
18	5.92	16	0.9889	-0.029	-0.005	0.098	0.010	0.005	0.043
24	12.29	22	0.9511	0.142	-0.043	-0.039	-0.126	-0.027	0.089

We can see from the table above that the residuals are uncorrelated which indicates they are white noise.



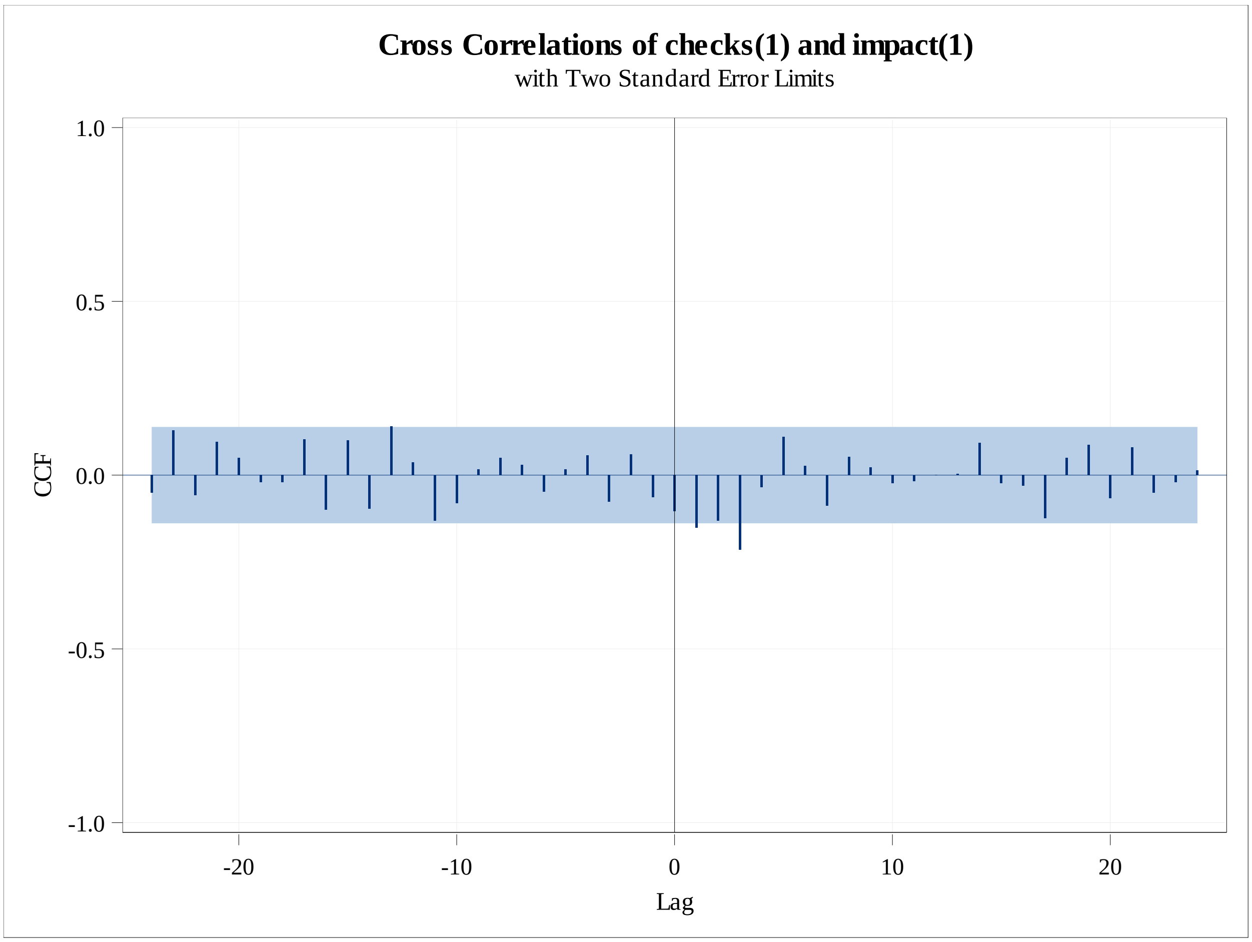
From the plot above, we can see that the residuals ACF and PACF don't have any significant spikes. Also, below we see no significant spikes in the white noise probability plot. This also confirms the residuals are likely to be white noise.



Based on the histogram and QQ plot, the normality assumption of the residuals looks reasonable here. This indicates that $ARIMA(1,1,1)$ is a good fit and can be considered as the tentative model.

Modeling the impact event

Next, we create an impact variable which is equal to zero for year before 2006 and equal to 1 afterwards. The CCF plot is shown below:



This plot makes sense from the data. There is a clear step change that is immediately felt, and then the series returns to acting normally.

Therefore, the form of our tentative intervention model is:

$$(1-B)check_t = \omega_0 Impact + \frac{(1-\theta B)}{(1-\phi B)}\mathrm{a}_t$$

$$a_t \sim WN(0, \sigma_\alpha^2)$$

The policy at the bank was placed into effect in 2006 and was kept from then on. This describes step function. We will define the step function as follows:

$$X_t = \begin{cases} 0 \;\; for\ t < Jan. 2006 \\ 1\ for\ t \geq Jan. 2006 \end{cases}$$

We used X_t to select an appropriate model.

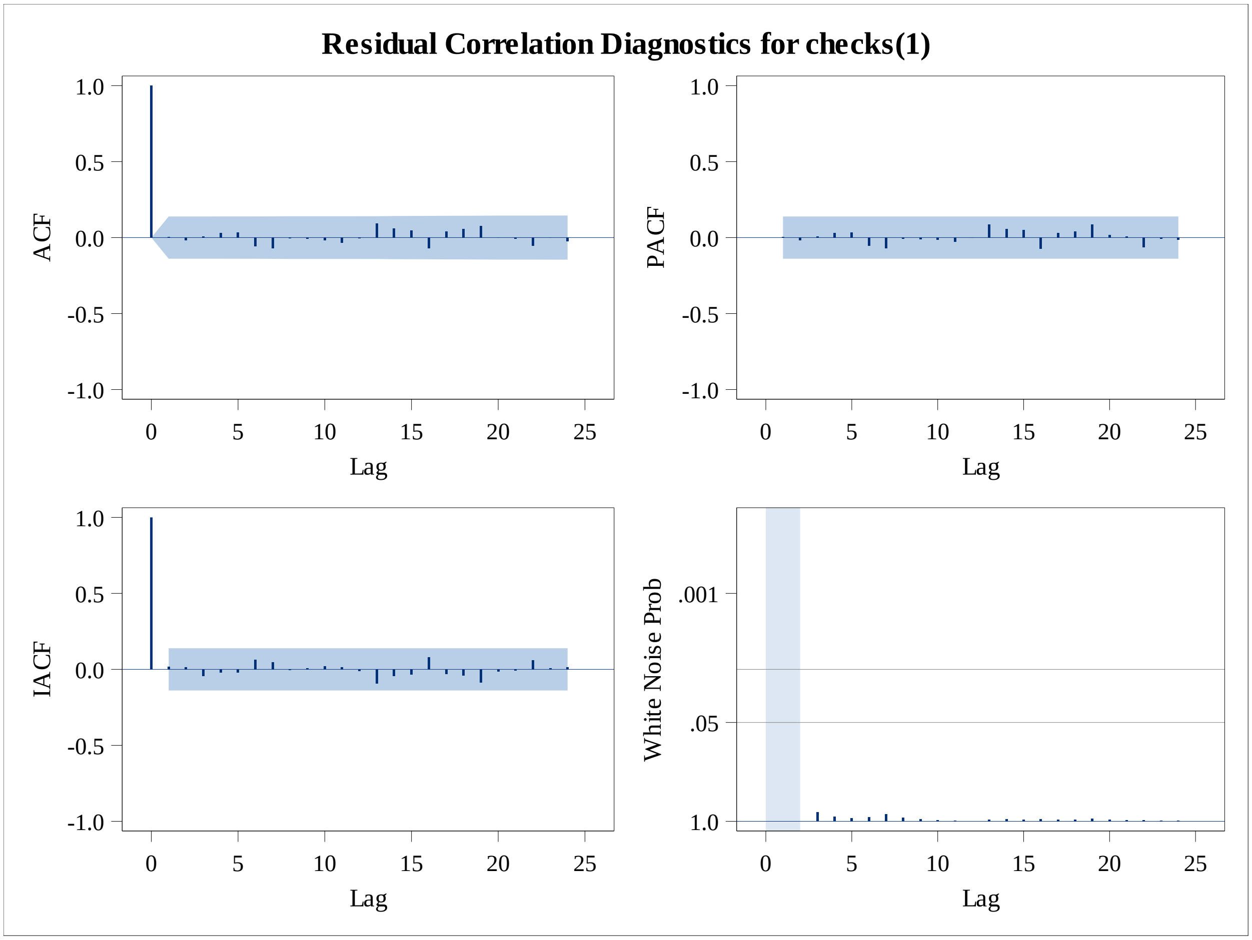
The estimated parameters are shown in the table below:

Conditional Least Squares Estimation							
Parameter	**Estimate**	**Standard Error**	**t Value**	**Approx Pr > |t|**	**Lag**	**Varia ble**	**Shift**
MU	-0.08043	0.32059	-0.25	0.8022	0	checks	0
MA1,1	0.90412	0.04345	20.81	<.0001	1	checks	0
AR1,1	0.56199	0.08420	6.67	<.0001	1	checks	0
NUM1	-63.11865	18.61030	-3.39	0.0008	0	impact	0
NUM1,1	86.25845	18.58923	4.64	<.0001	1	impact	0

Therefore, our model is:

$$\hat{Y}_t = -0.08 + (-63.119 - 86.258B)X_t + \frac{(1-0.90412B)}{(1-056199B)}\mathrm{a}_t$$

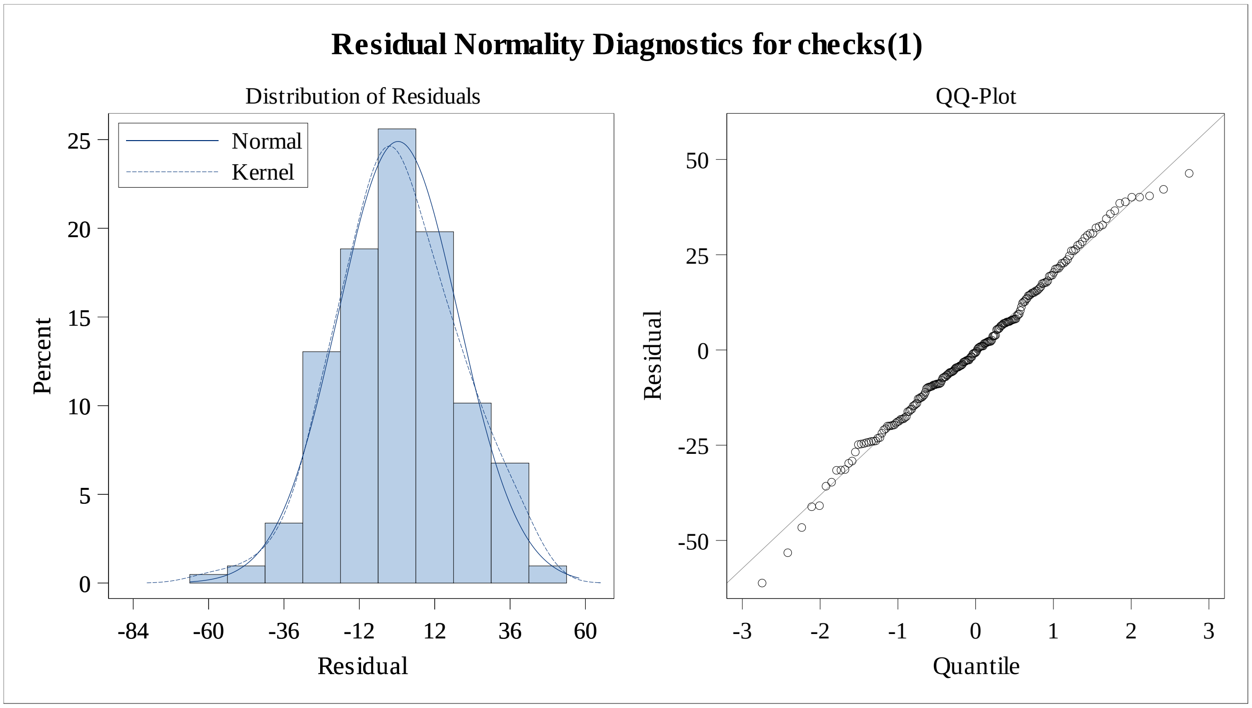
We must now check the assumptions on this model to make sure this is a good fit. From the plot below, we can see that the residuals ACF and PACF don’t have any significant spikes. Also, we see no significant spikes in the white noise probability plot.



Autocorrelation Check of Residuals									
To Lag	**Chi-Square**	**DF**	**Pr > ChiSq**	**Autocorrelations**					
6	1.23	4	0.8737	0.005	-0.019	0.009	0.030	0.035	-0.056
12	2.69	10	0.9878	-0.071	-0.006	-0.010	-0.017	-0.035	-0.005
18	8.15	16	0.9442	0.094	0.059	0.046	-0.071	0.041	0.056
24	10.39	22	0.9824	0.077	-0.003	-0.009	-0.054	-0.000	-0.026
30	14.19	28	0.9858	-0.091	0.029	-0.037	-0.021	0.065	-0.028
36	17.00	34	0.9934	0.046	0.083	0.027	0.022	-0.032	-0.013

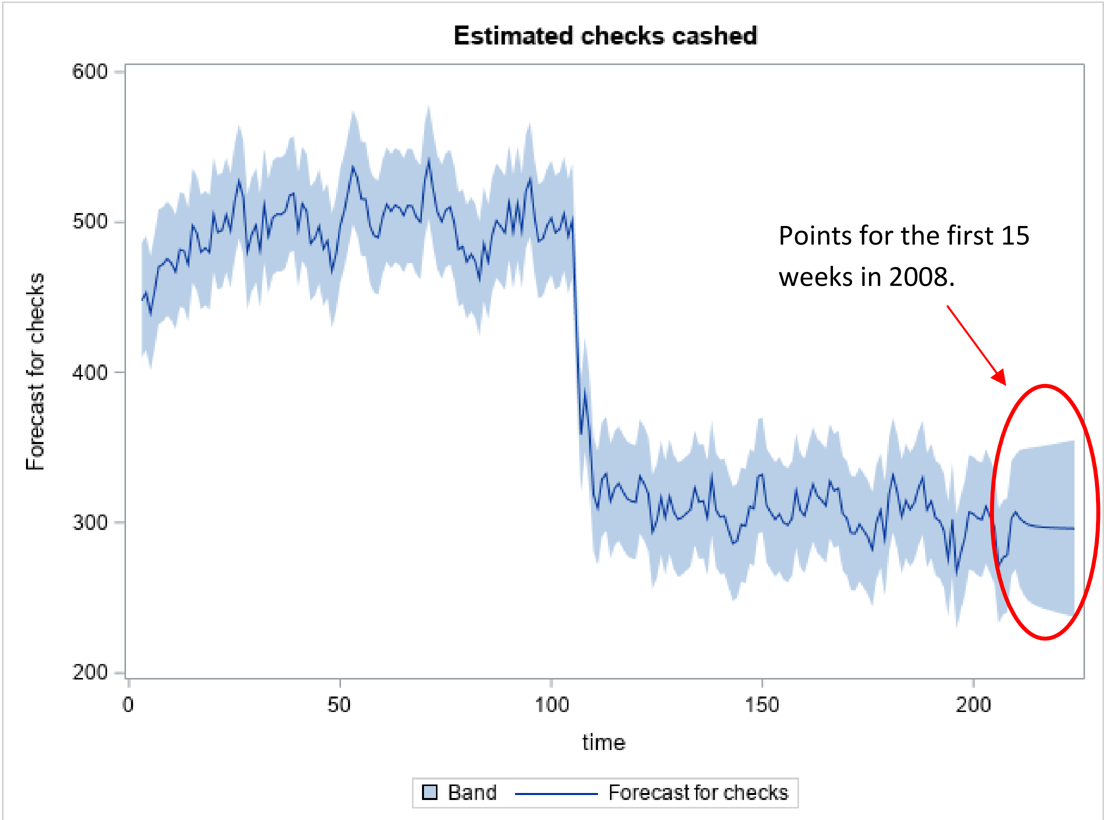
We can see from the table above that the residuals are uncorrelated which means they are white noise. There is no significance in the autocorrelations.

The normality assumption of the residuals looks reasonable here based on the histogram and qq plot. Therefore, our model is a good fit, and will be used for forecasting purposes.



Forecasting

Using our model, we forecasted the estimated number of checks cashed for the next 15 weeks of 2008 (weeks 210 to 224). We can see from the forecasts that as time increases, there is expected to be a slight decrease in the number of checks cashed. The graph confirms this with a slight decreasing pattern at the end.


Estimated checks cashed
Forecast for checks
600
500
400
300
200
0
50
100
150
200
time
Band
Forecast for checks
Points for the first 15 weeks in 2008.

Forecasts for variable checks				
Obs	**Forecast**	**Std Error**	**95% Confidence Limits**	
210	307.1970	19.4211	269.1324	345.2615
211	302.7765	23.2469	257.2134	348.3396
212	300.2570	24.9436	251.3684	349.1456
213	298.8058	25.8921	248.0583	349.5534
214	297.9550	26.5260	245.9650	349.9451
215	297.4417	27.0121	244.4988	350.3845
216	297.1180	27.4230	243.3698	350.8661
217	296.9008	27.7930	242.4275	351.3740
218	296.7435	28.1395	241.5912	351.8959
219	296.6199	28.4716	240.8166	352.4232
220	296.5152	28.7943	240.0794	352.9510
221	296.4211	29.1104	239.3659	353.4764
222	296.3330	29.4213	238.6683	353.9978
223	296.2483	29.7281	237.9823	354.5143
224	296.1654	30.0312	237.3053	355.0255

Conclusion:

Electronic checks had a permanent impact on the number of paper checks being cashed. Our model not only shows there was an anticipation effect, but it also indicates the number of paper checks being cashed will continue to decrease slowly. With the way technology has evolved over the last few years, it would not be surprising if another event nearly eliminates paper checks all together.