# Nonparametric Regression of Presidential Approval Rating with Correlated Observations

**Andrew V. Carter**

*Department of Statistics and Applied Probability*
*University of California Santa Barbara*
*Santa Barbara, CA 93106-3110*
*e-mail:* carter@pstat.ucsb.edu

**and**

**Lilian Lu**

*e-mail:* luyuying0128@gmail.com

**and**

**Yulei Yuan**

*e-mail:* inorysmw@gmail.com

**Abstract:** The Presidential Approval Rating is the result of a consistent series of surveys performed by Gallup Inc. It provides an interesting example of a nonparametric regression problem where there is a true unknown population value of the parameter of interest. Unfortunately, the way the data series is reported results in correlated observations. We explore methods for finding a bandwidth in a Nadaraya–Watson kernel estimator that are robust to local positive correlation between observations. We apply these results to some inferential questions regarding the popularity of President Barack Obama.

**MSC 2010 subject classifications:** Primary 62G08; secondary 62M10.
**Keywords and phrases:** nonparametric regression, kernel estimator, correlated errors.

## 1. Introduction

The presidential approval rating is ideally the proportion of the population that approves of the job that the president is doing. We are interested in exploring how this proportion changes over time. The data we decided to use comes from Gallup Inc. (Gallup, 2016) There are a number of relevant issues that might arise in survey data like this. What exact question is being asked? What is the population? What effects are there from differential nonresponse? By focusing on a data series from a single sources, we hope to avoid these concerns because the methodology is consistent from day to day. The specific data series we used was aggregated by Peters and Wolley (2017).

Investigating this series is also appealing because it is easy to believe that at any instant in time there is a true unknown population proportion. Our assumption is that typically people change their minds on a question like this only slowly over a long period of time (months). The exception may be in the case of a historical event that changes people's perceptions over night. There is clear evidence of this in President Bush's approval rating around September 11, 2001.

From this dataset, we consider two possible questions of scientific interest. Does the approval rate ever go below 40 percent? The other question is, how much do historical events affect on the approval rating? In particular, how did the approval rating for President Obama change during the week when Osama bin Laden was killed?

Since interviewees were answering a yes-or-no question, the data follows a binomial distribution. However, because the sample sizes are reasonably large, we decided to appeal to the Central Limit Theorem and treat observations as normally distributed. An advantage in choosing to work on a normal model instead of a binomial one is more flexibility in estimating the variance. If there exists correlation between observations, variance estimation will become an issue. Thus, our model is

$$y_t = \mu(x_t) + \sigma\epsilon_t$$

where $y_t$ represents survey results and $x_t$ represents dates. The population mean, $\mu(x_t)$ is what we try to estimate through kernel methods. $\sigma$ is normal standard deviation and $\epsilon_t$ are normally distributed errors. In our dataset, these errors are possibly correlated.

This correlation across error terms is the main data analysis challenge that this Gallup data poses. From the description at Peters and Wolley (2017) and Gallup (2016), the reported percentages are actually already the result of a three-day moving average of each night's survey results. While a kernel estimator may give a reasonable estimate, correlated errors greatly complicate the choice of bandwidth (see Altman (1990) and Hart (1991).) Opsomer, Wang and Yang (2001) reviews a number of techniques that can be used for these types of problems. For kernel estimators in particular, Chiu (1989) used an updated version of the Mallow's $C_p$ criterion based on an estimate of the spectral density of the errors. In this vein, we explore updates to the typical cross validation or prediction error techniques proposed by Altman (1990) which are a function of the correlations between observations. We also consider the partitioning techniques of Chu and Marron (1991), and a direct method of estimating the times series components in the model.

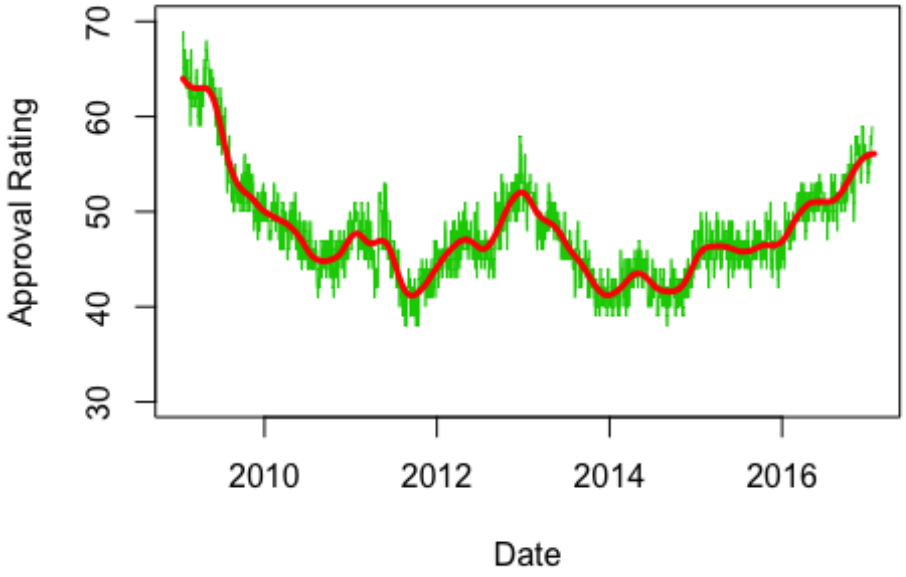


FIG 1. *Estimate of p using bandwidth 80 along with the observations of the original data.*

## 2. Kernel Estimation

Our approach to nonparametric regression will be to use kernel weighting functions. For our estimation, a quartic kernel will be applied of the form (1):

$$K_h(t) = \begin{cases} \left(1 - (\frac{t}{h})^2\right)^2 & \left|\frac{t}{h}\right| < 1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where $h$ is the bandwidth of this kernel function.

By inspecting the scatter plot of the data, a bandwidth of 80 seems a reasonable choice for the bandwidth. Figure 1 shows the estimate of the approval rating using a bandwidth of 80 with the quartic kernel. This is a reasonable fit for this dataset. it is not over smoothing the data while keeping all the features within the trend of the data.

### *2.1. Basic cross validation assuming iid errors*

Determining a good smoothing parameter, or in other words, a bandwidth, is crucial for many non-parametric techniques. Cross validation is one of the main methods that are used for determining reasonable values of the bandwidth from the data, when the errors are identically and independently distributed (i.i.d.). Therefore, for this data set, we first apply basic cross validation to estimate the best bandwidth, assuming independent errors. The code can be found in the appendix.

Cross validation minimizes the criteria function

$$CV = \frac{1}{n}\sum_{t=1}^{n}\frac{(y_t - \hat{\mu}_t)^2}{(1 - S(t))^2} \tag{2}$$

where $n$ is the sample size, $(y_t - \hat{\mu})^2$ is the residual for observation at time $t$, and $S(t)$ is the diagonal element of the influence matrix. In principle, $S(t)$ is analogous to the number of surveys included in the average. Specifically,

$$S(t) = \frac{1}{\sum_{t=1}^{n} K_h(t^* - t)} \tag{3}$$

We pick the bandwidth $h$ that minimizes this criterion.

Using this method, a bandwidth of 3 is obtained. Figure 2 shows the estimation of the data using bandwidth $h = 3$. Clearly the estimate is so rough that a much larger bandwidth is needed to add more smoothness. It means that basic cross validation is not the most appropriate approach to find the best bandwidth for this dataset. See section 2.3 for further discussion.

### *2.2. Basic Mallows's $C_p$*

The second method of finding the optimal bandwidth for our data is Mallows's $C_p$. Mallows's $C_p$, which is the "estimator of expected squared prediction error (ESPE) based on a correction to the observed squared residuals"(Altman, 1990) is defined as

$$C_p = \sum_{t=1}^{n} r_t^2 + 2\sum_{t=1}^{n} S(t)\hat{\sigma}^2 \tag{4}$$

(Mallows, 1973). In this formula, $r_t^2$ is the squared residual for each observation and $S(t)$ is the parameter which determines the number of surveys will be counted in the kernel function. Here, $S(t)$ is defined at previous section and $\hat{\sigma}^2$ is the unbiased estimator of population variance.

When we are assuming independent errors, using equation (4) shows that the optimal bandwidth will be 7, which is slightly larger than the result of cross validation. The corresponding bandwidth selection plot is figure 3 while the plot comparing bandwidth as 7 and 3 is figure 2. We can see there is no significant difference between $h = 7$ and $h = 3$. Overall, when these two methods are under

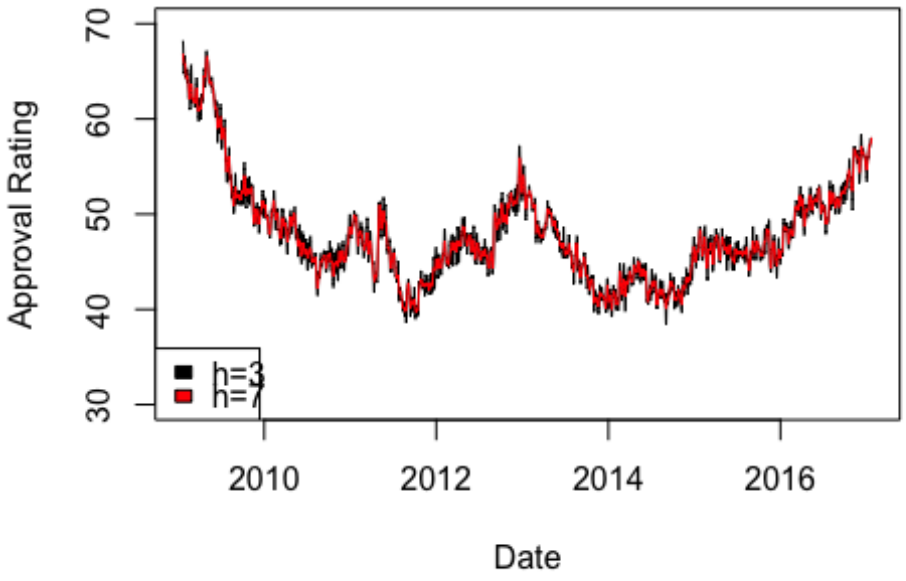


FIG 2. *Estimation of the approval rating using a bandwidth of 3(from basic cross validation) and 7(from Mallows's $C_p$)*

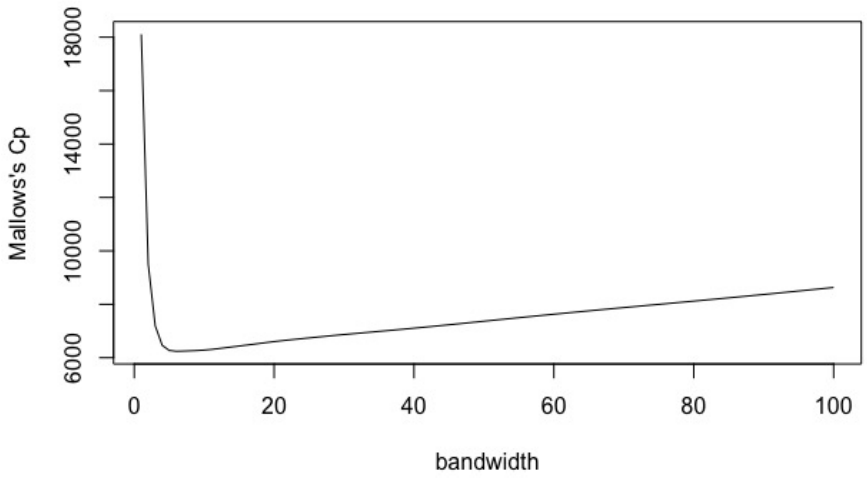


FIG 3. *Using Mallows's $C_p$ to find the optimal bandwidth from 1 to 100. The minimum point happens x=7.*

the assumption that errors are identical and independently distributed, cross validation and Mallows's $C_p$ give smaller bandwidth than we would consider reasonable. Therefore, we must consider the case where errors might be not i.i.d.

### *2.3. Problems with assuming independence*

The reason for the heavy underestimation of the bandwidth is the positive correlation of the data, as Altman (1990) has claimed that "the standard techniques for bandwidth selection, such as cross validation and generalized cross-validation, are shown to perform very badly when the errors are correlated." In a sense, positive correlation between successive observations is interpreted as signal by these automatic techniques. Therefore, we should consider alternatives to basic cross validation or basic Mallows's $C_p$ in order to minimize the effect of correlation on bandwidth selection.

## 3. Partitioned Estimate

Since our data is likely correlated across subsequent days, a partitioned estimate is a method that can deal with the existence of correlation. We assume that if two observations get further away in time then the correlation between these two

will be diminished. Specifically speaking, in our case, we divide our observations into 7 nearly equal-sized samples, which means that in each group, the gap between observations is 7 days. We expect that surveys taken 7 days apart will have minimal correlation, and it is convenient to think of these series as weekly estimates. After dividing our data, we could do an analysis then with nearly uncorrelated observations within each group. We will use Mallows's $C_p$ to calculate the best bandwidth.

### *3.1. Mallows's $C_p$*

In order to select a proper bandwidth, we apply Mallows's $C_p$ during this section. In partitioned estimate, the strategy of variance estimation comes from Rice (1984) who calculated $\hat{\sigma}^2$ through

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n-1}(y_i - y_{i+1})^2}{2(n-1)}. \tag{5}$$

This formula can calculate the nearly unbiased variance estimation if the mean does not change dramatically between consecutive observations. It has the assumption of independent observations and a smooth mean function. Based on this formula, Tecuapetla-Gómez and Munk (2017) discussed the variance estimation in difference-based covariance estimation that

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n-h}(y_i - y_{i+h})^2}{2(n-h)} \tag{6}$$

with $h$ as the gap for eliminating correlation. In Rice's formula, he used $h = 1$ because he was interested in consecutive observations and assumed observations independent. However, Tecuapetla-Gómez and Munk (2017) suggested that, using formula (6), with $h > 1$, we can get an unbiased variance estimate. In our case, we choose $h = 7$ because we believe that gap is large enough to eliminate the correlation. We calculate $\hat{\sigma}^2 = 2.274$. If we use this result in equation (4) to select the best bandwidth from 1 to 300, we would get figure 4. The bandwidth that achieves the minimum is $h = 5.4$ weeks which is 37.8 days.

Figure 5 shows separate estimates from each of the seven series using our best bandwidth. According to the plot, each group has a similar yet distinguishing estimation. Figure 6 shows the estimate on all of the data when the bandwidth is 37.8 days.

It is true that partitioned estimate is effective in removing the effect of correlation. However, it might not be the best method to apply for our dataset. According to Chu and Marron (1991), using the partition estimated we will end up with a poor bandwidth. As their paper states, "... the asymptotic mean of this bandwidth (of partitioned estimate) revels that there is a significant distance between the partitioned cross-validated bandwidth and the optimal bandwidth which minimizes the mean average squared error."(Chu and Marron, 1991). After dividing into 7 groups, within each group the observations are

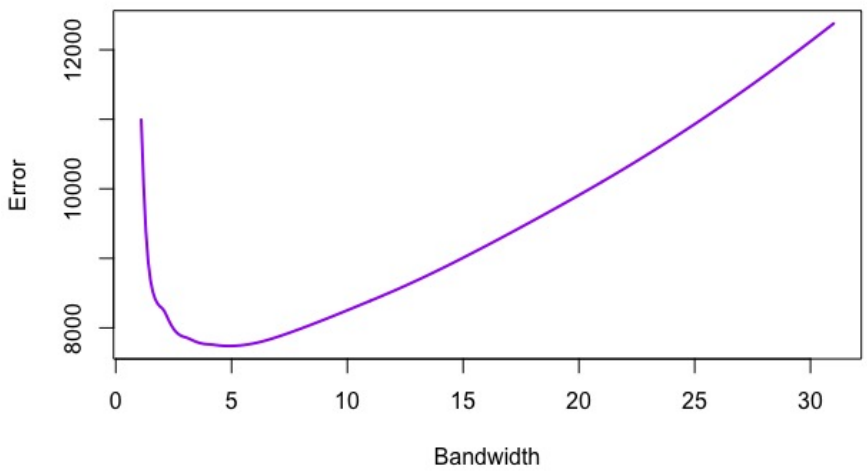


FIG 4. *Selecting a bandwidth in partition estimate from 1 to 300. The minimum happens at x=5.4. The optimal bandwidth is 5.4 weeks.*

nearly independent. However, we are not certain about the correlation across the groups, and if we would like to use the results from the partitioned estimate it is a difficult question about how to aggregate the partitioned estimates. The removal of dependence through a partitioned estimate is relatively effective, but the corresponding results are not our intended ones. The best bandwidth from partitioned estimate is not the optimal bandwidth when we are analyzing all observations together.

### *3.2. Variance Estimation*

Taking a look at the Mallows's $C_p$ method of choosing a bandwidth, it is clear that the value of $\hat{\sigma}^2$ affects the resulting bandwidth. Section 3.1 has already mentioned the variance estimation in partitioned estimate. However, since partitioned estimate is not our best choice while dealing with correlated observations and the variance estimation above works only in partitioned estimate, we use another formula for $\hat{\sigma}^2$ that

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^{n} r^2}{n - \sum_{t=1}^{n} S(t)}. \tag{7}$$

The result is 2.081. We believe the smaller value of $\hat{\sigma}^2$ is caused by the positive correlation of the data. Unlike partitioned estimate, since we do not process

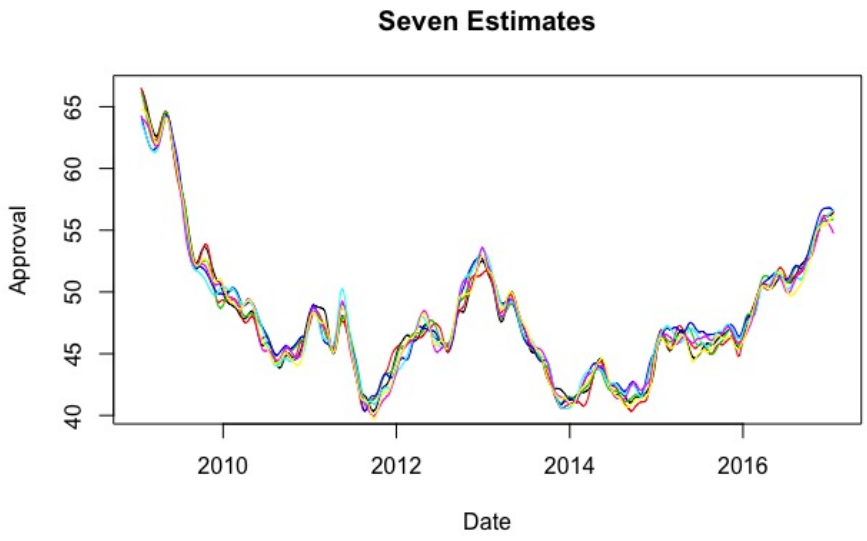


FIG 5. *Seven estimations of the approval rating from the partition data. The bandwidth is h=5.4 weeks.*

the data and get rid of the dependence, the existence of correlation affects the value of $\hat{\sigma}^2$. In this calculation, it neglects the positive correlation and causes the result to be smaller than the one of partitioned estimate.

Using this result and formula (4), we try to find out the best bandwidth from 1 to 100. According to the calculation,we get $h = 39$ as the best bandwidth.

Generally speaking, we have two different variance estimations and different corresponding bandwidth selection. The difference of results is caused by the existence of correlation. When we used equation (7) to calculate variance estimation, there is an underlying assumption that observations are independent. However, based on their collected method, our observations are not independent. Since we believe every three observations are correlated, the result based on the assumption of independence is inaccurate. On the other hand, when we used partitioned estimation, dividing data into 7 groups makes observations within each group share weak correlation and we can consider them as independent. Without considering the validity of partitioned estimate but only the accuracy of the variance estimation, calculating $\hat{\sigma}^2$ of partitioned estimate requires fewer assumption and should give a more accurate answer.

## 4. Correlation Model

Typically, the correlation function is unknown and it has to be estimated from the data. However, the collection process of our data provides a specific corre-

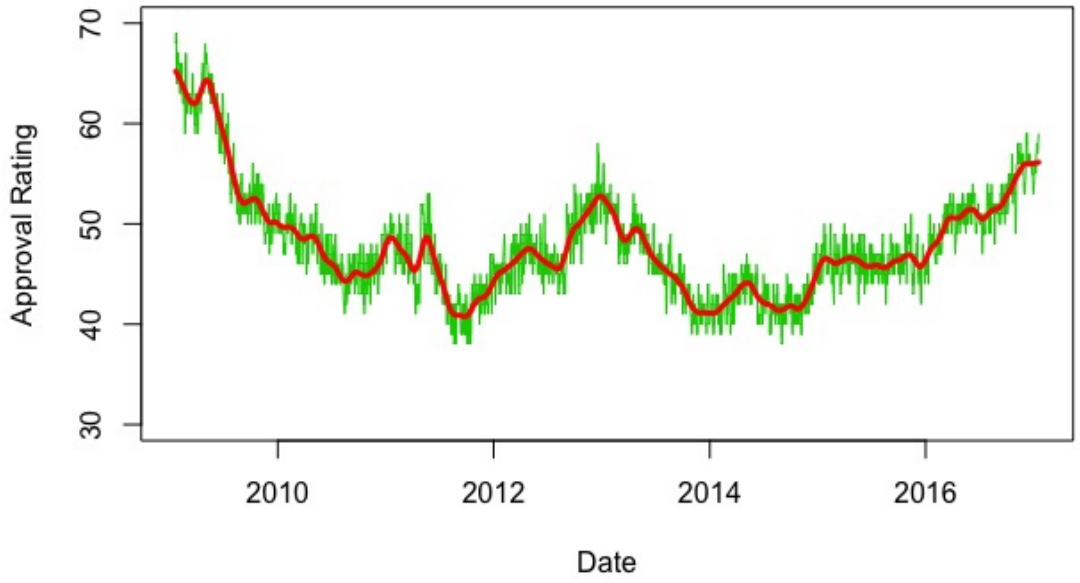


FIG 6. *Using the result from the partitioned estimate to plot the estimate with* $h = 37.8$ *on the original data.*

lation model. From the survey process description Gallup (2016), the data is reported using a 3-day moving average window. Thus it is pretty reasonable to apply an ARMA time series model or more simply an MA(2) model with the form

$$\epsilon_t = z_t + z_{t-1} + z_{t-2} \tag{8}$$

where $z_t \sim \mathcal{N}(0, \sigma^2)$ are uncorrelated.

In a general MA(2) model,

$$Var(y_t) = \sigma^2(1 + \theta_1^2 + \theta_2^2) \tag{9}$$

In our MA(2) model $\theta_1 = \theta_2 = 1$, and the variance is $3\sigma^2$. The autocorrelation is

$$\rho_1 = \frac{\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2},\ \rho_2 = \frac{\theta_2}{1 + \theta_1^2 + \theta_2^2},\ \text{and } \rho_h = 0 \text{ for } h \geq 3 \tag{10}$$

Then we have $\rho(0)$=1, $\rho(1)$=$\frac{2}{3}$, $\rho(2)$=$\frac{1}{3}$ ,$\rho(-1)$=$\frac{2}{3}$, $\rho(-2)$=$\frac{1}{3}$.

Altman (1990) argues that we should update our influence function $S_t^*$ for

the presence of correlated errors

$$S_t^* = \sum_{j=-h}^{h} K_h(t+j)\rho(j). \tag{11}$$

We could also have used MA(2) parameters estimated from the residuals which would give a total correlation of 2.94. The value is smaller than our 3 calculated above because the likelihood estimates for the MA(2) model are constrained to yield an invertible model. Here, we choose to use the $\theta_1 = \theta_2 = 1$ because it fits the mechanisms described and will generally lead to a larger bandwidth.

### *4.1. Modified Cross-validation*

An alternative way to estimate bandwidth using Cross-validation for dependent error from a stationary correlated process is also suggested by Altman (1990), in which the modified Cross-validation is estimated as

$$r_{CV}^2(t) = \frac{(y_t - \hat{\mu}_t)^2}{(1 - S_t^*)^2} \tag{12}$$

where $S_t^*$ was calculated (11).

Figure 7 shows that the the best bandwidth value is 61. Figure 8 shows the estimate of the approval rating when choosing the bandwidth equals 61.

What Figure 8 shows clearly denotes that this method would significantly reduce the effect of correlation. However, this might still be an underestimation of the bandwidth because as we demonstrated before, a bandwidth of 80 (Figure 1) shows a better graph without over-smoothness. Maybe more techniques are required to reduce of the effect of correlation for this dataset.

### *4.2. Modified Mallows's $C_p$*

Altman (1990) has brought up an appropriate adjustment for Mallows's $C_p$ when we have known correlation function. The formula is

$$C_p = \sum_{t=1}^{n} r_t^2 + 2\sum_{t=1}^{n} S_t^* \hat{\sigma}^2, \tag{13}$$

where $S_t^*$ is defined in equation (11).

Using this adjusted Mallows's $C_p$ to select bandwidth, we get figure 9 through R programming. According to the figure 9, we get to see that the optimal bandwidth has become 71, which is much bigger than the one of independent case. The plot of $h = 71$ is similar to figure 8.

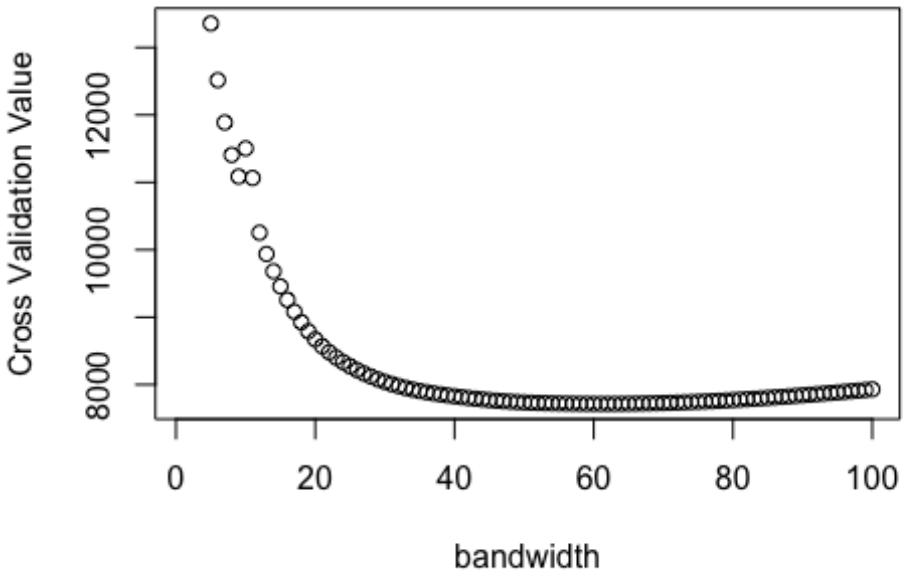


FIG 7. *The trend of cross-validation result corresponding to the choice of bandwidth using modified cross validation*

## 5. Times Series Model for Residuals

It seems that an entirely reasonable model for correlated residuals is an ARMA time series model. In fact, given our knowledge that we have surveys that averaged over multiple days, it seems like an MA model would work best.

$$y_t = \mu(x_t) + \sigma\epsilon_t$$

where $\epsilon_t$ is characterized as a stationary time series with

$$\epsilon_t = u_t + \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + \cdots$$

for $u_t \sim \mathcal{N}(0, 1)$.

As we have argued, the bias in any kernel estimator is not affected by the covariance of the residuals, but the variance will be. A positive correlation between observations that are close together in time suggests that we would want a larger bandwidth than our naive estimator.

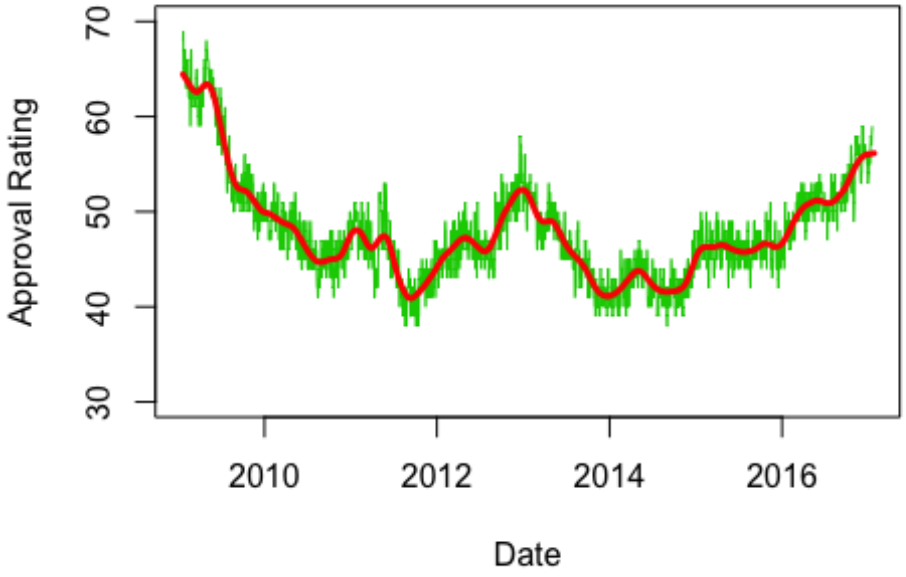


FIG 8. *Estimate of p using a bandwidth of 61, which is obtained from the modified cross validation.*

We can try to estimate the covariance of successive observations, but this would be biased by the changing mean $\mu(x)$. It is probably a better idea to estimate the auto-covariance from a set of preliminary residuals.

Figure 10 suggests that an MA(2) model for the residuals is realistic. The standard estimators applied to our residuals give us the results in table 1. We proceed by borrowing the "backfitting" technique used in additive and semiparametric models (see Hastie, Tibshirani and Friedman (2001) p. 259, for example). We fit the parametric model to the residuals, calculate residuals from this model, and then create a new data set where the original residuals are replaced by the residuals from the parametric model.

From the time series model, we can find estimates of the innovations $u_t$. These "residuals" should be nearly independent. Figure 11 shows the estimated innovations for this series.

The next step is to estimate the mean after replacing our original residuals by the these innovations. We have the derived observations

$$y_t^* = \hat{\mu}(x_t) + \hat{u}_t$$

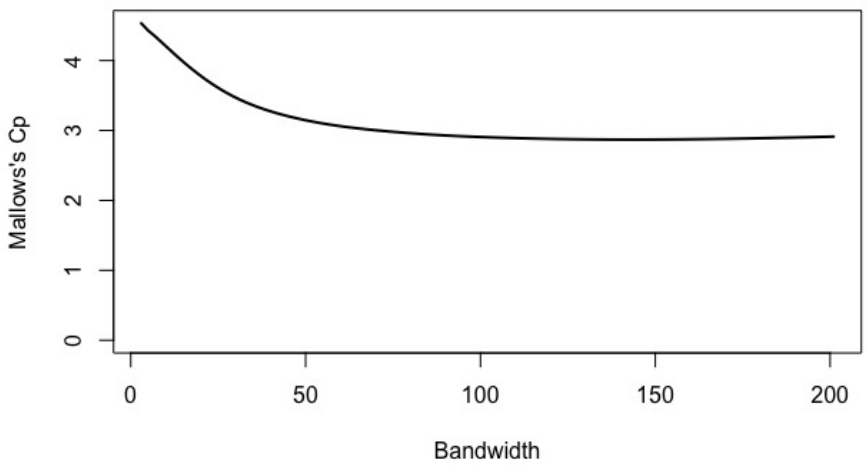


FIG 9. *Using modified Mallows's $C_p$ formula to find the optimal bandwidth. The minimum happens at x=71.*

TABLE 1
*The coefficient estimates from an MA(2) model applied to the residuals.*

| Coefficients | $\alpha_1$ | $\alpha_2$ |
|---|---|---|
| | 0.8093 | 0.7017 |
| std. error | 0.0142 | 0.0137 |
| Estimates | | |
| $\hat{\sigma}^2 = 1.382$, | log like.= -4458, | AIC = 8922 |

to which we can apply our standard bandwidth selection procedure. For instance, we could use Mallows $C_p$. Figure 12 shows $C_p$ over a range of realistic bandwidths, and we find that $h = 33.5$ minimizes the criterion. Figure 13 shows the resulting fitted value.

It is possible to iterate this whole operation another time to improve the analysis. We fit the time series model to the residuals from the fitted data using the $h = 33.5$ bandwidth. The estimates are in table 2. These results were not significantly different from our first iteration, and we do not see a big change.

### *5.1. Another Model for the Errors*

We made a relatively strong assumption regarding the model errors being from a MA(2) process. If instead we tried an ARMA(3,2) model, we get the coefficient estimates in table 3. The AR coefficients are smaller than the moving average

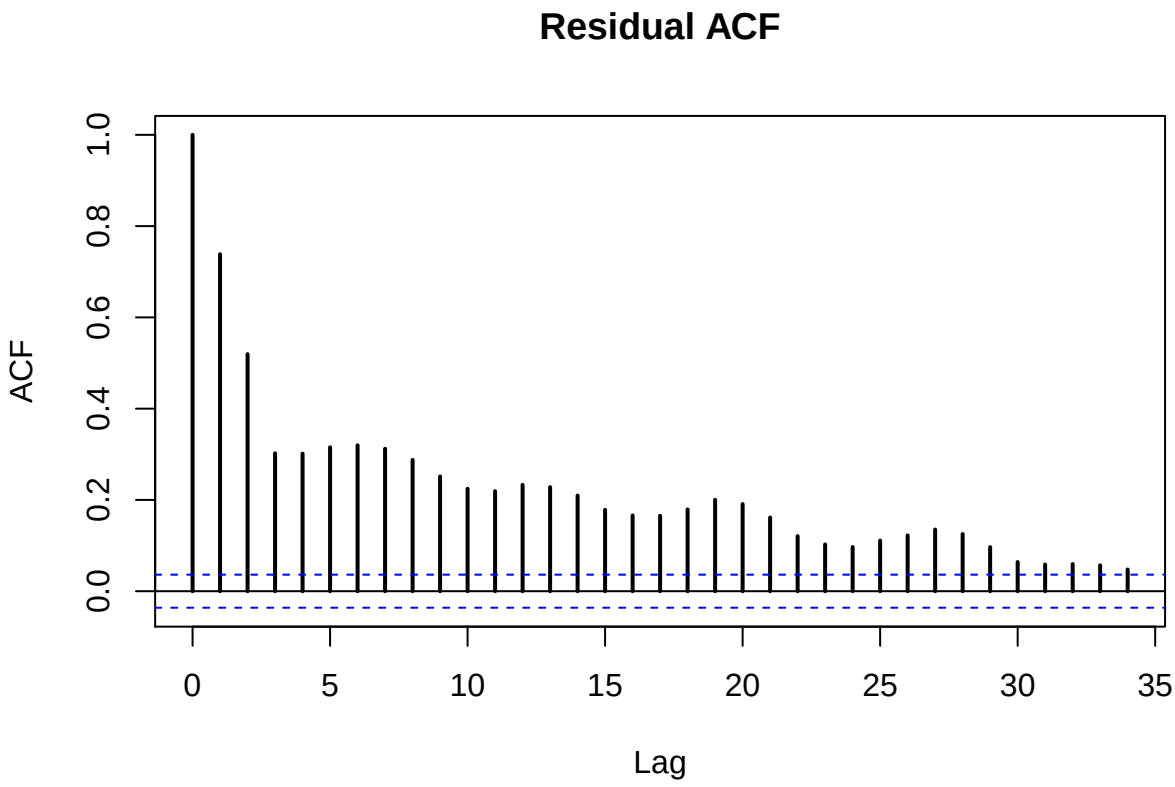


FIG 10. *The autocorrelation of the residuals from the initial estimate of the mean.*

TABLE 2
*Coefficient estimates from a new MA(2) model applied to the residuals from the chosen fit.*

| Coefficients | $\alpha_1$ | $\alpha_2$ |
|---|---|---|
| | 0.7823 | 0.6765 |
| std. error | 0.0145 | 0.0146 |
| Estimates | | |
| $\hat{\sigma}^2 = 1.267$, | log like.= -4332.77, | AIC = 8671.54 |

TABLE 3
*Coefficient estimates for ARMA(3,2) model applied to original residuals*

| Coefficients | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\alpha_1$ | $\alpha_2$ |
|---|---|---|---|---|---|
| | 0.1117 | 0.0836 | 0.1373 | 0.7782 | 0.6908 |
| std. error | 0.0375 | 0.0354 | 0.0303 | 0.0309 | 0.0264 |
| Estimates | | | | | |
| $\hat{\sigma}^2 = 1.319$, | log like.= | -4391.3, | AIC = | 8794.61 | |

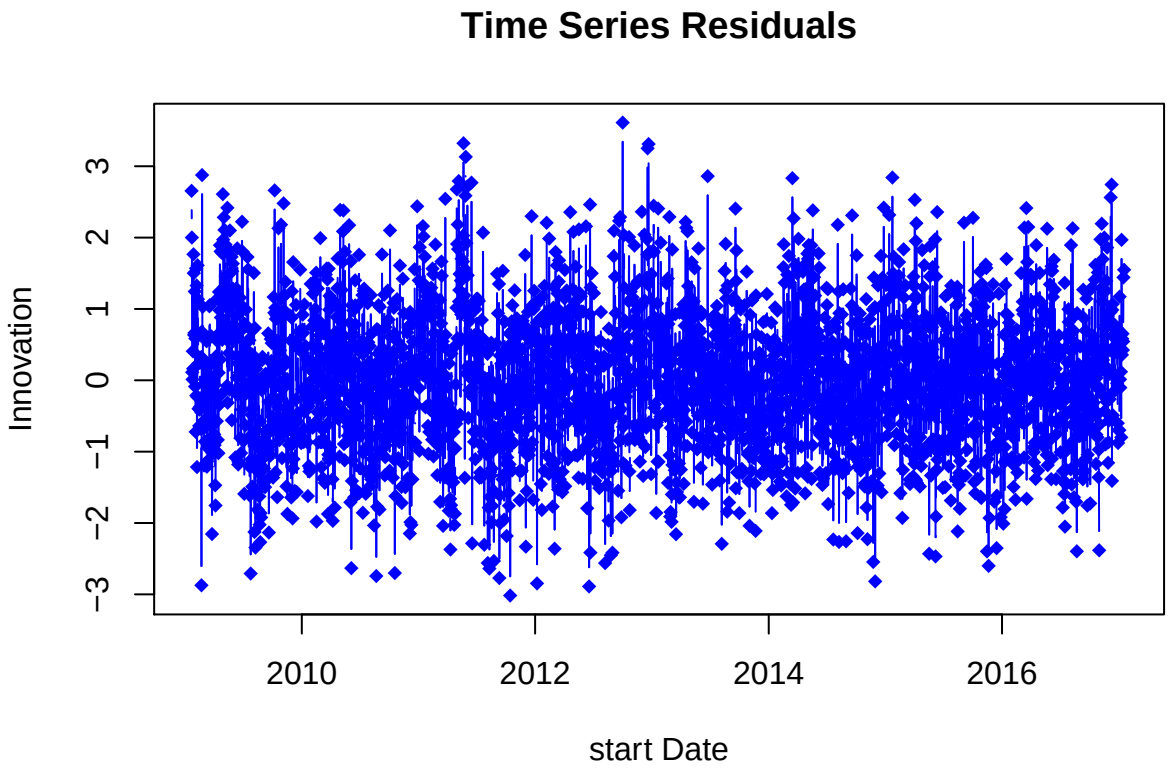


FIG 11. *The estimated innovations from the time series fitted to the initial residuals. These have the appearance of independence that we want.*

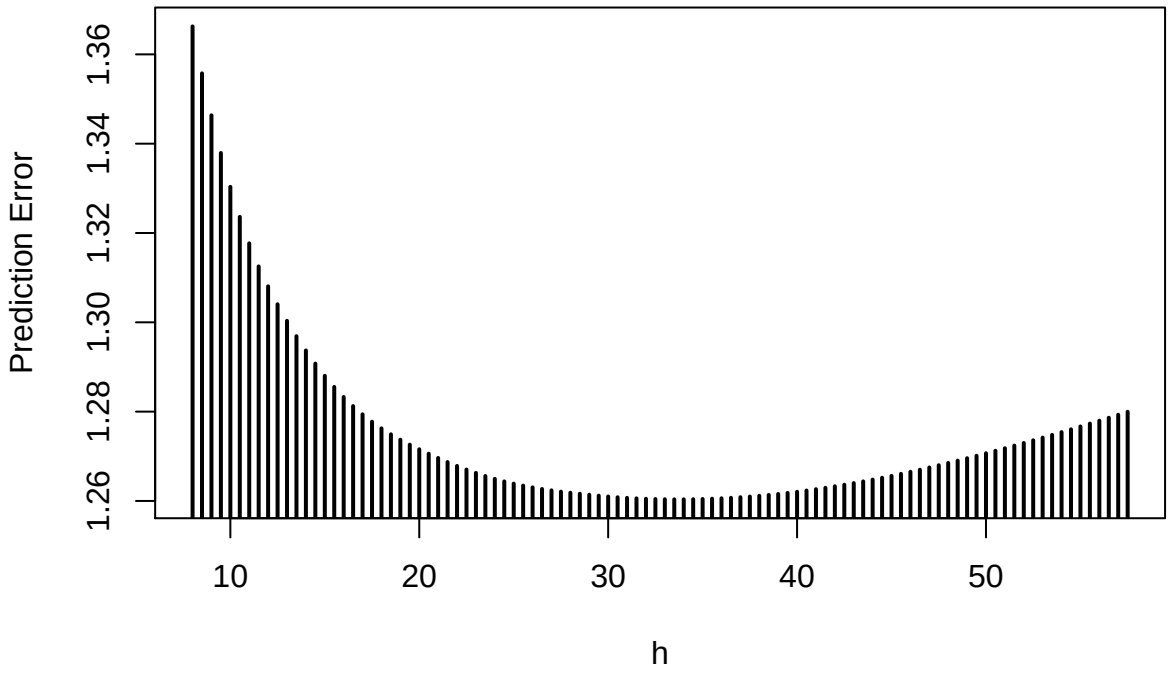


FIG 12. *The Mallows $C_p$ for a range of bandwidths.*

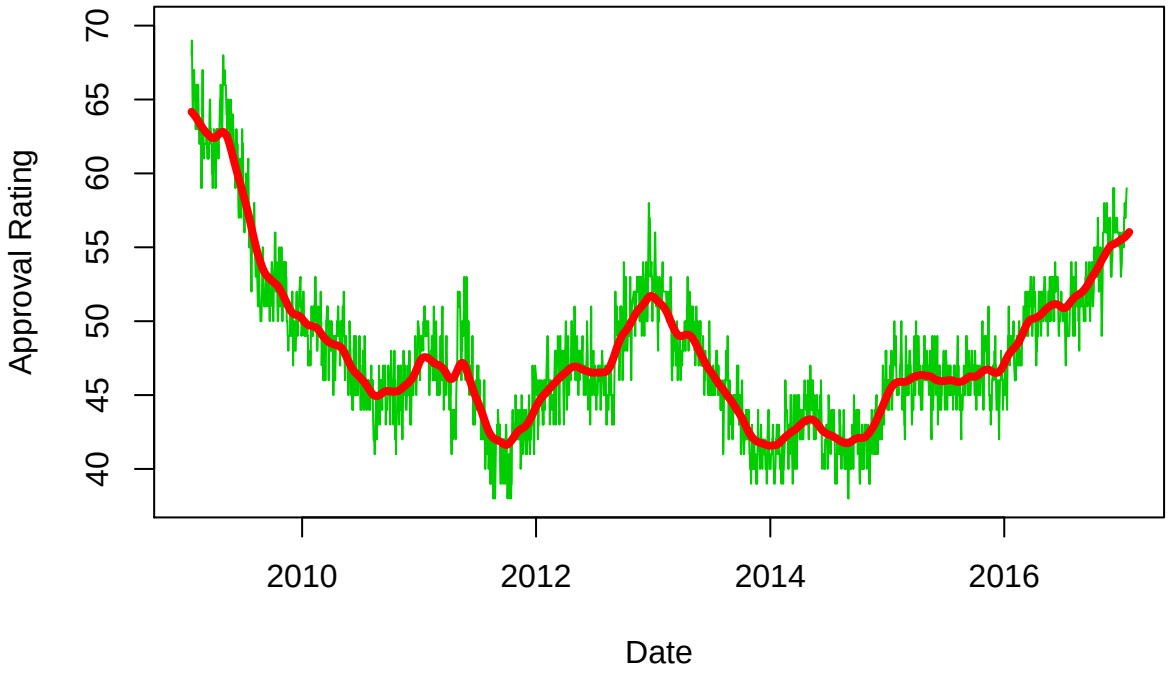


FIG 13. *The fitted approval rating using $h = 33.5$ on the derived observations. The green line represents the original observations.*

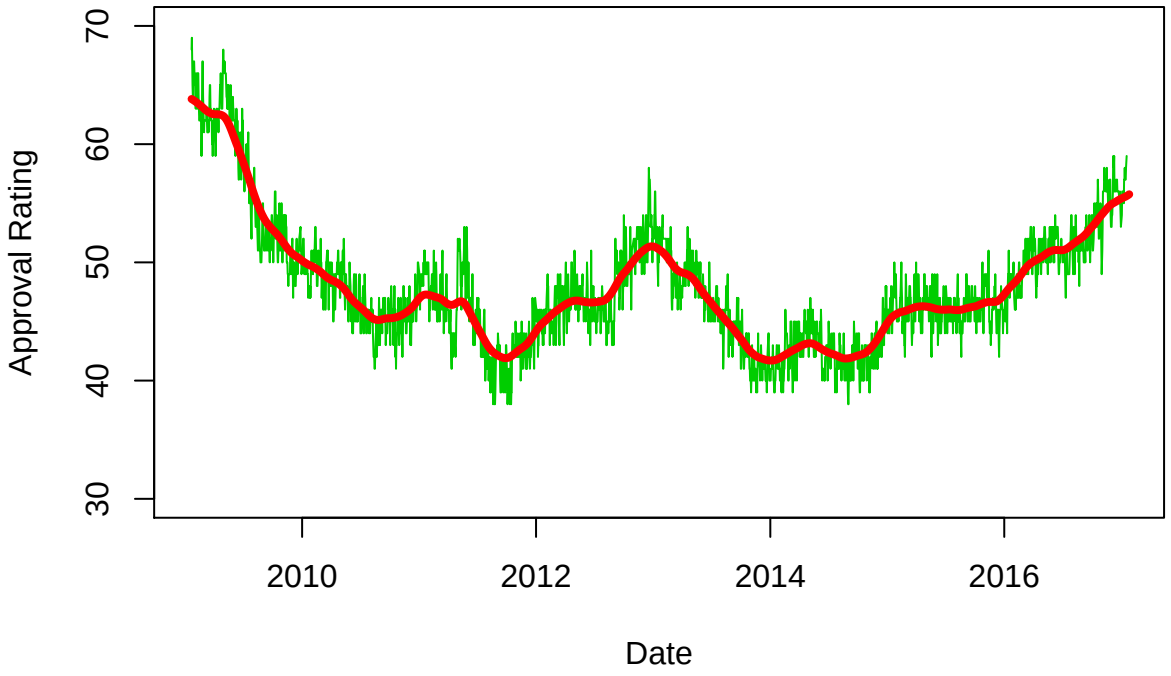


FIG 14. *Estimated approval rating using an ARMA(3,2) model for the errors.*

ones, but they are still greater than their margin of error. The AIC is reduced in this larger model.

Performing the same procedure as before on these residuals, we chose the bandwidth $h = 40$ to minimize the $C_p$, and Figure 14 shows the resulting fit. It makes sense that this fit is smoother than the fit from the MA(2) model because it is reading more of the variations over time as part of the residual process rather than changes in the population approval rate.

## 6. Conclusion

There were three proposed strategies for handling correlated errors in a kernel estimator. As in Chu and Marron (1991), the partitioned estimators sounds like a good idea, but there are certainly drawbacks. However, the selected bandwidth does generally confirm the results of our other two approaches.

The modified criterion of Altman (1990) was a successful approach, especially if the covariance matrix of the residuals is know. The Gallup data is probably unusual in that we have a concrete methodological reason for knowing the covariance structure of the errors.

If there is greater uncertainty regarding the generating process for the errors, then a procedure that included some time series estimation seems relevant. Our proposed algorithm for fitting these parameters iteratively seems like it may be useful, and it deserves further study.

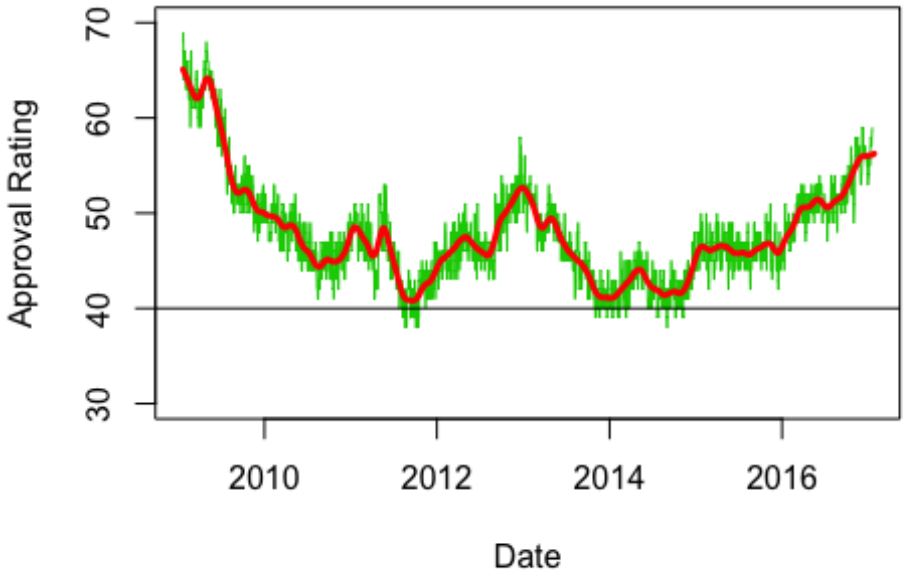


FIG 15. *compare the minimum approval rate from the best estimation with 40 percent*

We will use our estimate from the whole data using the bandwidth of $h = 61$ to address the questions that motivated our analysis. Figure 15 shows that there are several surveys with approval rates below 40 percent. They are distributed around the end of 2011, the end of 2013, the beginning of 2014, and at some random times in 2014. However, they are likely the result of the noise in the observation as our estimate never goes below 40. Under our best model, there is no time that we would estimate the approval rating as below 40%.

According to reports in 2011 (Pew Research Center, 2011), reports of the killing of the al Qaeda leader Osama bin Laden resulted in an increased approval rating for President Obama. Our estimate for the presidential approval rating

on May 1, 2011 was $\hat{\mu} = 47.04$. In figure 16, we can see the estimate of the approval rating along with the data from the month before and after the news event. Unfortunately, the bandwidth chosen by cross validation is too large for our purposes here. This bandwidth is tuned to minimize the error averaged over all dates, and so it expects slower changes in the underlying mean. This estimate is not sufficiently sensitive to short term changes in the approval rating that we would like to study in this instance. There is still about a 1 point increase in the estimate from April 1 to June 1, 2011. If we use a smaller bandwidth (such as $h = 15$), then the estimate would be increased to 3 or 4 points. The Pew Research Center (2011) study found a 7 point increase. We conclude that a better answer to this question would require an estimator that was better tuned to this specific issue.

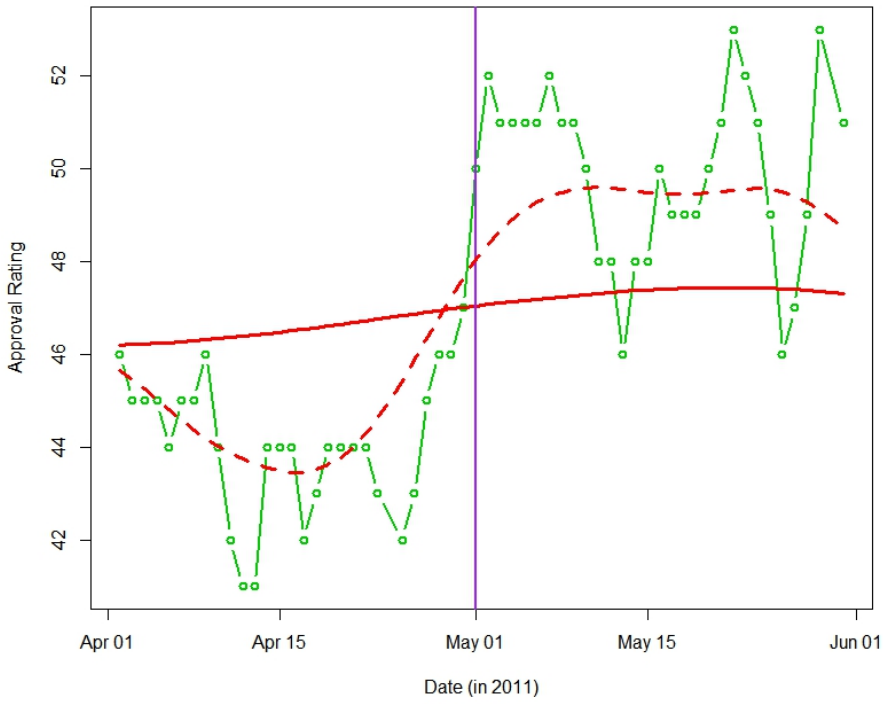


FIG 16. *President Obama's approval rating for April and May of 2011. The vertical line indicates May 1 when President Obama announced that Navy Seal's had kill bin Laden. The estimate of the approval rate shows little change over the two months because the bandwidth chosen by cross validation $h = 61$ is as wide as this entire subset of the data. Alternatively, a smaller bandwidth (such as $h = 15$ shown here as a dashed red line) shows greater evidence of an effect from the news.*

**Acknowledgements**

The data used in this analysis was compiled by Gerhard Peters at http://www.presidency.ucsb.edu/data/popularity.php. It is adapted from survey results published by Gallup, Inc (http://www.gallup.com/interactives/185273/presidential-job-approval-center.aspx?utm_source=genericbutton&utm_medium=organic&utm_campaign=sharing).

## Appendix A: R code used in Analysis

The data from the Presidential Job Approval website (Peters and Wolley, 2017) is used in our analysis, and each survey is associated with the first day that the interviews were done. All the surveys are actually performed over multiple days.

```
ObamaApproval ← read.csv("ObamaApproval.csv", stringsAsFactors
    =FALSE)
start.date ← as.Date(ObamaApproval$Start.Date,format="%m/%d/%Y
    ")
lng.dtes ← seq(from=as.Date("2009-01-21"),to=as.Date("
    2017-01-25"),by=1)
# Data is extended to include "missing" on days when no survey
     is taken
surveys ← sapply(lng.dtes,function(x){any(start.date == x)})
approvals ← rep(0, length(lng.dtes))
approvals[surveys] ← rev(ObamaApproval$Approving)
```

The kernel fitting function designed to use data in this format.

```
#Kernel Fitting Function
fit.loc.ave.krnl ← function(appv, dtes, knl){
   h ← length(knl)
   appv.add ← c(rep(0,(h-1)/2),appv, rep(0,(h-1)/2))
   dtes.add ← c(rep(0,(h-1)/2),dtes, rep(0,(h-1)/2))
   svs ← filter(dtes.add, knl, method="convolution",sides=1)
   ht.mu ← filter(appv.add, knl, method="convolution", sides
       =1)/svs
   return(list(mu = ht.mu[seq(1-h, -1)], sv = svs[seq(1-h, -1)
       ]/max(knl)))
}
#Quartic Kernel
qknl ← function(h){
   ksnl ← seq(from=0, to=h, by=1)
   m ← length(ksnl)
   hlf ← ( 1- (ksnl/h)^2)^2
   c(hlf[m:1],hlf[2:m])
}
mu.hat ← fit.loc.ave.krnl(approvals, surveys, qknl(150))
plot(start.date, ObamaApproval$Approving,type="l",lwd=1,col=3,
```

```
      xlab="Date",ylab="Approval Rating",ylim=c(30,70))
lines(lng.dtes,mu.hat$mu,lwd=3,col=2)
```

Using R code to calculate the variance estimation of partitioned estimation.

```
m<-length(surveys)
srvys<-matrix(surveys[-m],ncol=7, byrow = TRUE)
apvls<-matrix(approvals[-m], ncol=7,byrow=TRUE)
tot=0
for (k in 1:7){
  for (i in 1:416){
    if ((srvys[,k][i]==TRUE) & (srvys[,k][i+1]==TRUE)){
    temp=(apvls[,k][i]-apvls[,k][i+1])^2
    }
    else {
    temp=0
    }
    tot=tot+temp}
}
vr1<-1/(2*m-14)*tot
```

Another method of variance estimation.

```
mu.hat<-fit.loc.ave.krnl(approvals, surveys, c(1:22,21:1))
vr2<-sum((approvals[surveys] - mu.hat$mu[surveys])^2)/(2919-
    sum(1/mu.hat$sv))
```

The code which are used to find the optimal bandwidth using modified Mallows's $C_p$.

```
for( j in 1:100) {
  knll<-qknl((j+5)/2)
  mu.hath<- fit.loc.ave.krnl(approvals, surveys, knll)
  sss<-1/mu.hath$sv[surveys]
  res<-approvals[surveys]-mu.hath$mu[surveys];
  Ccv<- sum(knll[seq(-1,3)+floor((j+1)/2)+1]*c(1/3,2/3,1,2/3,1/
      3))
  Ph[j]<-mean(res^2) + 2*vr1*mean(Ccv*sss);
}
which.min(Ph)
```

The code to produce figure 10.

```
rsds<-approvals - mu.hat$mu
rsds[!surveys]<-NA
acf(rsds, main="Residual ACF", lwd=2, na.action=na.pass)
```

The "backfitting" type procedure for finding the bandwidth.

```
ts.fit ← arima(rsds,order=c(0,0,2), include.mean = FALSE) #fit
    MA(2) model
in.ut ← resid(ts.fit) #Take the innovations from this fit
r.apprv ← rep(0, length(lng.dtes))
### Construct synthetic observations from uncorrelated
    innovations
r.apprv[surveys] ← in.ut[surveys] + mu.hat$mu[surveys]
### Find the best bandwith to fit synthetic observations
Ph.ma ← rep(-999,100) #blank vector
n ← sum(surveys)
for( j in 1:100) {
  mu.hath ← fit.loc.ave.krnl(r.apprv, surveys, qknl((j+15)/2)
      )
  RSS ← sum( (r.apprv[surveys] - mu.hath$mu[surveys])^2)
  trS ← sum(1/mu.hath$sv[surveys])/n
  Ph.ma[j] ← RSS/n + 2*ts.fit$sigma2*trS # Mallow's Cp
}
h.star ← (which.min(Ph.ma)+15)/2
```

## References


Altman, N. S. (1990). Kernel smoothing of data with correlated errors. *JASA* **85** 749–759.

Pew Research Center (2011). Public "Relieved" By bin Ladens Death, Obamas Job Approval Rises, Polling Report, The Pew Research Center for The People & The Press/Washington Post. http://www.people-press.org/2011/05/03/public-relieved-by-bin-ladens-death-obamas-job-approval-rises/.

Chiu, S.-T. (1989). Bandwidth selection for kernel estimate with correlated noise. *Statistics & Probability Letters* **8** 347 - 354.

Chu, C. K. and Marron, J. S. (1991). Comparison of Two Bandwidth Selectors with Dependent Errors. *Ann. Statist.* **19** 1906–1918.

Gallup (2016). Gallup Daily: Obama Approval Rating. http://www.gallup.com/poll/113980/gallup-daily-obama-job-approval.aspx. accessed: June 12, 2017.

Hart, J. D. (1991). Kernel Regression Estimation With Time Series Errors. *Journal of the Royal Statistical Society. Series B (Methodological)* **53** 173-187.

Hastie, T., Tibshirani, R. and Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics*. Springer New York.

Mallows, C. L. (1973). Some comments on $C_p$. *Technometrics* **15** 661–675.

Opsomer, J., Wang, Y. and Yang, Y. (2001). Nonparametric regression with correlated errors. *Statistical Science* 134–153.

Peters, G. and Wolley, J. T. (2017). Presidential Job Approval. http://www.presidency.ucsb.edu/data/popularity.php.

Rice, J. (1984). Bandwidth Choice for Nonparametric Regression. *Ann. Statist.* **12** 1215-1230.

Tecuapetla-Gómez, I. and Munk, A. (2017). Autocovariance Estimation in Regression with a Discontinuous Signal and $m$-Dependent Errors: A Difference-Based Approach. *Scandinavian Journ. Statist.* **44** 346–368. 10.1111/sjos.12256.