Cycling Involvements

Jake Bowers
University of California, Berkeley
jbowers@csm.berkeley.edu

February 23, 2002

Abstract

DRAFT — COMMENTS APPRECIATED

This paper shows that decomposing a time-series into periodic components can provide politically useful information about the shape of aggregate political participation in the United States. Specifically, it provides statistical tests for the periodicity of the aggregate time series of political participation and explains how this decomposition and associated tests work. Between 1973 and 1994 there appears to be an annual cycle in the reporting of political participation by respondents to a series of polls conducted by Gallup 10 times per year. This seasonality has been noted by in one other publication, by Rosenstone and Hansen (1993), but was explained as tied to a summer political cycle. In this article I suggest that this discovery has more to do with annual cycles in the composition of the Gallup sample than politics. I am currently trying to obtain detailed information on the monthly mail volume into and out of Congress. With this information, I will be able to test more directly if, despite the changes in sample composition of the Gallup polls, the political participation of Americans ought to be see as an "output" of Congressional mobilization or an "input" or in what way the flow of participation into Congress is related to the flow of mobilization out of it.

Keywords: Frequency domain time series analysis; Fourier analysis; periodogram; squared coherency; aggregate political participation

1 Why should we care about periodicity?

> ...I am trying to make sense of the periodic outbreaks of mass participation in public affairs and of collective action in general. (Hirschman, 1982, page 79)

Periodicity is feature of participation as it changes over time that can help us understand it better. In political science the periodicity of time-series has been mostly ignored in favor of examinations of trend and lag structure (For the two exceptions that I know of see Beck, 1991; Rosenstone and Hansen, 1993).

Detecting and explaining the periodicity in political participation has substantive implications. Say, for example, that periodicity maps onto presidential electoral cycles. If this were the case, then we might think that people participate when national (instead of local) institutions provide opportunities, or when attention is focused on politics by national media coverage of elections. This understanding of political participation would indicate perhaps yet another sign of the decline in "civic society" that worries communitarian and civic republican scholars. What if the periodicity in political participation instead is discovered to cycle at a very slow rate? This might suggest that what appears to us as a declining trend is actually just the downward swing of a slowly shifting system (perhaps like the one envisioned by Hirschman (1982)).

In their 1993 book, Rosenstone and Hansen (1993) showed that Americans tended to write to Congress, attend meetings, and to sign petitions more in the summer than in other months. Although they did not provide any information about how they arrived at this estimate, they argued that the summer is special because it is when bills come to the floor for debate and voting and it is the season when petitions must be signed in order to get candidates and issues on ballots. Their discussion of seasonality aimed to provide more evidence for their general argument that the political participation of individual Americans is driven, in large part, by mobilization on the part of politicians and activists. Although I do not dispute their general claim, I analyze the same time-series that they used [extended by 4 years from 1990 to 1994] in order to shed more light on (1) how one might rigorously *test* for the existence of cyclicality in the time-series of political participation and (2) the meaning of an annual cycle discovered within the 12 different participation activities.

I present several different methods for detecting periodicity and then also show how one might detect dynamic relationships between two time-series which have periodic components.

2 The Data

From 1973 to 1994, the Roper organization fielded 10 national surveys each year of about 2,000 Americans.[1] During each of these 207 surveys, respondents answered a series of questions about their participation in politics in the following format:

"Now here is a list of things some people do about government or politics. (HAND RESPONDENT CARD) Have you happened to do any of those things in the past year? (IF "YES") Which ones?

1. "Written your congressman or senator"
2. "Attended a political rally or speech"
3. "Attended a public meeting on town or school affairs"
4. "Held or run for political office"
5. "Served on a committee for some local organization"
6. "Become[Served as] an officer of some club or organization"
7. "Written a letter to the paper"

[1]See the Notes section at the end for more details on how this dataset was constructed.

8. “Signed a petition”
9. “Worked for a political party”
10. “Made a speech”
11. “Written an article for a magazine or newspaper”
12. “Been a member of some group like the League of Women Voters, or some other group [which is] interested in better government”

3 Detecting Periodicity: The Ocular Method

The first, and easiest, method for detecting periodicity in a time series involves merely looking at the data. Figure 1 shows a set of data for which an quick look is all that is necessary. The percentage of the voting age population that actually turns out to vote has a clear 4 year cycle. In fact the distinction is so clear that one might even think of the vote turnout in presidential election years and in non-presidential election years as two separate series! A smoothed line with a wide bandwidth has been fitted to this series in order to display that the decline in turnout over time.[2]

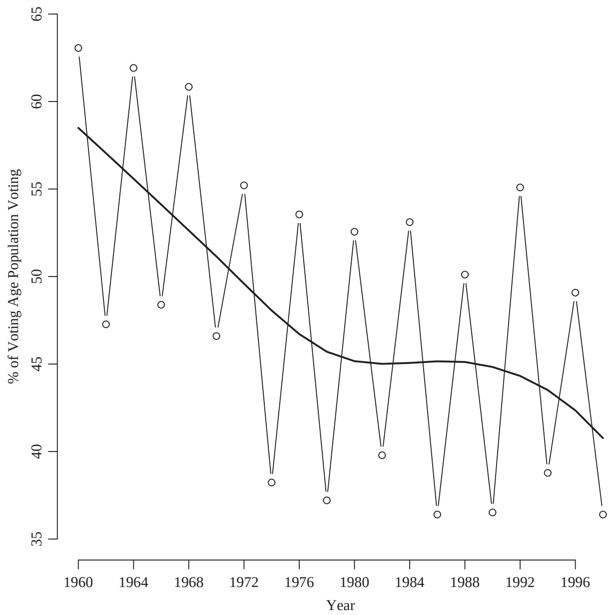


Figure 1: Voting Turnout 1960-1998

Figures 2 and 3 display the series for the 12 series of the Roper data. What was clear from looking at information about vote turnout is no longer clear. The smoothed lines (using 50% of the nearest neighbors as a bandwidth) do show non-linear trends of decreasing proportions of the citizenry participating, but if any periodicity exists in these series, it is not discernible to the naked eye. It is worth noting that each of the different activities is plotted on a different scale so that the fluctuations that might indicate periodicity are magnified. It is interesting to see that while nearly

[2]For details on the smoothing and analysis, see the “Notes on Analysis” section at the end of the paper.

26% of the sample reported attending a local public meeting in 1973, in the same year about 1% reported holding or running for an elected office, and about 7% reported working for a political party. Thus, certain acts appear more difficult than others, or at least less popular than others. Despite decreasing (nonlinear) trend across all acts, each one looks slightly different. Due to this diversity in political participation, the rest of the analyses in this paper display the results for all twelve acts. The disadvantage will be that the reader will feel overwhelmed in information. The advantage will be that any one finding will have essentially twelve replications on different types of behavior.

4 Detecting Periodicity: Regression on Lagged Values

If the ocular method doesn't work on the seemingly noisy data of the participation series, then one must turn to other methods. One technique for characterizing the temporal structure in time-series that has been popular in economics has been to inspect the Autocorrelation function (ACF) of the data. One can think of the ACF as the result of regressing the series upon lagged values of itself, and the Partial Autocorrelation function (PACF) as the result of regressing the series upon lagged values of itself *controlling for the influence of intervening lagged values.*

We can calculate the coefficients which define the ACF using the following formula:

$$ACF = r_k = \frac{\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1} T(x_t - \bar{x})^2} \quad (1)$$

Just as a correlation coefficient is a standardized version of a covariance, the ACF is a standardized version of an AutoCovariance, which is defined as:

$$c_k = \frac{1}{N}\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x}) \quad (2)$$

And $r_k = c_k/c_0$

The following figures show ACFs and PACFs for the 12 series which have been detrended and demeaned via subtracting the smoothed trend lines plotted in Figures 2 and 3. The plots cut off the ACF at lag 0 since that value is always 1. The dotted lines display the 95% confidence region for the null hypothesis that there is no temporal structure in the data — i.e. that, just as white light is combination of light at all frequencies such than no single frequency can predominate, the time-series is white noise, with no single temporal aspect dominating. Each vertical line plots the value for the ACF or PACF — longer lines display stronger relationships at that lag. One would expect about 1 out of every 20 coefficients to be longer than the confidence region merely by chance (it is a 95% confidence region after all).

The ACFs and PACFs show that there appear to be strong relationships between the proportions reporting participation in subsequent surveys (this is shown by the strong relationships at lag 1). The plots also show clusters of negative relationships around lags 12-17. Since this occurs in for multiple activities, these relationships are probably not just chance fluctuations. But, it is not

Figure 2: Political Participation, Series 1-6

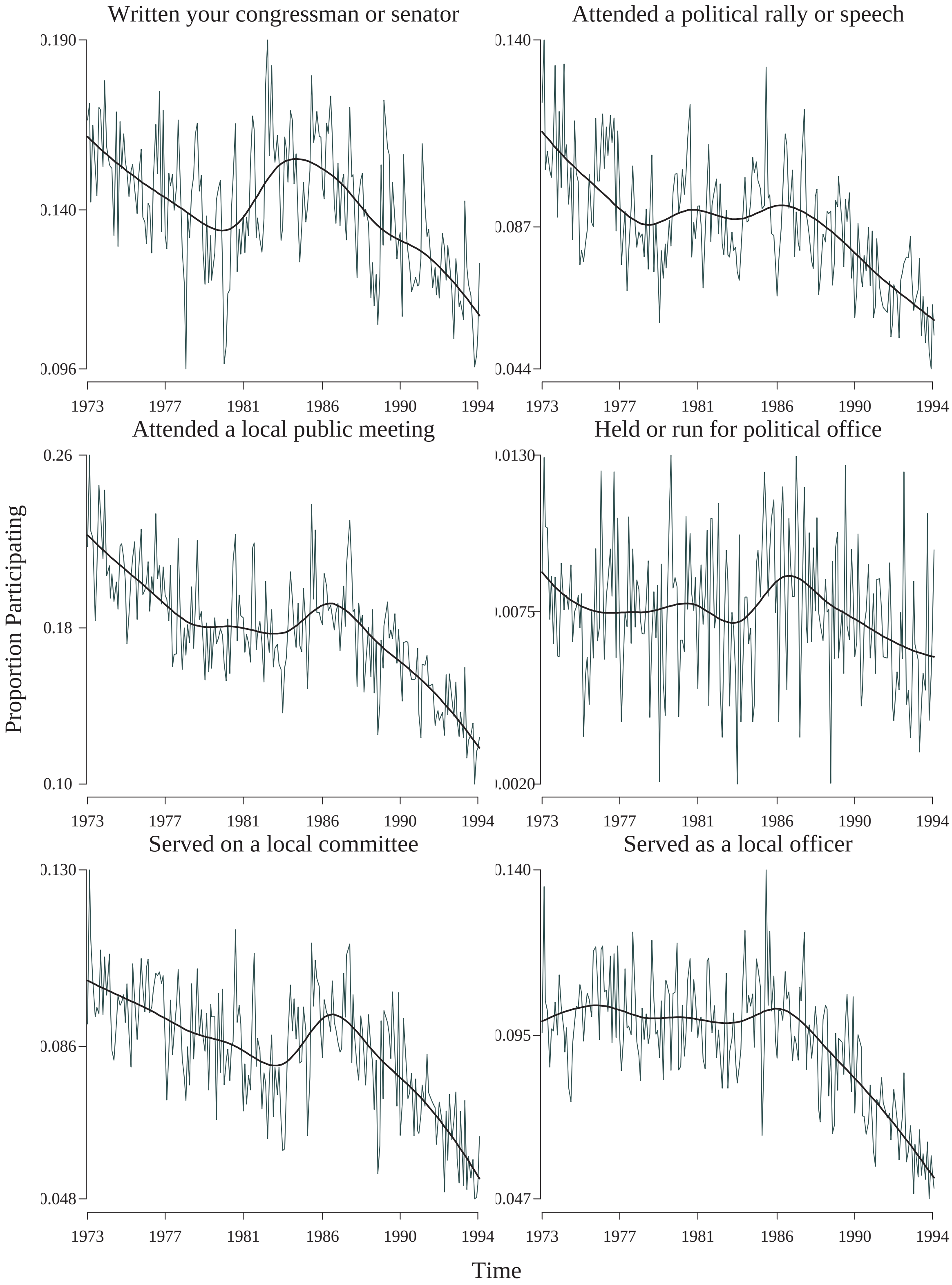


Figure 3: Political Participation, Series 7-12

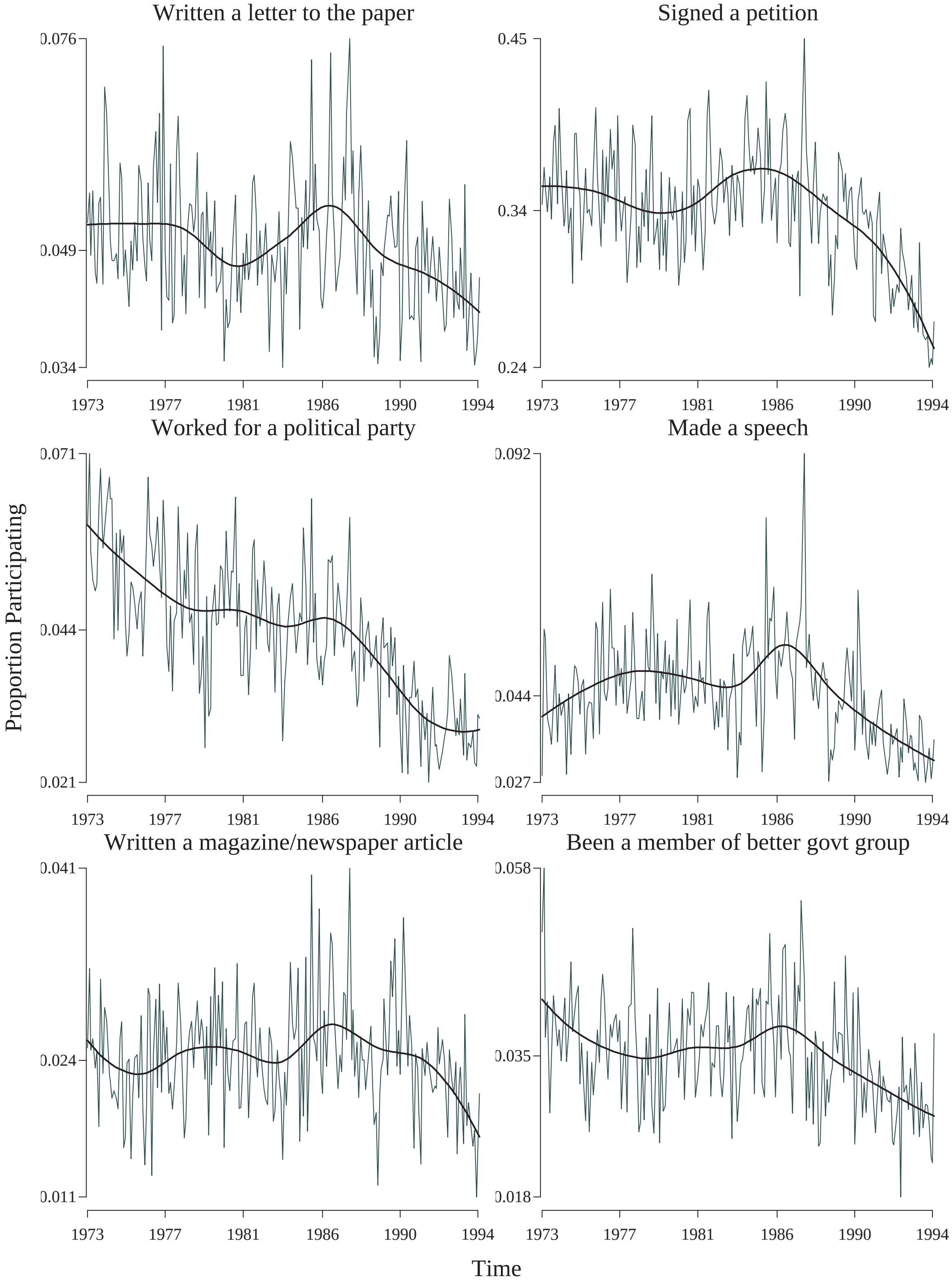
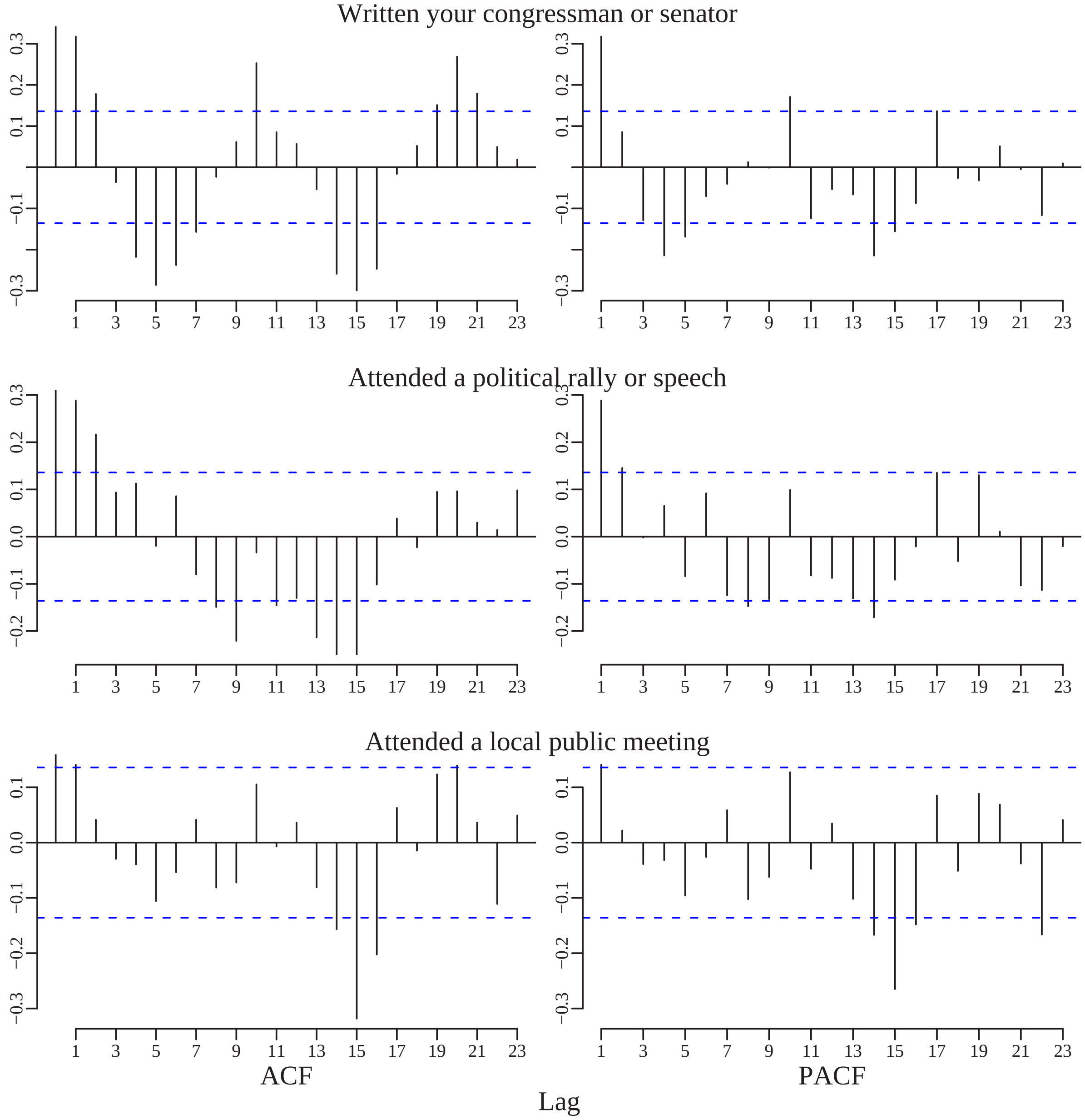


Figure 4: ACFS and PACFS for Series 1-3



exactly clear what this might mean. The sampling frequency of these time-series is 10 times per year. Thus, an annual cycle should show up at lag 10 rather than 12-17. Overall, the ACFs and PACFs do not display much information that is easily interpretable about periodicity.

Figure 5: ACFS and PACFS for Series 4-6

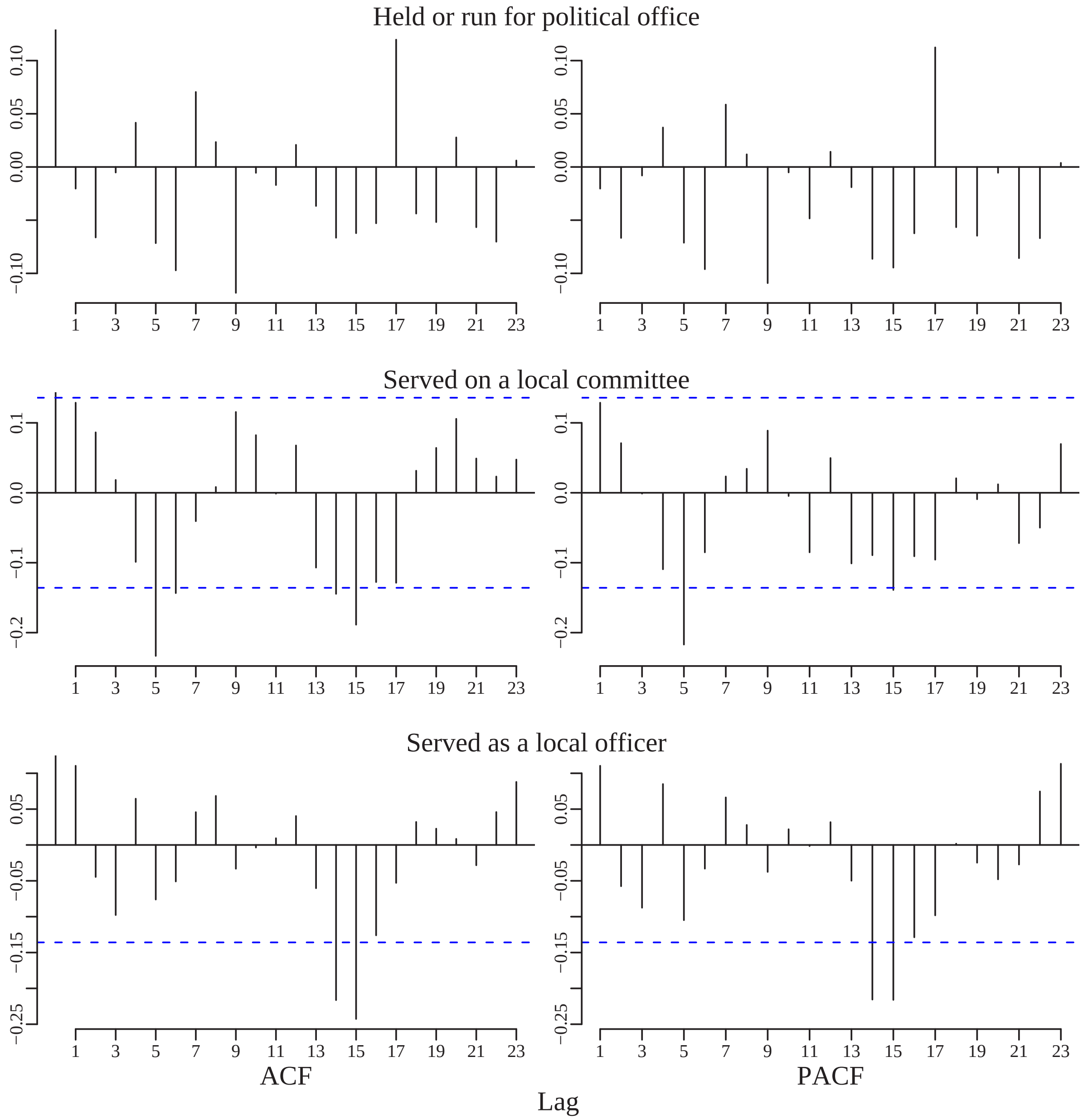


Figure 6: ACFS and PACFS for Series 7-9

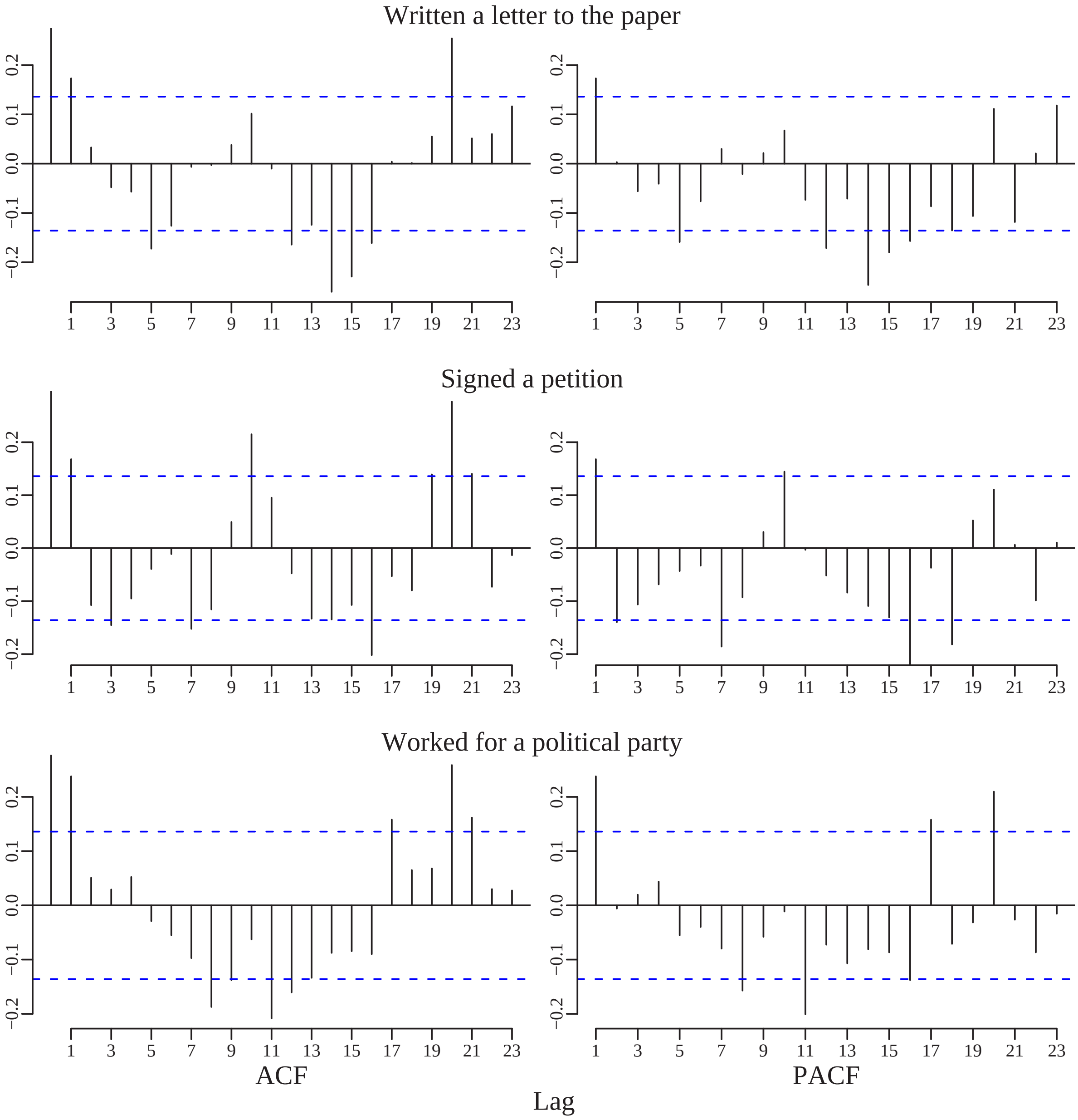
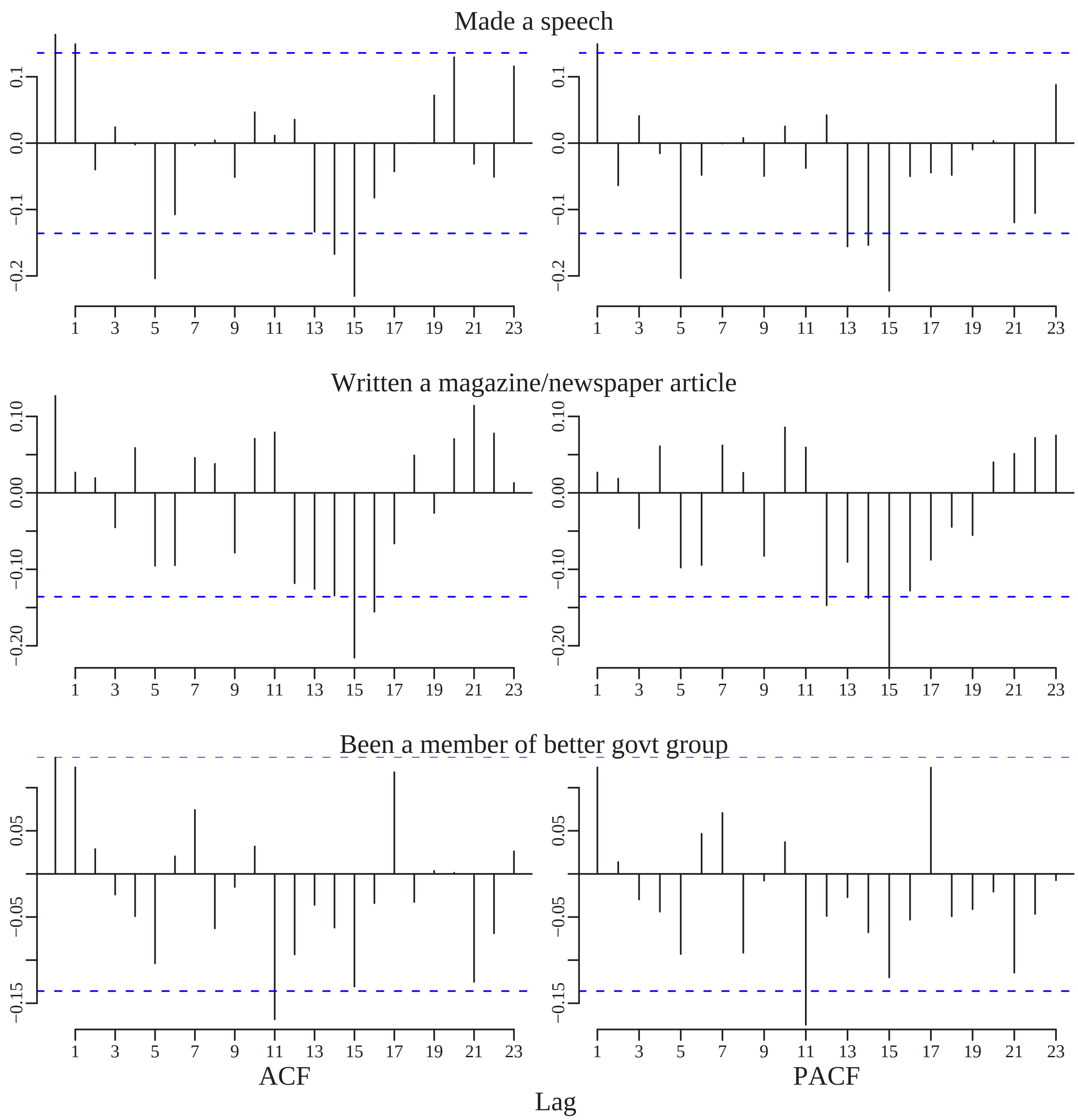


Figure 7: ACFS and PACFS for Series 10-12



5 Detecting Periodicity: Regression on Sinusoids

Since neither the ocular nor the ACF/PACF approach seemed to provide much help in detecting periodicity, it makes sense to try something else — namely using mathematical functions that are periodic themselves to characterize the time-series. This method is the most basic approach to what is known as frequency domain time series analysis while the ACF/PACF approach tends to be the basic approach of what is known as time-domain time series analysis. [insert more here on the distinction, what kinds of data and questions are useful for each kind — actually, perhaps move this to the front] Since sine and cosine functions are periodic, one can do a least squares fit of a time-series with combinations of sine and cosine functions as independent variables as follows:

$$y_t = \alpha + \sum_{\lambda} (a_\lambda \cos(2\pi t\lambda) + b_\lambda \sin(2\pi t\lambda)) + u_i \tag{3}$$

where λ is the frequency in cycles per year. In this case, I chose $\lambda = 1/10, 1/5, 1/4, 1/3, 1/2, 1, 2, 3$ to represent 10 year, 5 year, 4 year, 3 year, 2 year, 1 year, 1/2 year, and 1/3 year cycles.

Figures 8 and 9 plot the estimates for a and b from equation 3 for the values of λ that I plugged in. The coefficients are plotted as circles with bars extending ± 2 standard errors around the point estimates. Each panel has a horizontal line at zero. To the extent that a point and it's associated error bar do not touch the zero line, we can say that such a coefficient estimate is not zero. One pattern which leaps out of these plots [all plotted with the same y-axis] is that there appears to be an annual pattern — i.e. that the sin and cosine functions at a 1 year cycle tend to be strongly positive and different from zero. Of course, each λ has two coefficients here. So it is a bit difficult to interpret the situation, say, where the cosine function is strongly different from zero but the sine function is nearly equal to zero.[3] It is also interesting to note very few strong relationships with electoral cycles at frequencies of 4 or 2 — or at least that where these relationships clear zero, they are not as strong as the estimates for the annual periodicities. Another notable aspect of these plots is the extent to which the sinusoids tend to have any explanatory power [such as the series capturing writing to members of the house and senate] versus series that seem to have no periodic relationships at least at the frequencies I choose [such as the series capturing the proportion of Americans who hold or run for electoral office]. Basically this difference is due to the differences in scale and variance of the series themselves [probably should plot these on different scales].

The advantage of this approach is that one can fit periodic functions directly using well-understood techniques and can produce statistical tests of the extent to which the different periodic components contribute to the series overall. Thus, if the question is about periodicity, this method is much more direct. The disadvantages of this approach include those attendant to any use of OLS, such as violation of assumptions (at least for estimating the standard errors of the coefficients). It was also annoying to have two coefficients returned for each frequency — one would be enough. Finally, one may not choose the correct frequencies at which to run this regression — and because one must have two coefficients for each frequency, the number of frequencies that are testable can eat up degrees of freedom quite rapidly.

[3] Actually, as we'll see, this can be interpreted as saying that the imaginary part of the Fourier coefficient is nearly zero. But for now, two coefficients per frequency is clearly one piece of information too many.

Figure 8: OLS Coefficients for Series 1-6

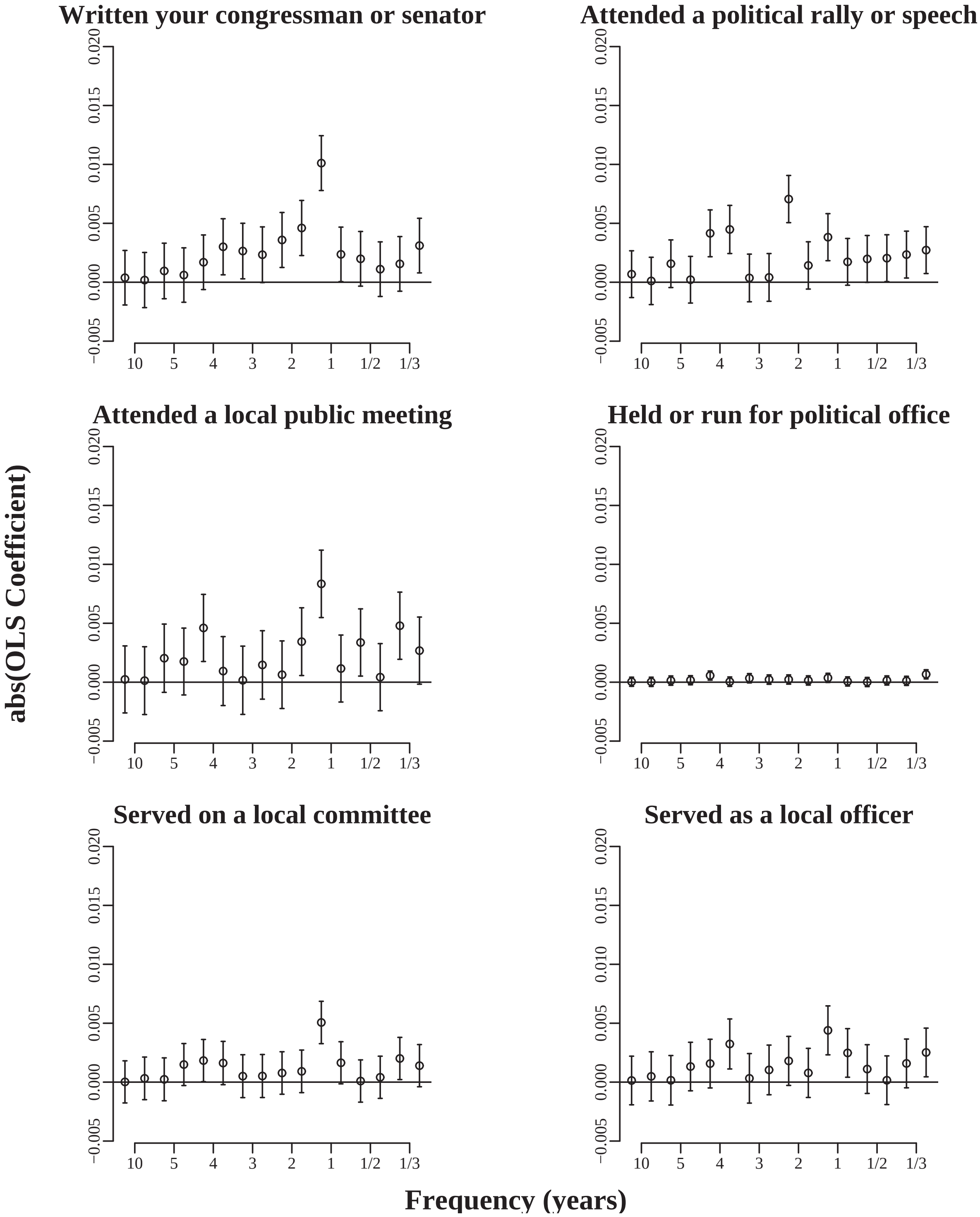
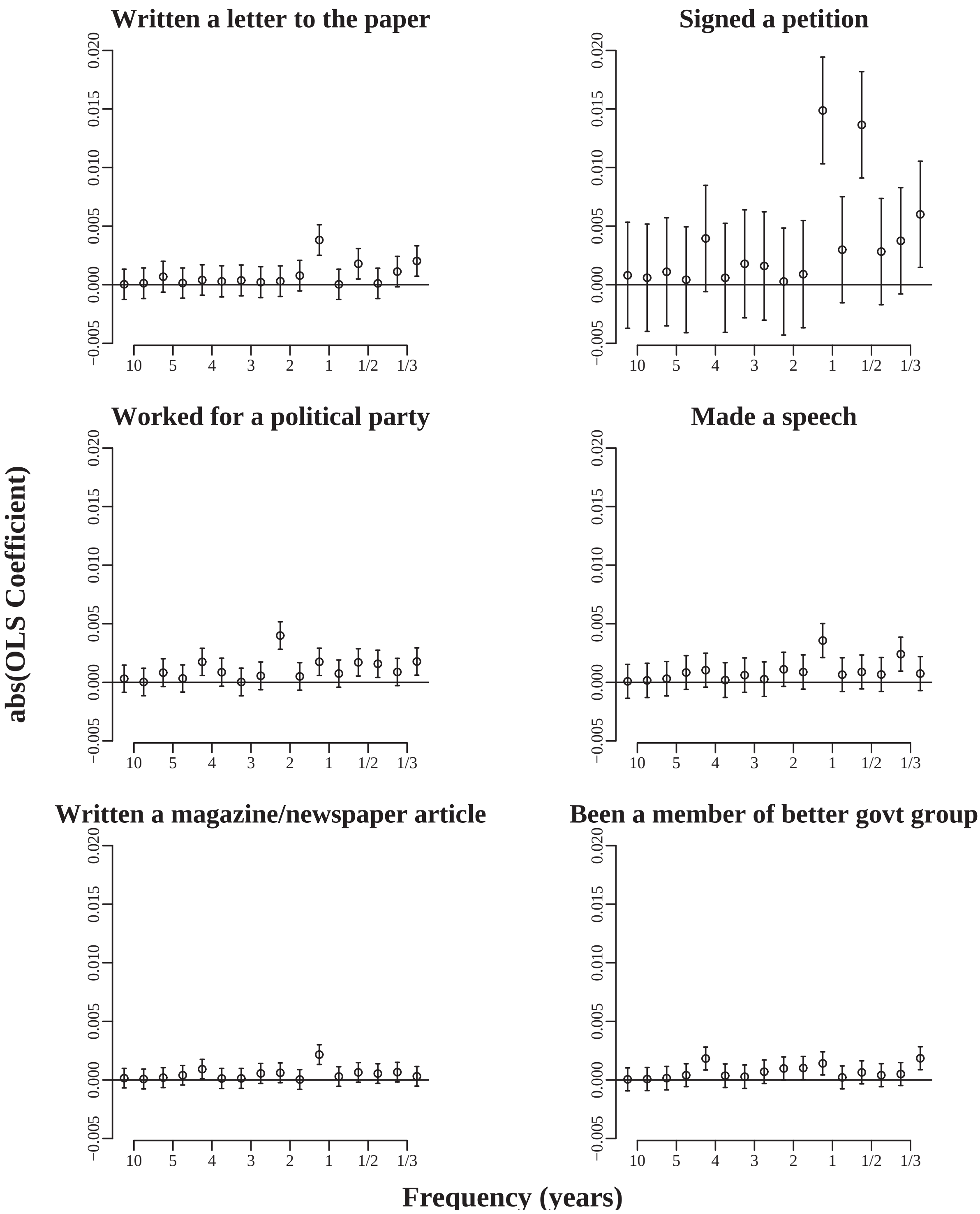


Figure 9: OLS Coefficients for Series 7-12



6 Detecting Periodicity: The Periodogram

The most elegant method for detecting periodicity in a stationary time-series is via the periodogram.[4] On can think of the periodogram as an analysis of variance in which the variance of a stationary time-series, $Y(t)$, is decomposed into periodic parts. This decomposition works by approximating the time series with a trigonometric polynomial, where

$$Y(t) = \sum_{\lambda=-T/2}^{T/2} D(\lambda)e^{-i2\pi\lambda t} \approx \sum \left(A(\lambda)\cos(2\pi\lambda t) + B(\lambda)\sin(2\pi\lambda t)\right). \tag{4}$$

Notice that the right-hand side of this equation looks a lot like the regression on sinusoids that I displayed above. It turns out that $D(\lambda)$ is an elegant and useful way to combine the information in A and B into one number. The raw periodogram, $I(\lambda)$, turns the coefficients contained in $D(\lambda)$ into magnitudes: $I(\lambda) = \frac{1}{T}|D(\lambda)|^2$. It turns out that one can get consistent estimates of the amount of variance explained in a series by smoothing over nearby values of the raw periodogram. [more here? or in the appendix?] Figures 10 and 11 show the smoothed periodograms[5]

In each plot the black lines represent the smoothed periodogram estimates for $D(\lambda)$ calculated for each series. The dark gray rectangle represents the 95% confidence region for the null hypothesis that these series are white noise. The lighter gray lines tracking the periodogram values are the pointwise 95% confidence intervals for the null hypotheses that the value of the periodogram estimate $I(\lambda)$ is actually the population value.[6] One advantage that the periodogram has over the ACF and PACF functions is that one can easily derive the sampling distribution of the periodogram estimates in general, not just for the situation in which the series is white-noise, whereas for the ACF and PACF functions is it more difficult to represent uncertainty about the actual estimates seen as null hypotheses. [fix this, too confusing] The frequencies at which the black lines emerge from the white-noise regions of the gray rectangles are those which have the most (non-negligible) power in determining the variance of the series. As we can see, most of these series are dominated by non-cyclical components [or at least components which are not easily distinguishable from white noise given the current length of the series]. However, the places where the periodogram does emerge from the white-noise region tends to be at the frequency of 1 cycle per year.

Table 1 summarizes these graphs by listing the top five "peaks" (ranked in terms of power) in the periodogram for each series, and marking the peaks which exceed the white-noise region. Most series only had one or two frequencies at which powerful peaks occur. The most powerful peaks were at (or very close to) an annual frequency (for 7 out of the 12 series). The periodograms revealed powerful 2 year cycles (peak frequency=.5 cycles per year) for "Attended a political rally" and "Worked for a political party". Petition signing seems to have a 6 month cycle (peak frequency=2 cycles per year).

[4]I haven't mentioned stationarity yet. Probably need to explain that both the ACF/PACF method, the regression method, and this method all require stationary time-series; and that I forced these series to be roughly stationary by removing the non-linear trend lines that I displayed earlier. Removing a linear trend would not have been enough.

[5]The periodograms shown employed smoothing over the 3 nearest neighboring points using a Daniell smoother [i.e. a moving average that gives half weight to the points at the end of the smoothing window] with a 10% taper.

[6]Actually "population" here is a bit misleading in the context of time-series. Rather, the idea is more that each of these series can be thought of as if it were sampled from an "ensemble" of series. Thus, the confidence intervals refer to the confidence one might have that the "ensemble" value is close to the estimated value.

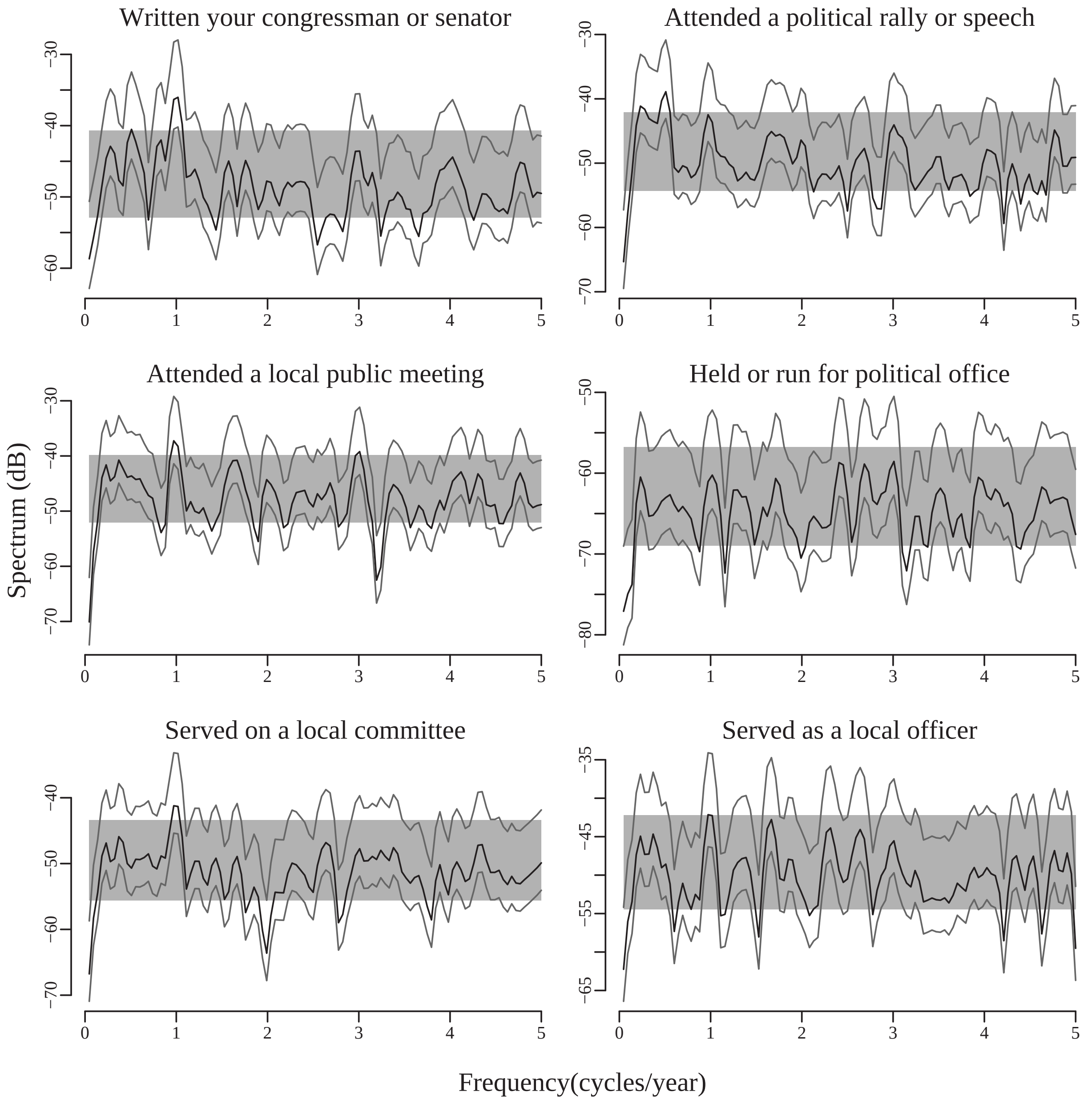
Table 1: Peak Frequencies (cycles/year) for Smoothed Periodograms

	Rank of Peak				
	1	2	3	4	5
Written your congressman or senator	**1.02**	**0.51**	0.83	0.28	3.01
Attended a political rally or speech	**0.51**	**0.23**	0.97	0.37	3.01
Attended a local public meeting	**0.97**	**3.01**	0.37	1.67	0.23
Held or run for political office	3.01	2.41	2.69	1.02	0.23
Served on a local committee	**0.97**	0.37	2.64	0.23	4.35
Served as a local officer	**0.97**	1.67	2.31	2.64	0.37
Written a letter to the paper	**0.97**	**0.37**	3.01	4.54	1.99
Signed a petition	**1.99**	**0.97**	0.37	3.01	1.62
Worked for a political party	**0.51**	2.96	1.99	2.31	1.02
Made a speech	**1.02**	0.37	3.01	1.67	2.64
Written a magazine/newspaper article	**1.02**	0.37	2.87	2.69	4.49
Been a member of better govt group	3.01	0.23	1.02	1.71	1.34

Bold numbers are peaks outside (or touching) the 95% white noise confidence interval.

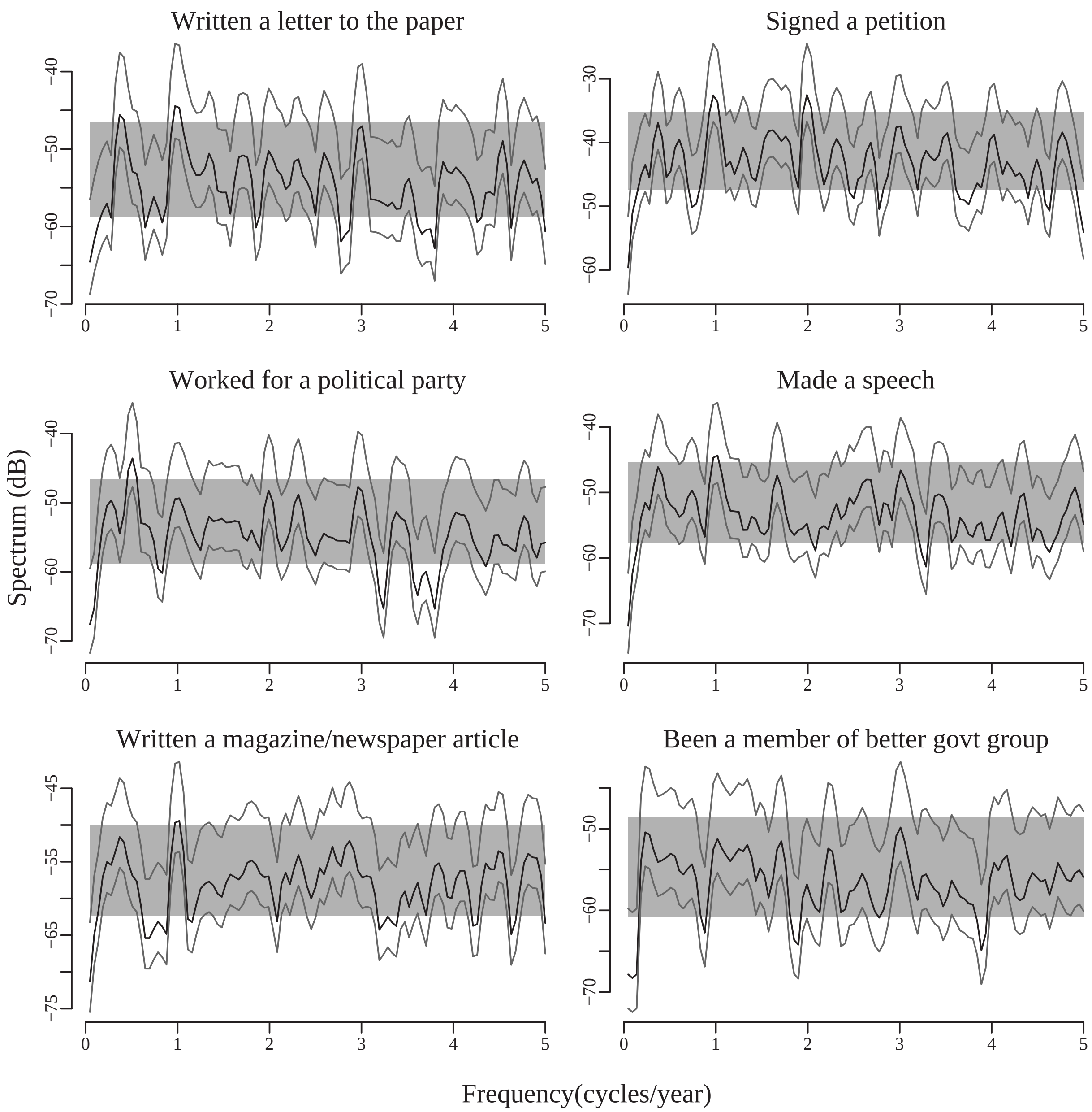
There is not much evidence of long/slow cycles [i.e. 4 year+ cycles]. Nor much evidence of short (within year) cycles. This is surprising given the fact that much of American political life is linked closely with elections. It does make sense that party work and rally attendance should be closely linked to a 2 year cycle — but inspection of the vote turnout figure earlier in the paper would have suggested much stronger 4 year cycles.

Figure 10: Smoothed Periodograms for Series 1-6



Note: The black lines represent the smoothed periodogram for each variable. The light gray region in each plot represents the 95% confidence region given a null hypothesis of white noise. The dark gray lines about the periodogram estimate represent the point-wise 95% confidence intervals for the estimated values.

Figure 11: Smoothed Periodograms for Series 7-12



6.1 Block Resampling

Although we know that the smoothed periodograms provide consistent estimates of the power of the periodic components of the series, in the case of real-world data analysis, it is always easy to wonder whether any given result should be trusted. Does the periodicity depend on the length of the series? Or where the series began? If this were a naturally stationary series, the answer would be clearly "no". In fact, these data are not naturally stationary, but were demeaned and detrended in order to do the analysis. Perhaps something about the specific composition of this series may be biasing the results.

One way to convince myself that the results are robust is to take pieces each series and to check the periodograms for each pieces. The idea is that, it the series is not stationary then periodograms calculated for different pieces would show different types of periodicity. If a series were stationary — such that the temporal structure in the series only depends on the relationships between the measurements NOT the actual moment in time that the measurements were made. The problem with such an approach would be that I would not be sure which pieces to choose, at what length, and where they should begin and end in the series. The obvious solution to this problem is to choose many such pieces, beginning and ending in many places. Davison and Hinkley (1987) suggest a number of such methods, known as "bootstrap" methods [?more on what the bootstrap is?]. However, they focus specifically on improving estimating of certain quantities, such as the coefficients for ARIMA models. They do also describe some techniques for estimating values in the frequency domain, such as $I(\lambda)$. But even they advise against using "phase scrambling." (page 411). They also describe procedure for calculating improved estimates of the confidence intervals for periodograms. However, my desire was whether the periodogram would even estimate *any* power at the same frequency within different subsections of the series rather than on a more robust estimate of the confidence interval around the current point estimates. Thus, inspired by Davison and Hinkley's chapter on "Complex Dependence," I implemented a form of block resampling in the following steps [repeated 1000 times per series] with the goal of checking that the periodogram of the whole is a reasonable representation of the periodicity in these series:

1. Choose at random a block length no less than 40 surveys and no greater than 207 surveys.
2. Choose at random a starting point between 1 and 208-40=168.
3. Calculate the frequency at which the smoothed periodogram has its maximum for the resulting block of data.

Figures 12 and 13 display density estimates with rug plots for the results of the block resampling.[7] For example, the block resampling in the case of writing to one's member of congress or senator overwhelming finds that the most powerful frequency is that 1 cycle per year over the 1000 re-samples of different blocks. There were two exceptions to this general confirmation that the artificial creation of stationarity was enough to truly allow the periodogram to work [really need to talk about stationarity up front — including note that I checked for stationary by fitting linear models to the series after subtracting out the nonlinear trend]. The two exceptions, "Served as a local officer" and

[7] [describe the density estimation]. "Rug plot" refers to the plotting of the actual values of a variable as small vertical lines on the x-axis of the plots, thereby providing a sense for the actual distribution of the results to complement the information provided by the density estimate.

"Signed a petition" had the 2nd highest peak switch with the first. Such that the strongest peak for serving as a local officer after the block resampling was at a six month cycle (2 cycles/year), and signing a petition had a much stronger annual cycle after the resampling.

Figure 12: Resampled Peak Frequencies: Series 1-6

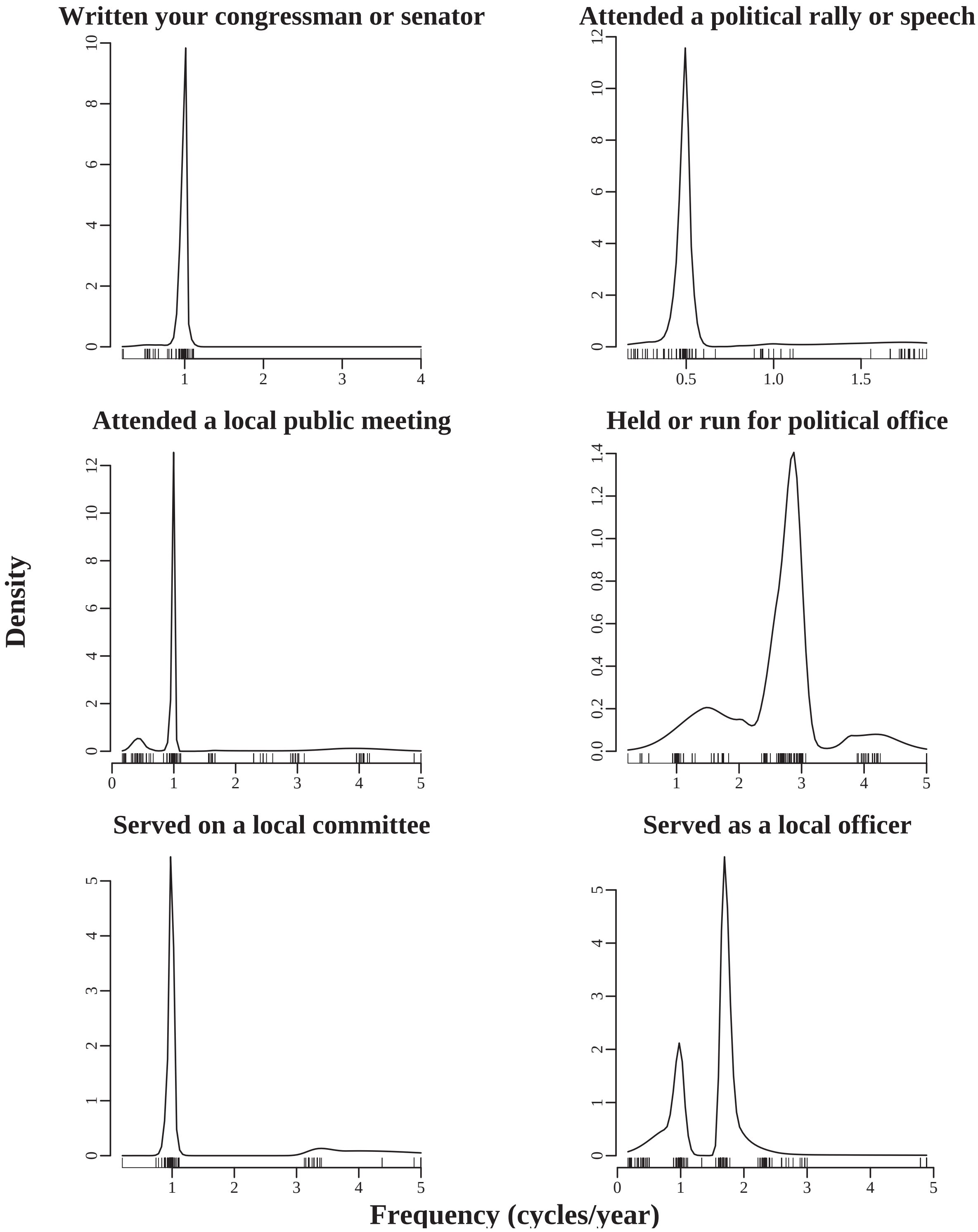
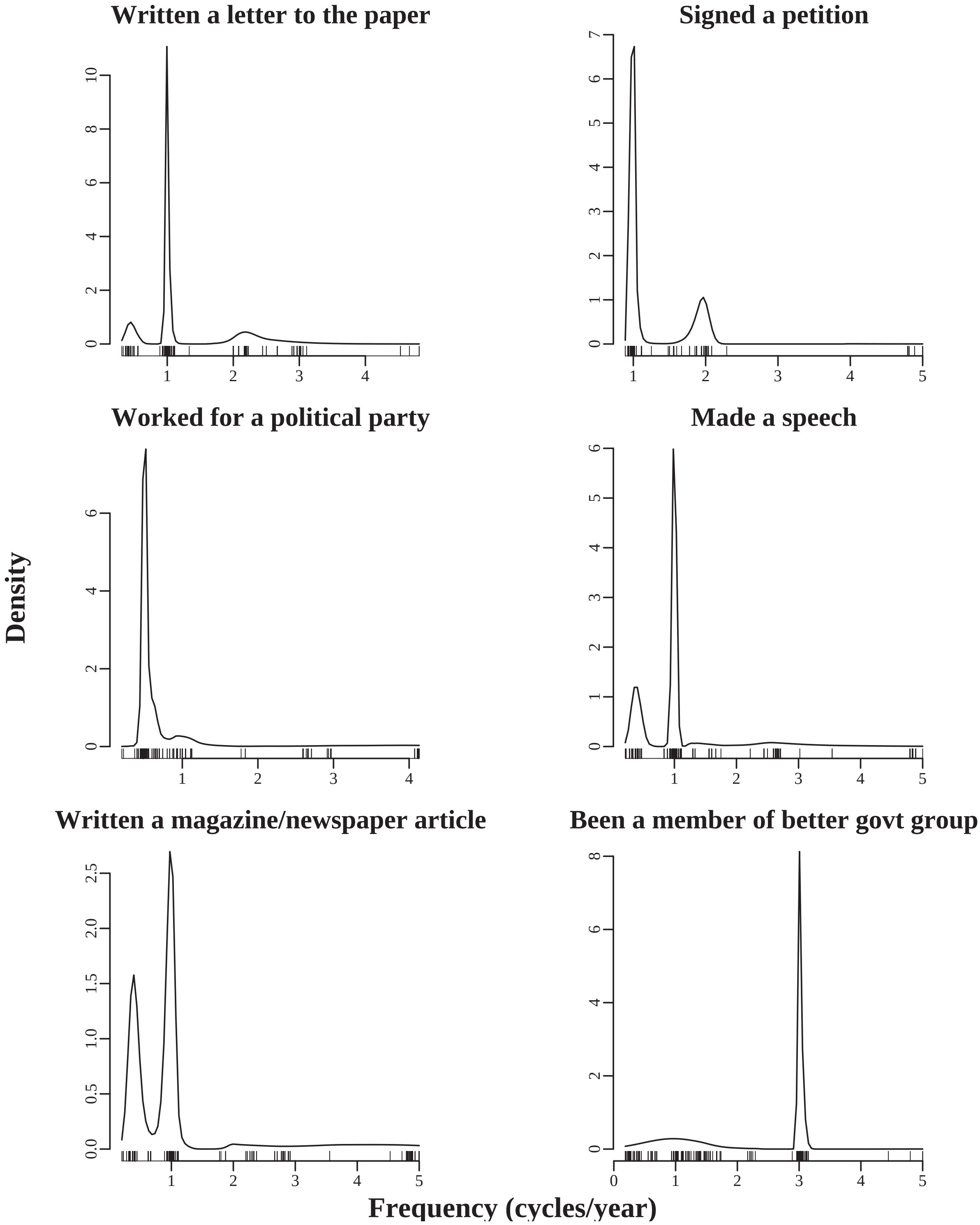


Figure 13: Resampled Peak Frequencies: Series 7-12



7 A Season for Participation

The finding of annual cycles is surprising. If there is a season for participation, when does it peak? How should we move back to the time-domain from the frequency domain?

For now, I will use predicted values from the regressions presented above to identify the months in which participation seems to peak. One might also decompose the series into seasonal or periodic pieces by smoothing over months controlling for smooth trend (See Loader (1999); Cleveland (1993); Cleveland et al. (1990)). One might also demodulate the series to construct a band pass filter [in essence to smooth over frequencies that are considered "noise".] (See, for example, (Bloomfield, 2000; Brillinger, 1988, 1981). For now, simple predicted values seem to work fine to show that participation peaks in the summer. Figures 14, 15, and 16 show the results. Each panel plots the actual, demeaned/detrended series in light gray and overlays the series that would have been predicted from the regression equation 3. At each point that the predicted values come to a peak (in a local sense), that point is labeled with the name of the month. Although my peak labeling algorithm was not completely successful at always labeling only the very highest peaks, it it still clear from these plots that the summer is when people seem to be reporting more participation.

Figure 14: OLS Predicted Values: Series 1-4

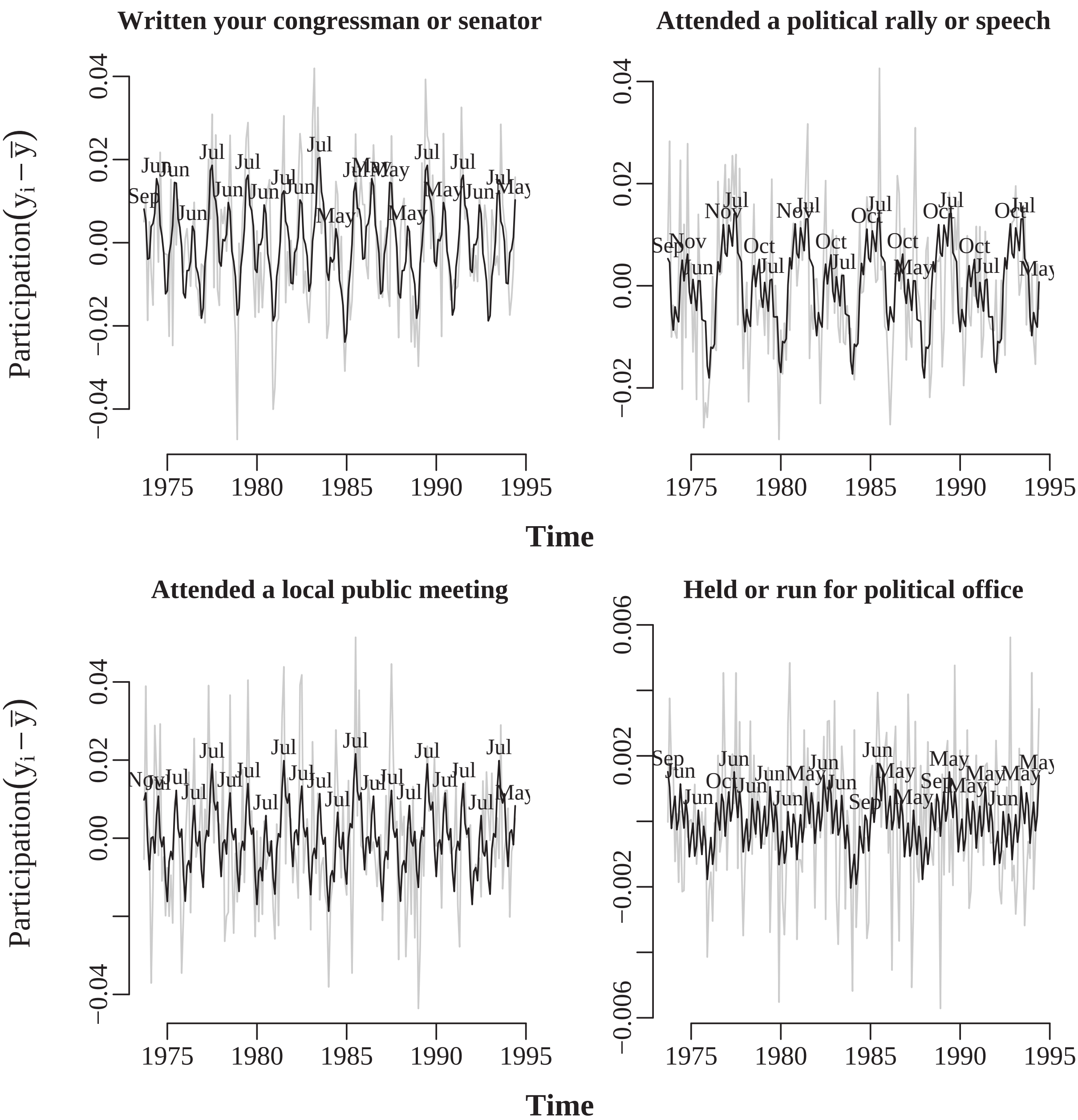


Figure 15: OLS Predicted Values: Series 5-8

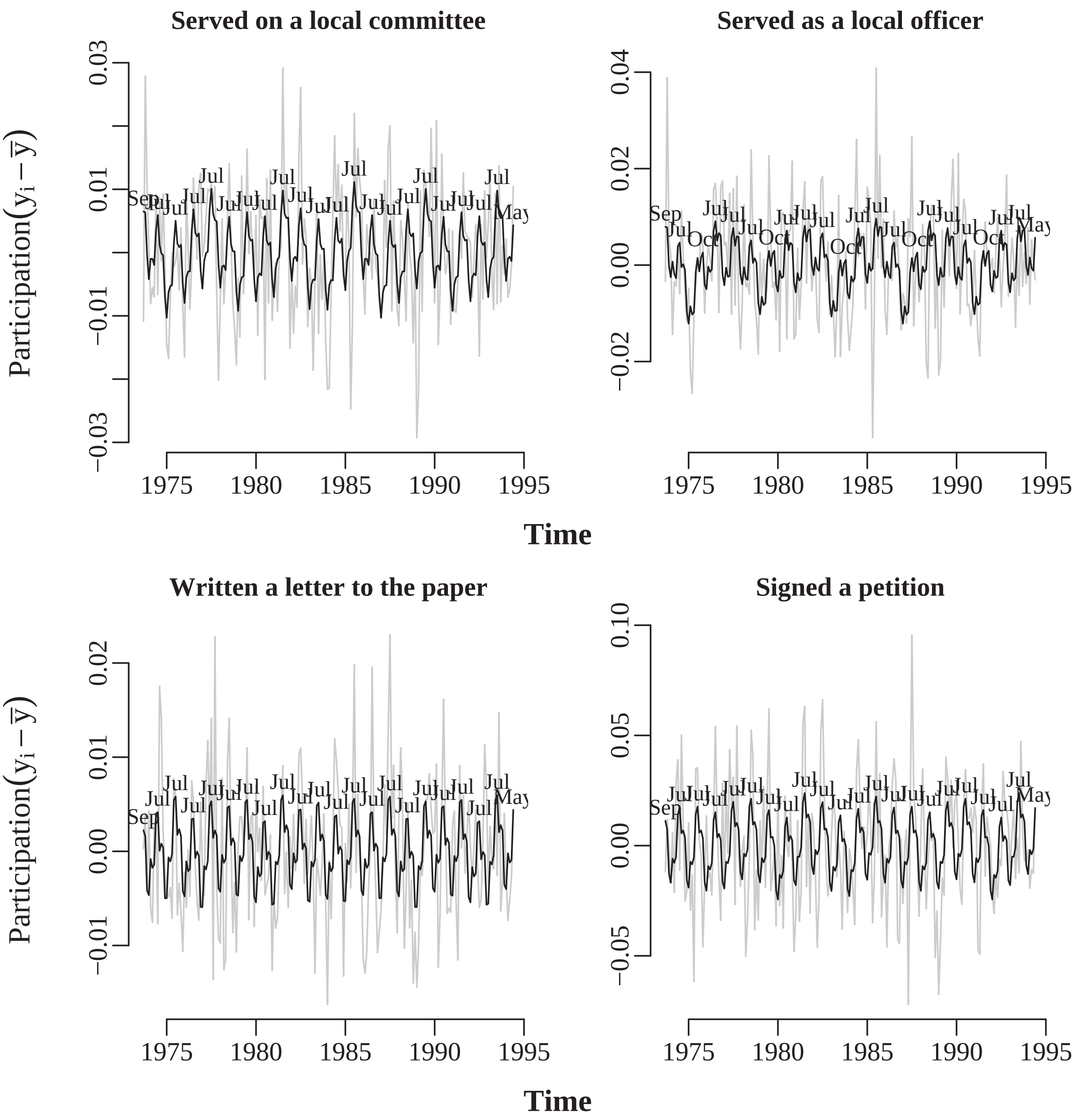
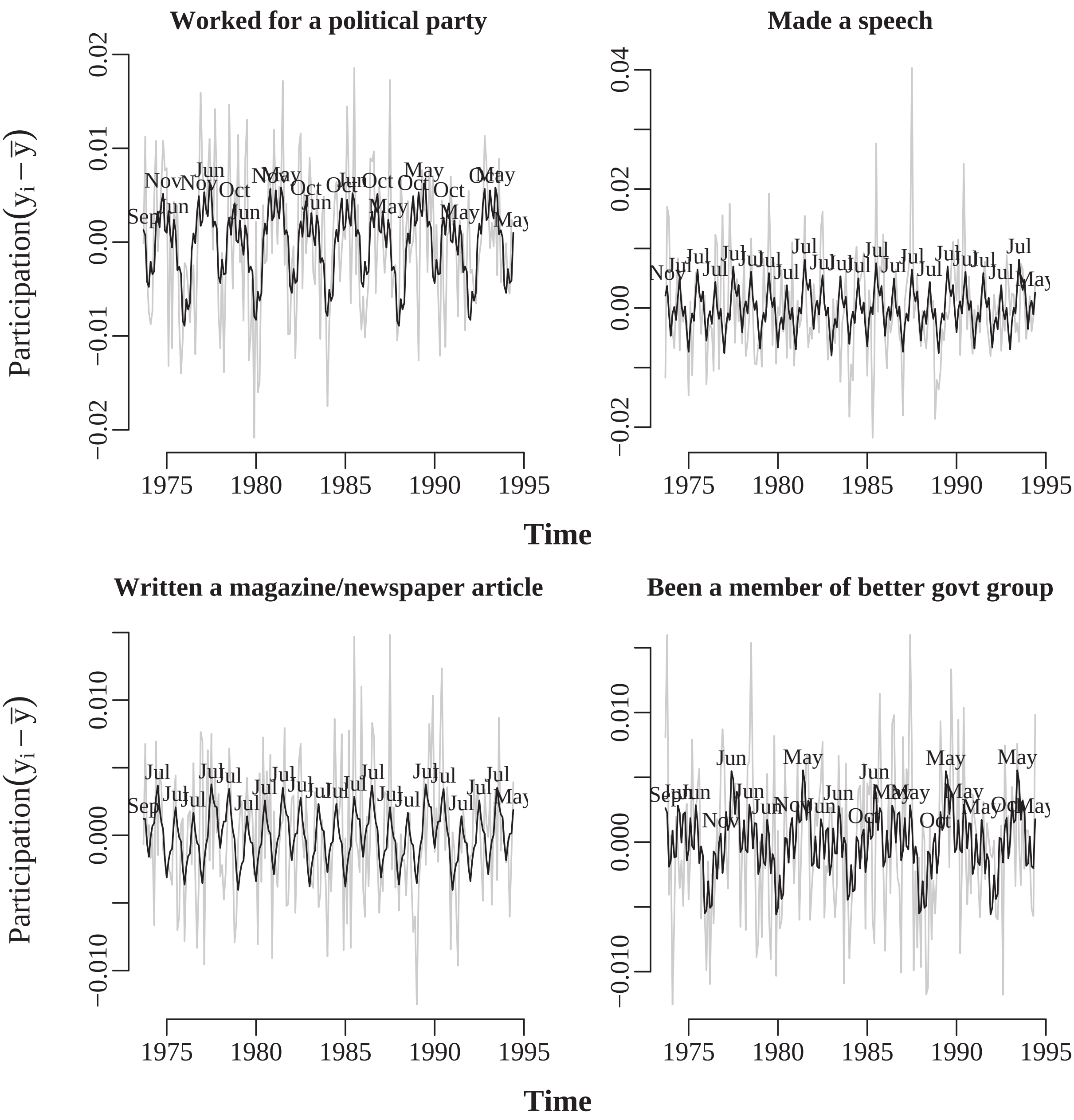


Figure 16: OLS Predicted Values: Series 9-12



8 What Drives Participation?

The story doesn't end there. In Rosenstone and Hansen (1993)'s account, this seasonality with peaks in the summer is evidence of close ties between moments of activity on Capitol Hill and the activity of citizens. They cite some very intriguing evidence — for example, they managed to get data on monthly volumes of mail sent from Congress, and they report that the volume of mail sent also peaks in the summer. Thus, they argue, the summer peak in participation (at least regarding letter writing) is an output where mail from Congress is the input. Unfortunately, they do not display the data on mail from Congress, nor do they try to relate it to the participation series in any analytic manner. I have begun to try to get data on 1) amount of mail received by Congress each month for the 1973-1994 period and 2) the amount of mail sent by the Congress over that period. Using the methods that I will present in this section, I ought to be able to analytically assess whether the aggregate participation (again, at least for letter writing) is an input or an output regarding the presumed peak of Congress' summer mailings.

While I wait for that data, I am skeptical that one would see *all* participation activities having an annual cycle peaking in the summer based on Congressional mailings — or even the moment at which there is most media coverage of Congressional debates and votes.[8] In fact, there is another, alternative explanation for what Rosenstone and Hansen report and which I found in the previous sections of this paper: that the sample of people answering the survey fluctuates in an annual cycle. Remember, that the actual survey question that was asked said:

> "Now here is a list of things some people do about government or politics. (HAND RESPONDENT CARD) Have you happened to do any of those things in the past year?

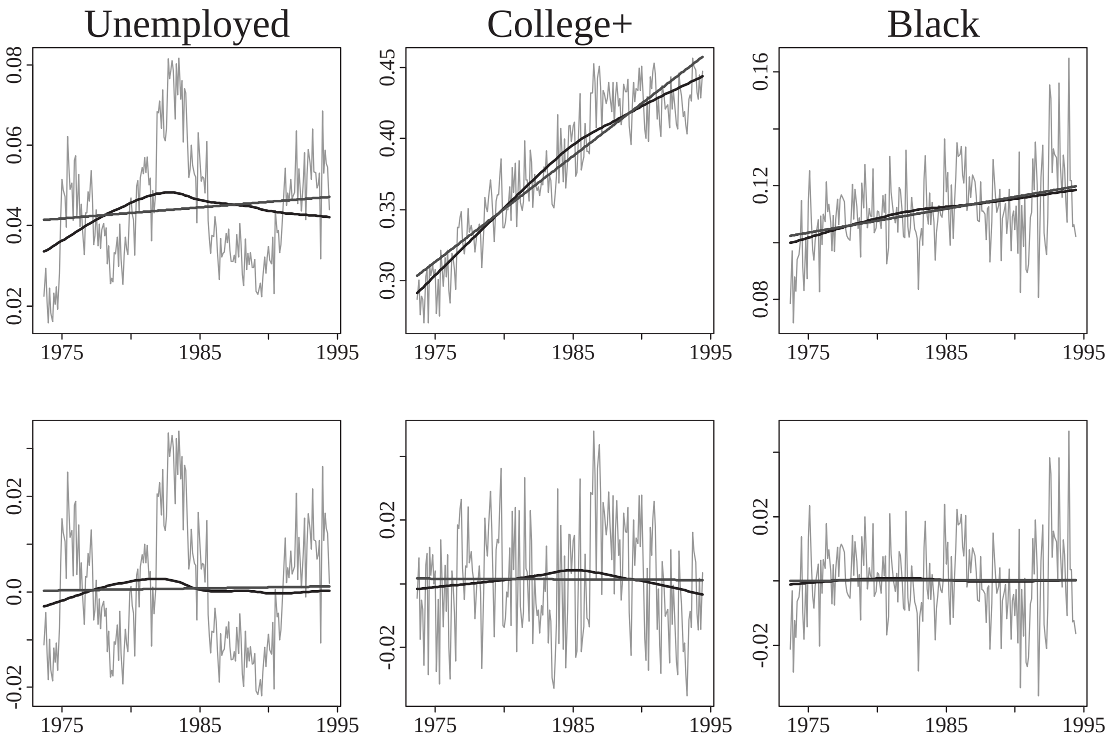
That is, the annual cycle that I've been discovering is in fact an annual cycle in reports about behavior over the past year NOT in reports of behavior at the moment of the survey. Rosenstone and Hansen (1993) realize this but argue, based on (Hansen and Rosenstone, 1983), that asking about the past six months and the past year produce the same results. Thus, they say that the Roper data are reflecting participation within a very short time of the interview data rather than participation across the whole year. The evidence they display in (Hansen and Rosenstone, 1983) is compelling, but as I check, I decided to see if there was any discernible relationship between the activities of participation and the composition of the sample. For sample composition, I only choose three variables: the proportion of the sample reporting that they were unemployed; the proportion of the sample reporting that they had a college degree; and the proportion of the reporting their race as African-American. I chose the measure of education and race based on previous research on political participation which emphasizes the importance of these two factors in predicting participation at the individual level (See,e.g. Verba, Schlozman and Brady, 1995). I chose unemployment because it knew it would have both short and long-term cyclical behavior — with the thought that those individuals who were caught at home by the Gallup researched in the summer might be different from those respondents who were at home and available to be interviewed at different months.

The top row of figure 17 shows the three variables in light gray with both linear and nonlinear trend lines in dark gray and black respectively. The second row shows the series after subtracting the

[8]Congress tends to use the spring in research and committee work and the summer for votes and debates (cite).

nonlinear trend. Although detrended unemployment series still shows some nonlinear relationship, it is very slight and the linear trend is flat. The medium-term periodicities in unemployment remain after removing the long-term trend. Although the upward trend in the proportion of respondents reporting a college education dwarfs any hint of periodicity in the top row, after removing the trend there does seem to appear some gentle cyclical behavior. The series containing the proportion of African-Americans in the sample seemed to be gently increasing over time, but the overall shape of this series does not change much after detrending.

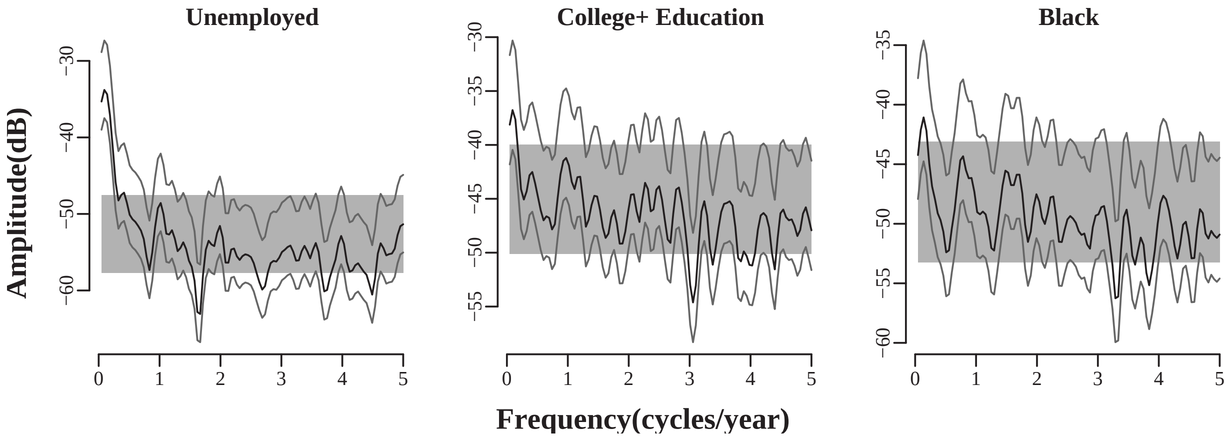
Figure 17: Change in the Roper Sample Over Time: Unemployment, Education, Race



Note: Unemployment is the proportion of people reporting that they had no job and were not students, retired, or housewives. College+ refers to the proportion of respondents reporting 13 or more years of education. Black refers to those respondents reporting African-American as their race.
The smooth lines are locally linear, and were estimated with a nearest neighbor bandwidth of .8 using locfit

Figure 18 displays the smoothed periodograms calculated for the three detrended sample composition measures. Each of them displays some very low frequency periodicity. Often high power at low frequencies can be an indication of non-stationary data. Fearing that the nonlinear detrending had somehow either induced trend or allowed some linear trend to remain, I also generated the periodograms for the series with a linear trend and mean removed so that I could guarantee lack of linear trend. However, the same pattern emerged. We saw in Figure 17 that unemployment appeared to have a perhaps 10-15 year cycle, the peak frequency ought be to be at roughly .1 (i.e. 1/10th of a cycle per year). The proportion of college education also looked to have a long cycle, which also seems to appear in the periodogram. The series for proportion African-American doesn't really display much power outside of the white-noise band at all.

Figure 18: Periodograms of Unemployment, Education and Race



9 Relationships between Sample Composition and Political Participation

There are two main methods for detecting relationships between multiple time-series: inspecting plots of cross-correlation functions [which are a bivariate form of the ACF] and inspecting plots of the squared coherency and phase of the two series [coherency and phase provide information about the frequencies at which two, or more, time-series may be most strongly related].

Figures 19 to 24 display the cross-correlation functions between each of the twelve participation acts and the three sample composition series. The dotted lines on each panel show the 95% confidence interval for the null hypothesis of white noise [i.e. no relationship between series]. There are two panels for each pair — the left panel shows the cross-correlation coefficients for negative lags, and the right panel shows the cross-correlation coefficients for positive lags [probably should redo this as one single panel per pair since the relationship at lag 0 is not clear from this plot]. One would expect about 1 out of every 20 coefficients to be stronger than the white-noise region merely by chance. Spikes at negative lags would indicate that the sample composition variable measured in previous surveys is having an effect on the present values of a given participation series. Spikes at positive lags indicate that the sample composition variable measured at subsequent, future, surveys seems to be strongly related to current values of the participation series.

Overall, there appears to be nearly no relationship between the participation series at the sample composition series. For example, there seems to be no relationship between the proportion of the population who are unemployed and the proportion of the population who attempt to contact their members of congress. There does seem to be a relationship between the proportion of the population attending rallies and unemployment [the top right two panels in Figure 19] — meaning that ?current attempts to contact members of congress seems to be related to unemployment as reported about 1.5 years in the future?. There is a similar relationship between petition signing and unemployment in the middle row, left two columns of Figure 20.

Other notable relationships include: the contacting members of congress preceding proportion college education in the sample by about 1.5 years; a strong relationship between the proportion of college educated in the sample and attendance at rallies [this relationship appears contemporaneous from Figure 21]. Overall, however, these plots do not tell us much about periodic relationships [although some of the cross-correlation functions do display seemingly sinusoidal patterns — more...]

[fix plots – titles and combine into single panels]

9.1 Cross-correlation Functions

Figure 19: Cross-Correlation Function: Unemployment with Series 1-6

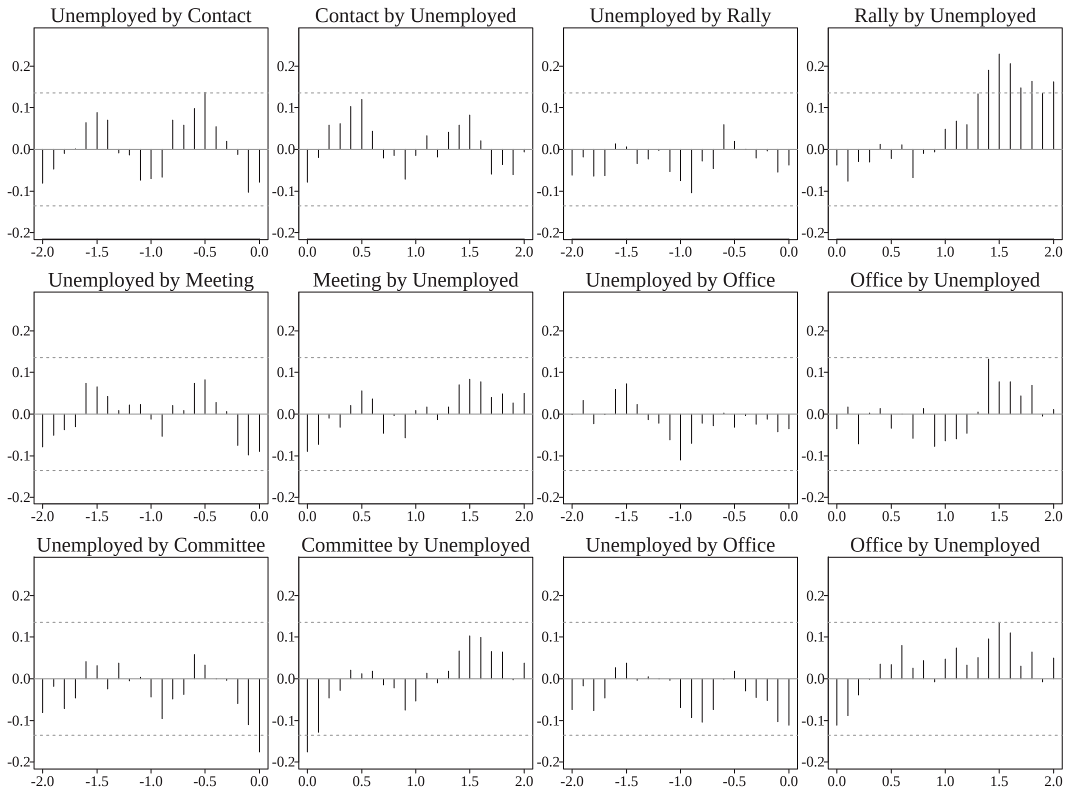


Figure 20: Cross-Correlation Function: Unemployment with Series 7-12

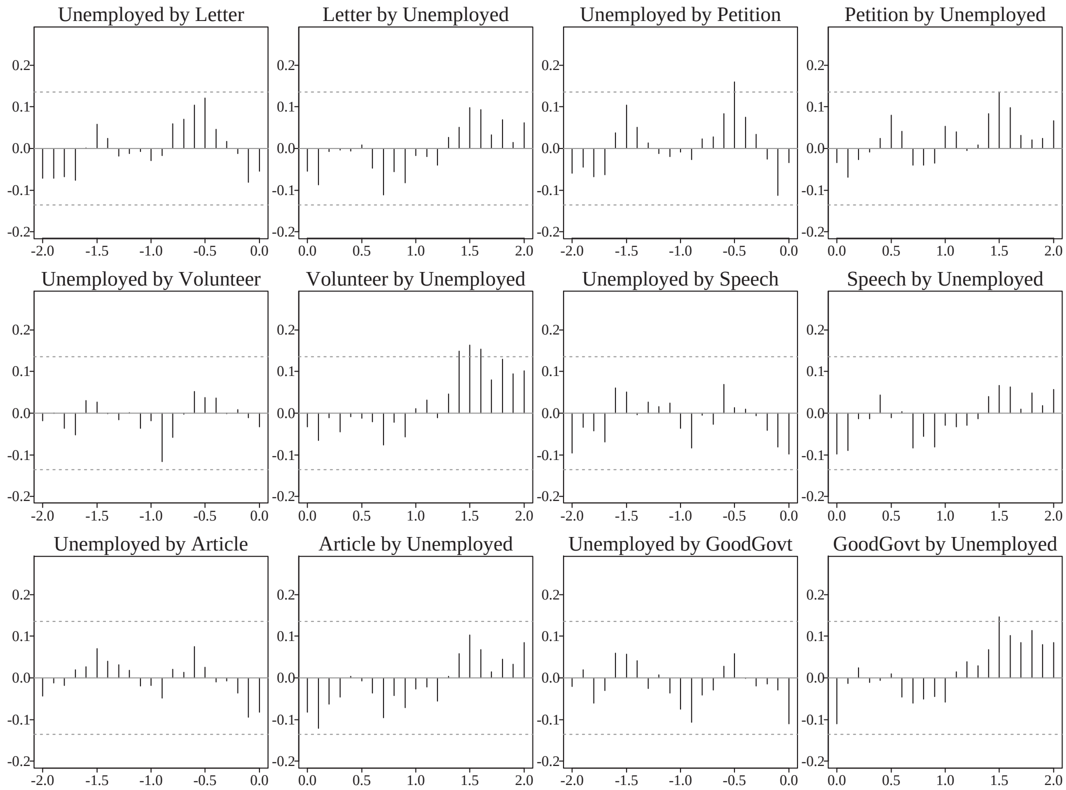


Figure 21: Cross-Correlation Function: College Education with Series 1-6 [fix titles]

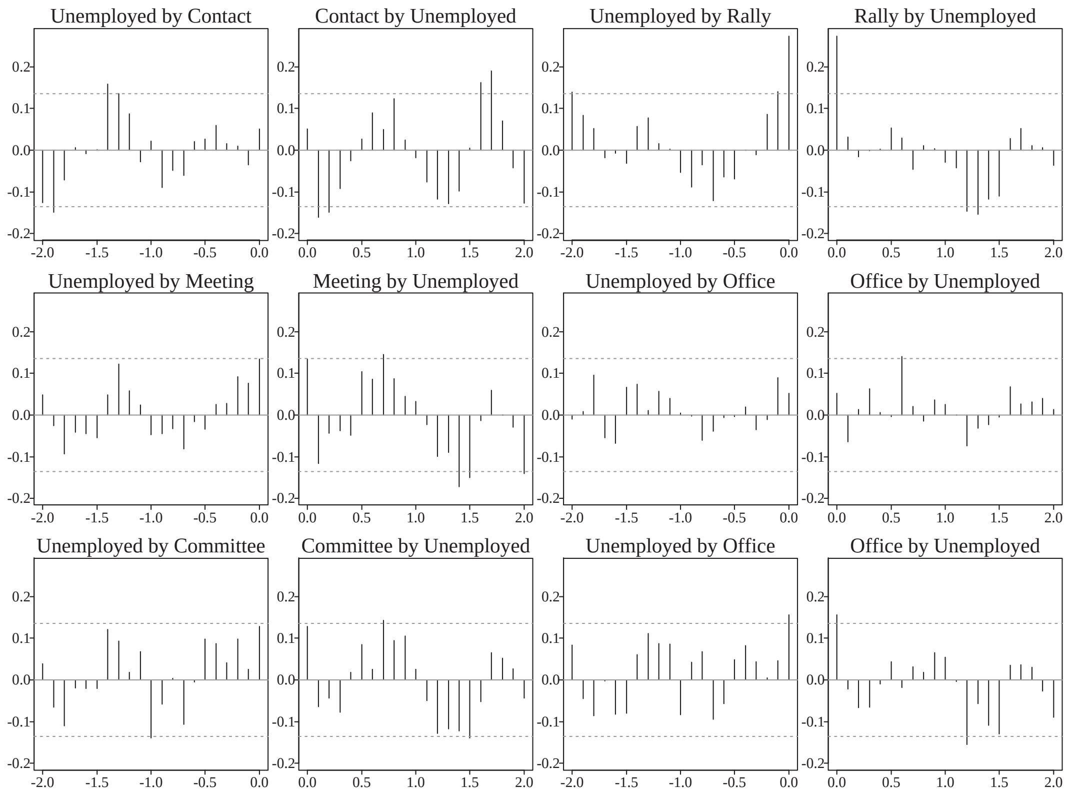


Figure 22: Cross-Correlation Function: College Education with Series 7-12 [fix titles]

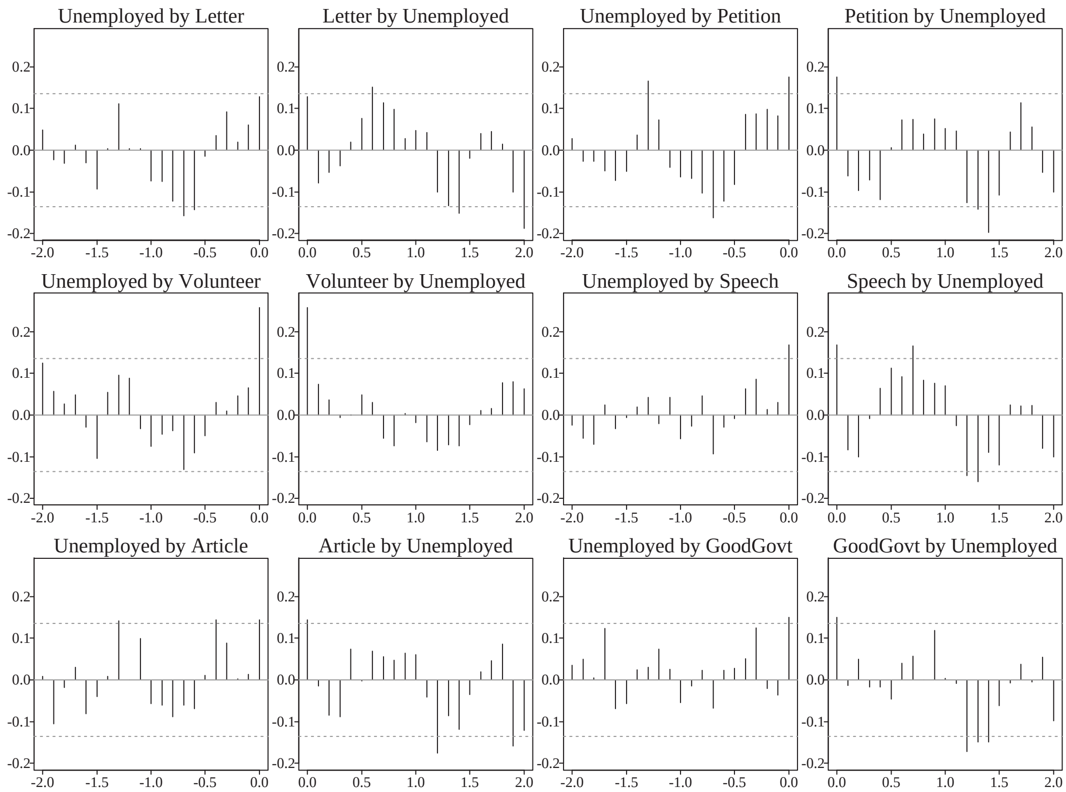


Figure 23: Cross-Correlation Function: Proportion Black with Series 1-6[fix titles]

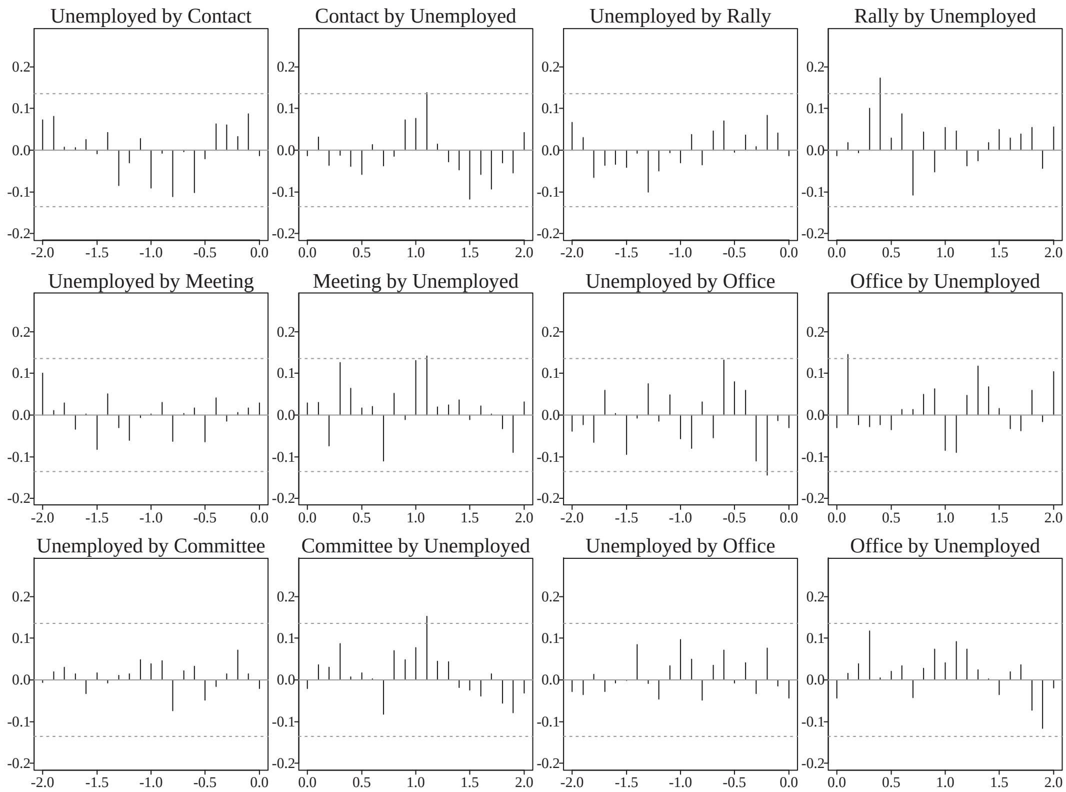
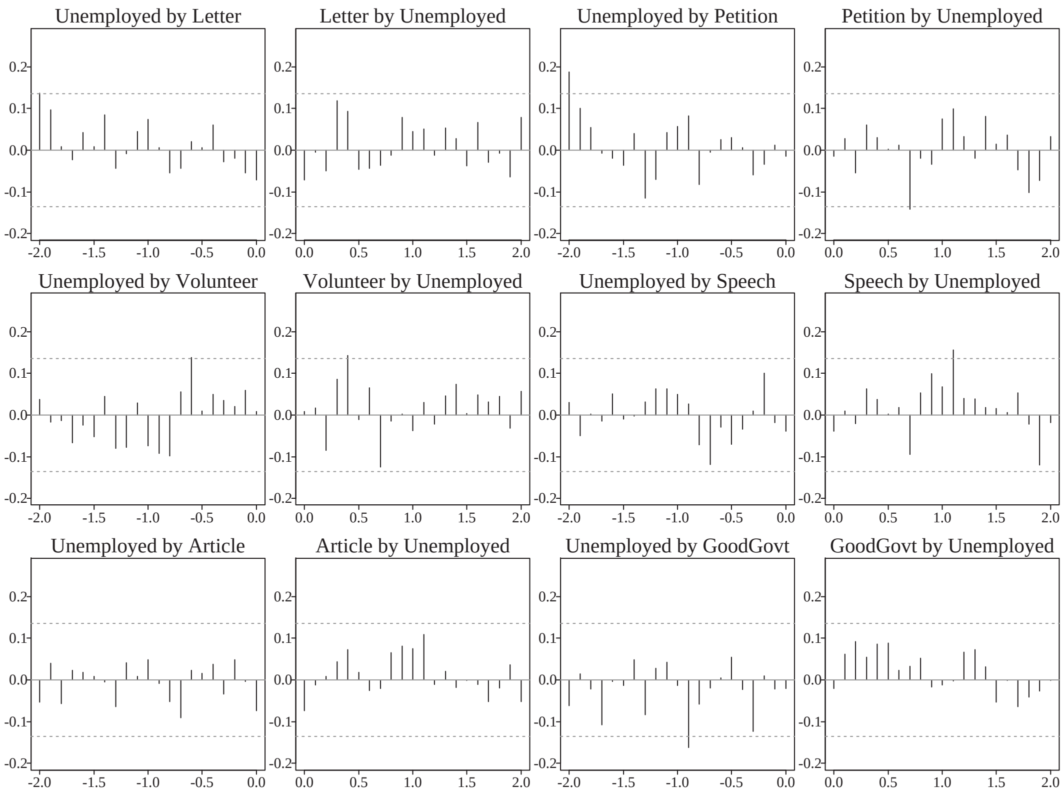


Figure 24: Cross-Correlation Function: Proportion Black with Series 7-12[fix titles]



9.2 Coherency

The frequency domain analog to the cross-correlation function is known as the "coherency" between series. The coherency provides estimates of linear time-invariant relationships between two stationary time-series. Just as the temporal structure in a univariate time-series can be represented by the "spectrum" of that series, as estimated by the discrete Fourier transform and displayed by the smoothed periodogram, so too can the relationship between two (or more) time-series be represented by the "cross-spectrum" which is estimated via the "coherency" and displayed as as the "squared coherency" and the "phase". [FIX: ?add discussion of spectrum up front to show that the periodogram is an estimate, and that the uncertainty estimates for the periodogram are based on the spectrum]

The cross-spectrum is usually defined as:

$$s_{X,Y}(\lambda) = \mathrm{E}(D(\lambda_X)\overline{D(\lambda_X)}), \tag{5}$$

where $D(\lambda)$ is the discrete Fourier transform (DFT) defined by the relation in (4). This time, the DFT is calculated for each of the two variables, say, $X(t)$ and $Y(t)$. The symbol $\overline{D(\lambda_Y)}$ refers to the complex conjugate of the DFT of $Y(t)$. The complex conjugate of a complex number $a + ib$ is defined as $a - ib$ and denoted $\overline{a + ib}$. A complex number multiplied times its own conjugate is a squared distance measure — i.e. it converts a complex number into a real valued magnitude measure.

The usual procedure for estimating the cross-spectrum is to use the *squared coherency*, which can be interpreted as the correlation between the random coefficients, $D(\lambda_Y)$ and $D(\lambda_X)$ and which can be written as follows:

$$\hat{\kappa}^2_{X,Y}(\lambda) = |\mathrm{Cor}(D(\lambda_X), D(\lambda_Y))|^2 = \frac{|\mathrm{E}(D(\lambda_X)D(\lambda_Y))|^2}{|\mathrm{E}D(\lambda_X)|^2|\mathrm{E}D(\lambda_Y)|^2} = \frac{|\hat{s}_{X,Y}(\lambda)|^2}{\hat{s}_{X,X}(\lambda)\hat{s}_{Y,Y}(\lambda)}, \tag{6}$$

where $\hat{s}_{X,Y}(\lambda)$ refers to the estimated cross-spectrum between the series $X(t)$ and $Y(t)$, $\hat{s}_{X,X}(\lambda)$ and $\hat{s}_{Y,Y}(\lambda)$ refer to the estimated univariate spectra of each of the series $X(t)$ and $Y(t)$. Basically, the squared coherency is the correlation between two series at a given frequency and it ranges from 0 [indicated no relationship] and 1 [indicated a strong relationship].

Following Bloomfield (2000)'s terminology, one can write approximate 95% confidence intervals for $\kappa_{X,Y}(f) \neq 0$ as:

$$\tanh(\mathrm{arctanh}(\hat{\kappa}_{X,Y}(f) \pm 1.96\sqrt{\frac{g^2}{2}})), \text{ where } g^2 = \frac{2}{\text{df from } \hat{s}_{X,Y}(f)} \tag{7}$$

And, approximate 95% confidence intervals for $\kappa_{X,Y}(f) = 0$ as:

$$1 - (1 - p)^{g^2/(1-g^2)}, \text{ where } p = .95 \tag{8}$$

That is, equation 7 shows how to generate pointwise confidence intervals around estimates of the squared coherency, given that the estimates are not zero. Equation 8 shows how to calculate confidence intervals for the situation where there is no relationship between the series, i.e. when the squared coherency is zero. [more here on normal approximation? or probably move all of this stuff to the appendix.]

Another important statistic summarizing information about the relationship between two stationary time-series in the frequency domain is the estimated phase of the bivariate (i.e. combined) series.

$$\text{estimated phase} = \hat{\phi}_{X,Y}(f) = \text{Arg}(\kappa_{X,Y}(f)) \tag{9}$$

[explain “Arg”, “tanh”, “arcsin” probably in the Appendix where all of this stuff should move]

Approximate 95% confidence intervals can also be written for the phase as:

$$\hat{\phi}_{X,Y}(f) \pm \frac{1}{2\pi} \arcsin\left[t_v(.05)\sqrt{\frac{g^2}{2(1-g^2)}\left\{\frac{1}{\hat{\kappa}^2_{X,Y}(f)} - 1\right\}}\right], \text{ where } t_v(\alpha) = 100\alpha\% \text{ of t-dist, and } v = 2/g^2 - 2 \tag{10}$$

The phase is only important where the squared coherency is not zero. However, where there does appear to be significant squared coherency between series, the shape of the phase indicates lagging and leading relationships between series. To the extent that it slopes down [over frequencies that have non-zero coherency] then the y_t series leads the x_t series [that fluctuations in y_t seem to precede fluctuation sin x_t], to the extent that it slopes upward [over frequencies that have non-zero coherency] then it indicates that the reverse [that fluctuation in the y_t series seem to follow fluctuations in the x_t series].

Figures 25 to 30 show plots of the squared coherency and phase for each of the three sampling composition variables with each of the twelve participation activities. A horizontal line is drawn at zero in each plot, and confidence intervals are plotted in light gray around the black lines indicating squared coherency and phase. [Add somewhere that one can only estimate coherency from a smoothed values of the periodograms – explain this]. The frequencies at which the squared coherency is estimated to be near 1 are those frequencies at which there appears to be a relationship between the two series. Similarly, in the regions in which coherency is high, the local slope of the phase can indicate leading versus lagging relationships.

For example, there appears to be a strong relationship around a cycle of one year between the proportion of respondents in the sample reporting writing to members of congress and the proportion of respondents in the sample reporting being unemployed. In that region the phase is less than 0 and slopes up. [what does it mean to be less than 0?] This positive slope indicates [I think!] that unemployment seems to follow writing to members of Congress!!?? Similarly holding or running for public office seems to be related to unemployment, but the phase is flat and negative – ?indicating that both series synchronize their fluctuations at an annual cycle?

[much more here — interpret each one]

9.3 Squared Coherency and Phase

Figure 25: Squared Coherency and Phase: Unemployment with Series 1-6

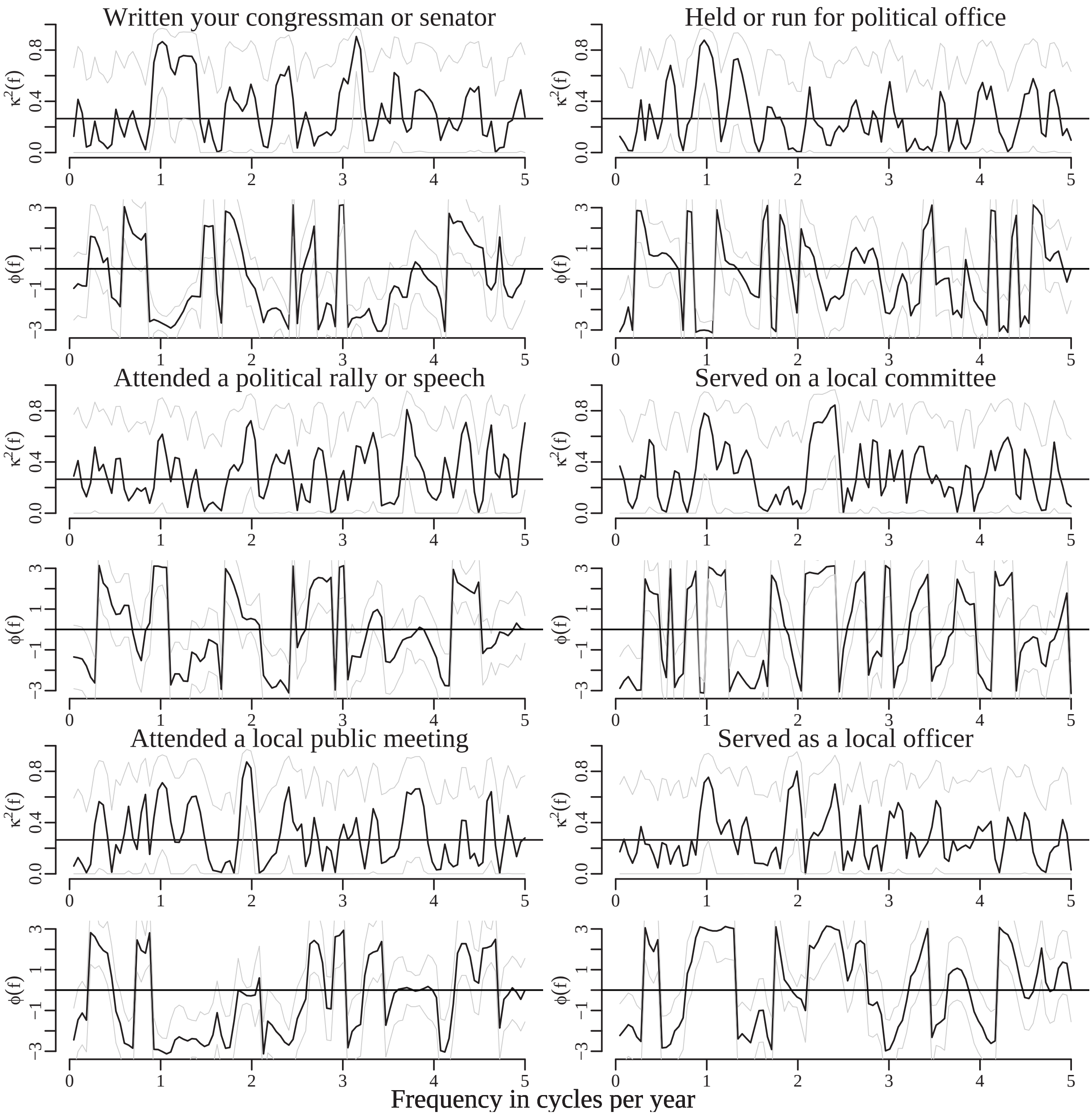


Figure 26: Squared Coherency and Phase: Unemployment with Series 7-12

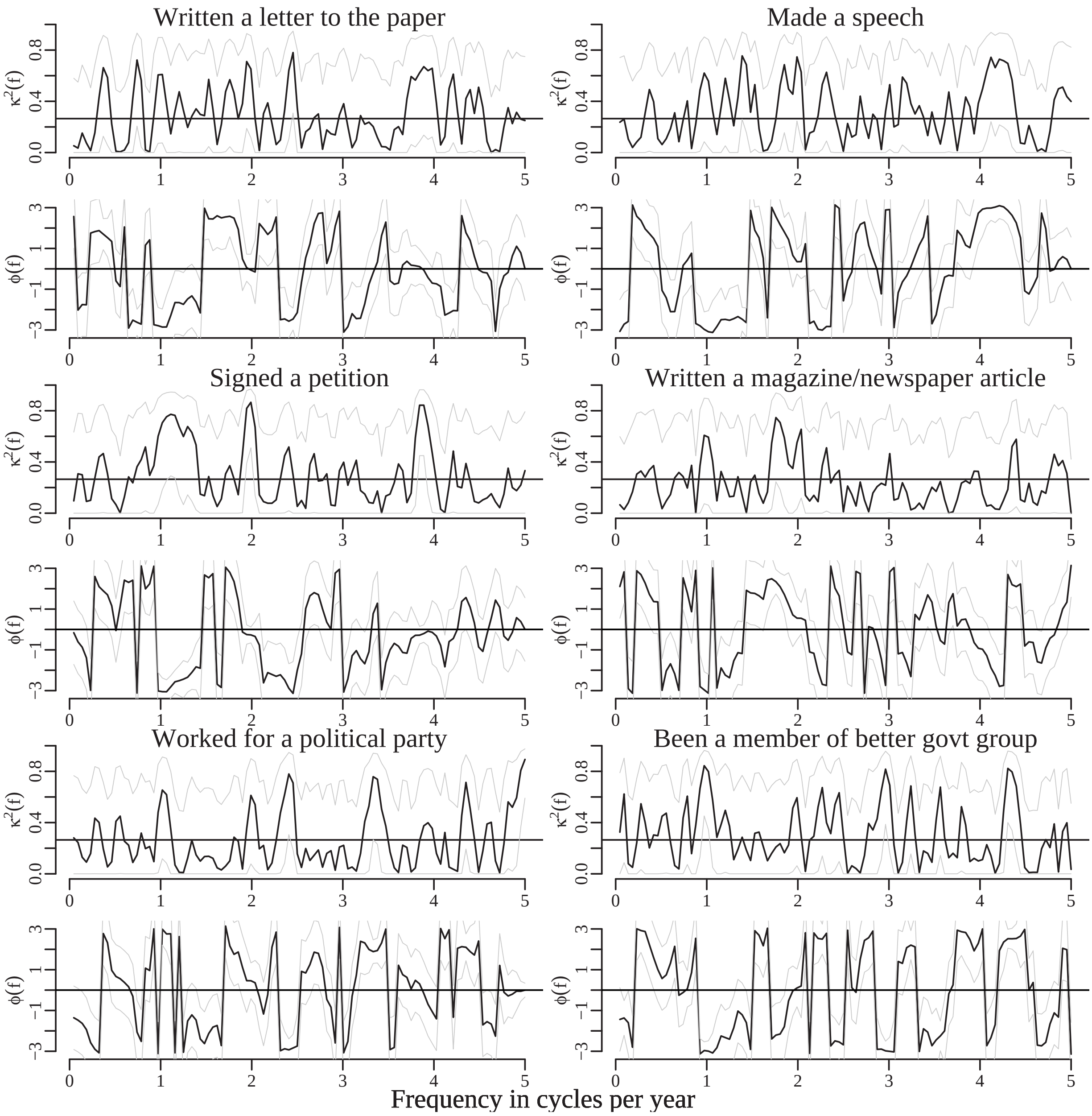


Figure 27: Squared Coherency and Phase: College Education with Series 1-6

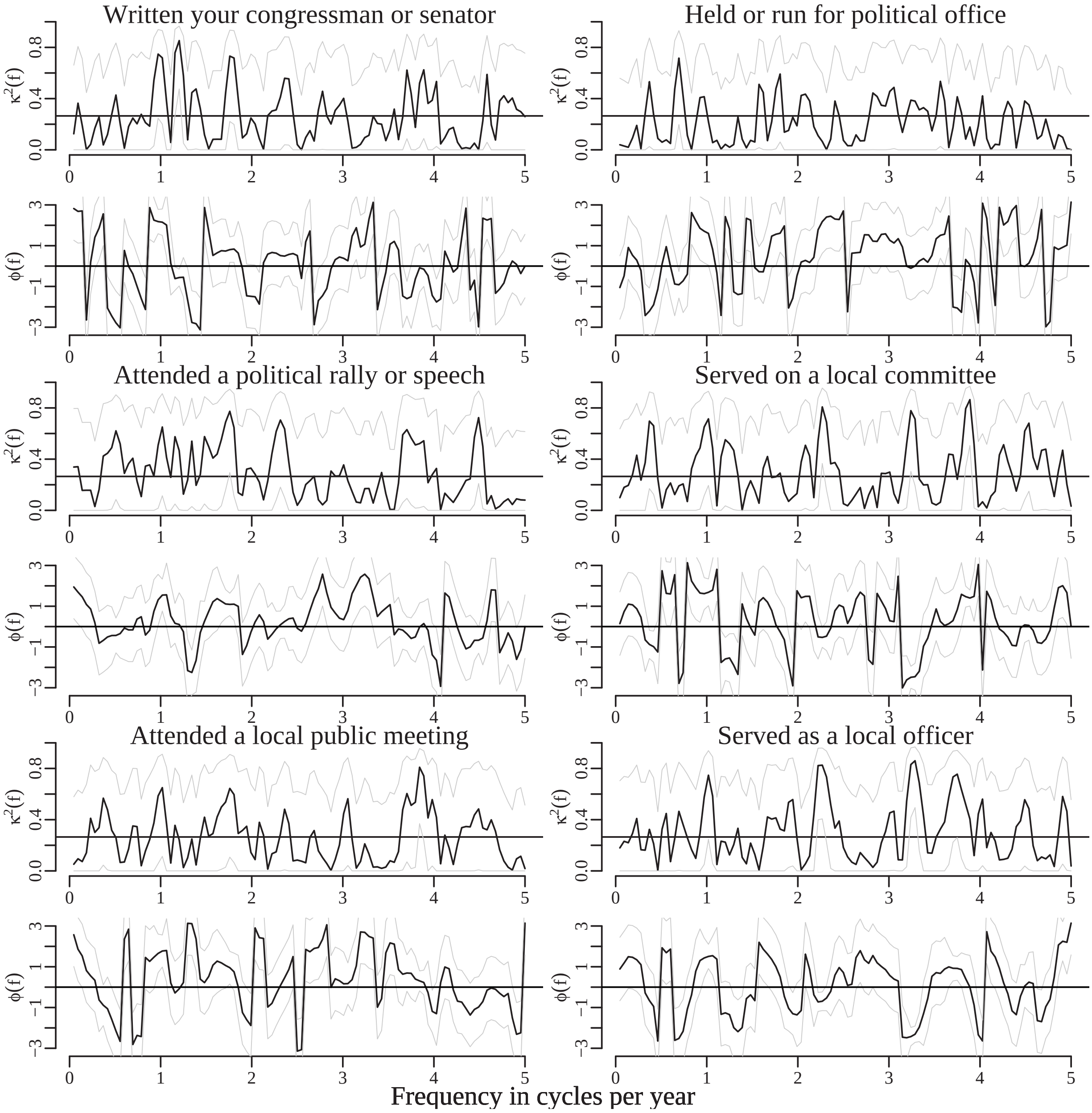


Figure 28: Squared Coherency and Phase: College Education with Series 7-12

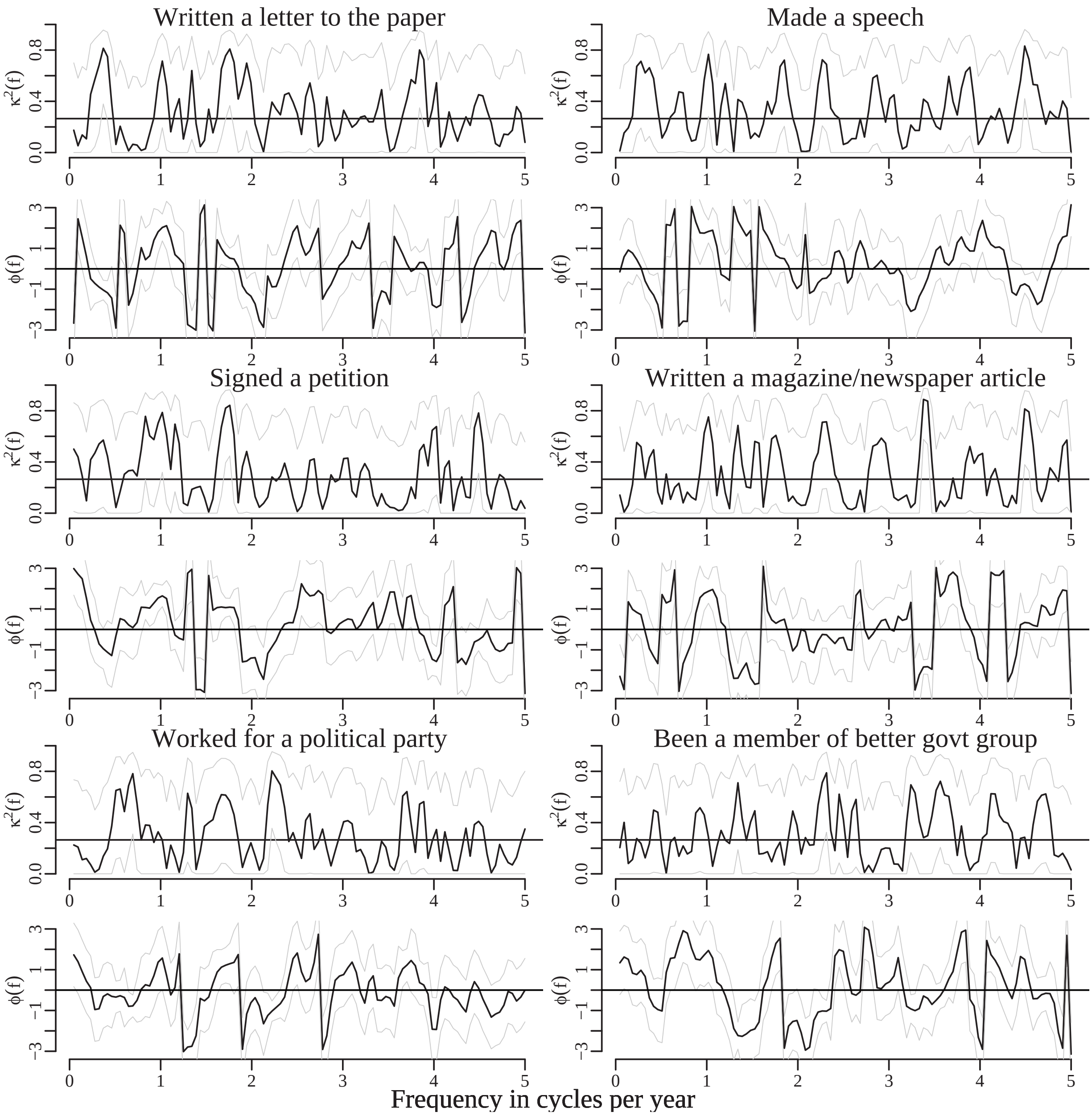


Figure 29: Squared Coherency and Phase: Proportion African-American with Series 1-6

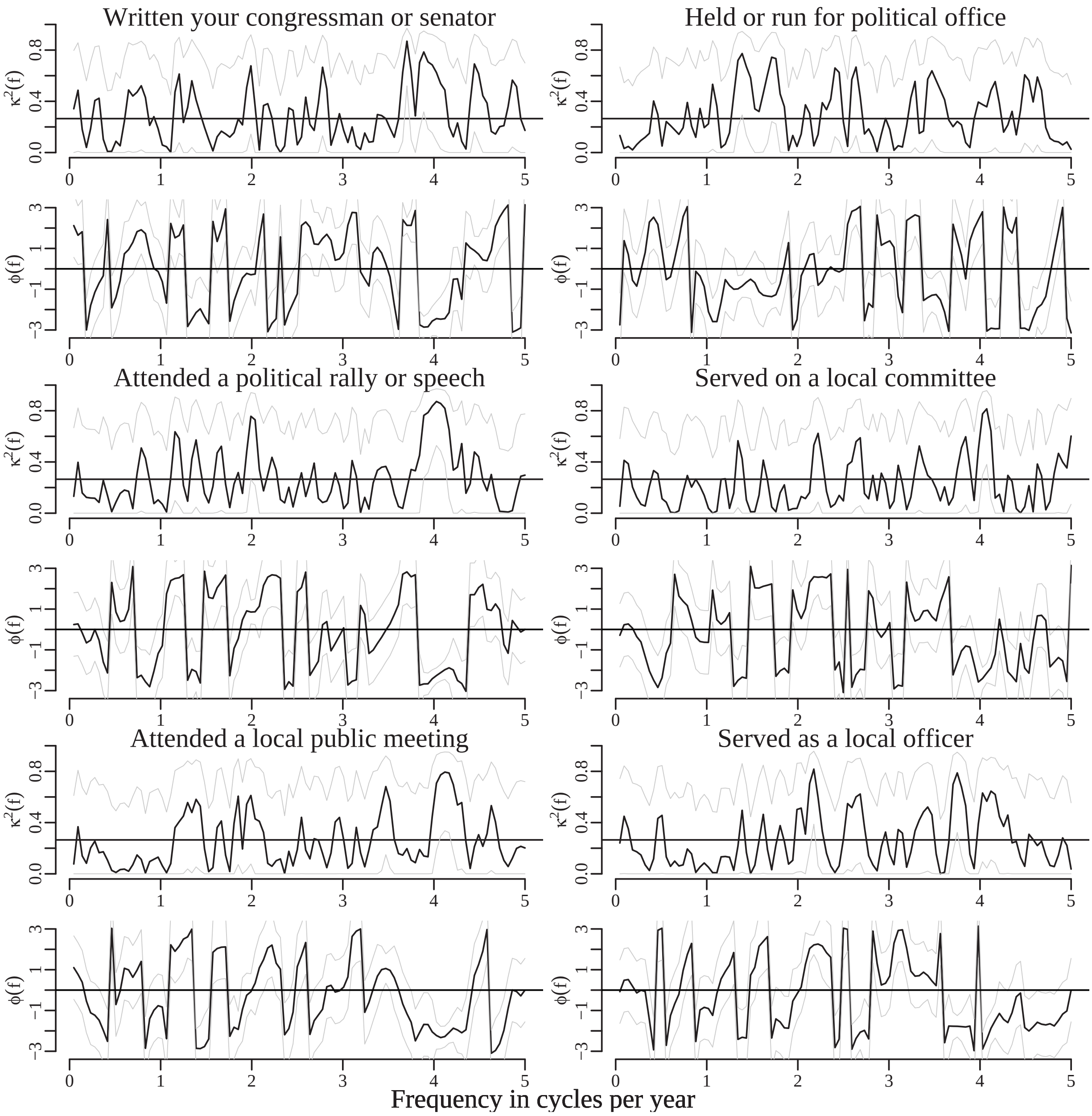
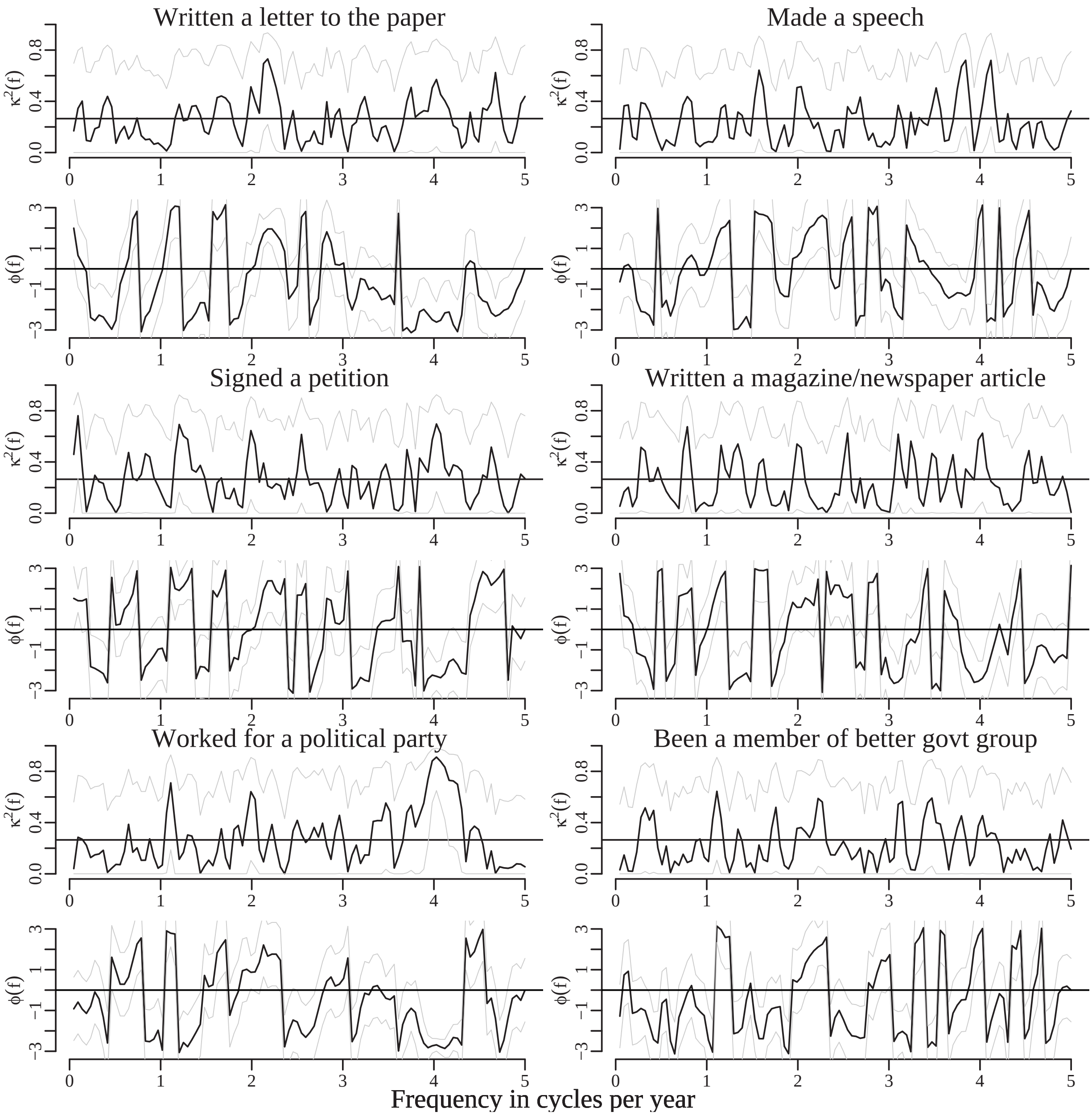


Figure 30: Squared Coherency and Phase: Proportion African-American with Series 7-12



10 Summary

> ...the change from the fifties to the sixties and then to the seventies and other such alternation in earlier periods raise the question whether our societies are in some way predisposed toward oscillations between periods of intense preoccupation with public issues and of almost total concentration on individual improvement and private welfare goals."(Hirschman, 1982, page 3)

This project is less ambitious and focuses on a shorter time period than that analyzed by Hirschman. From 1973 to 1994 aggregate political participation has decreased across a range of activities. Since so much attention has already been focused on this trend, this paper has explored another important characteristic of any time-series — periodicity.

Periodic oscillations definitely exist in this data, but they are predominantly annual peaking in the summer months.

The annual cycles are surprising given priors that electoral institutions ought to largely structure political participation. Previous inspection of the time-series also spurred one set of previous analysts to found an annual cycle,peaking in the summer, in 3 of the 12 activities analyzed here, and in a version of the time-series which ended in 1990 rather than in 1994 (Rosenstone and Hansen, 1993). Rosenstone and Hansen (1993) suggested that this annual cycle shows how tightly the political activity of Americans is bound to the movements of Congress and specifically the mobilization attempts surrounding the debates and vote on bills typical in the summer months. Since the actual survey questions ask about participation over the past year, this explanation sounds less plausible — and another explanation suggests itself — that the fluctuations are due to systematic variation in the sample composition in the Gallup samples. This paper has shown [I think!] that this is indeed the case.

I am currently trying to obtain detailed information on the monthly mail volume into and out of Congress. With this information, I will be able to test more directly if, despite the changes in sample composition of the Gallup polls, the political participation of Americans ought to be see as an "output" of Congressional mobilization or an "input" or in what way the flow of participation into Congress is related to the flow of mobilization out of it.

Appendices

A How it works

"Spectral analysis is essentially a modification of Fourier analysis so as to make it suitable for stochastic rather than deterministic functions of time."(Chatfield, 1996, page 105)

A.1 The Univariate Periodogram

Fourier analysis decomposes the variance of a time-series into periodic components. The main idea behind this analytic strategy is to approximate a stationary time-series by a trigonometric polynomial. For real valued time-series $Y(t)$ with $\mathrm{E}Y(t) = 0$ this Fourier polynomial can be written as follows:

$$Y(t) = a_0 + 2 \sum_{0<\lambda<T/2} \left(A(\lambda)\cos(2\pi\lambda t) + B(\lambda)\sin(2\pi\lambda t)\right)\left[+A(T/2)\cos(2\pi t T/2)\right], \tag{A1}$$

where the term in brackets is included only when T (the length of the series) is even.

The A and B terms in (A1) can be thought of as representing the "power" that oscillations at a particular frequency λ have in accounting for the variance of the series. Although sinusoids are not the only periodic functions available, their orthogonality (among other properties) has allowed the statistical properties of A and B to be easily developed.

It is often more convenient to express this polynomial using one complex number $D(\lambda)$ instead of the $A(\lambda)$ and $B(\lambda)$ above. This form depends on what is known as Euler's equation(Simon and Blume, 1994, page 883)[9] which shows that $e^{ix} = \cos(x) + i\sin(x)$. Equation A1 can be re-written as

$$Y(t) = \sum_{\lambda} D(\lambda)e^{-i2\pi\lambda t}, \tag{A2}$$

where $-T/2 < \lambda \leq T/2$. This equation gives a formula for the complex coefficients $D(\lambda)$. These coefficients are known as the Discrete Fourier Transform (DFT) of the time-series $Y(t)$ and are most often calculated via the Fast Fourier Transform (FFT) algorithm[10].

$$D(\lambda) = \sum_{t=1}^{T} Y(t)e^{-i2\pi\lambda t} \tag{A3}$$

The coefficients in (A1) and (A2) are related to each other by the following identities:

$$A(\lambda) = D(\lambda) + D(\lambda_{-j}) \tag{A4}$$
$$B(\lambda) = i\left(D(\lambda) + D(\lambda_{-j})\right) \tag{A5}$$
$$D(\lambda) = (A(\lambda) - iB(\lambda))/2 \tag{A6}$$
$$D(\lambda_{-j}) = (A(\lambda) + iB(\lambda))/2 \tag{A7}$$

(Stromberg, 1981, page 503), where the j and $-j$ refer to frequencies chosen on either side of 0.

[9] aka the Euler relation(Bloomfield, 2000, page 41)

[10] The FFT algorithm is available in most data analysis packages. In S the function is called fft

The main tool for displaying and analyzing the coefficients $D(\lambda)$ is the periodogram, which is built from the sequence

$$I(\lambda) = \frac{1}{T}|D(\lambda)|^2 \tag{A8}$$

The process of computing $I(\lambda)$ converts the complex numbers $D(\lambda)$ into real-valued magnitudes that we can compare with each other and which support hypothesis testing.

For all of the periodograms shown in this paper, λ was chosen to be of the form $f(j/T)$ where $j = 0, 1, \ldots, T/2$ and f is the sampling frequency of the data — 10 surveys per year. This scaling means that the periodograms are displayed for frequencies between 0 and 5 [meaning that the fastest cycle observable from these data would be 5 cycles per year=one cycle every 2 months]. For data with a sampling frequency of 1 (say, for annual observations) the periodograms would be shown for frequencies between 0 and .5 [meaning that the fastest cycle observable would be .5 cycles per year=a two year cycle].

It can be shown that the estimates of $I(\lambda)$ do not become less variable as the length of the series increases . To reduce this variability, most analysts work with a smoothed periodogram. In this paper, I averaged over the 3 nearest neighbors twice over.[11]

It can also be shown that if $Y(t)$ is a normal time-series, then the A and B coefficients will be jointly normal and that $D(\lambda)$ will have a normal distribution as well.[12] Based on these results, as $n \to \infty$, $I(\lambda)$ will be distributed χ^2_2. The smoothed periodogram will also have a χ^2 distribution but will have degrees of freedom proportional to the number of nearest neighbors.(Shumway and Stoffer, 2000, pages 235-242)

A.2 Squared Coherency and Phase

[insert stuff from text here! with better explanation especially about how to interpret the relationship between coherency and phase!]

There is a lot more to learn about frequency domain time-series analysis. If you want to know more, see the references.

B Notes on the Analysis

- In 1991, only 9 polls were done. In order to make the sampling frequency the same for all years, I created an observation for this survey, linearly interpolating the adjacent observations.

[11] This is known as a "modified Daniel" smoother which basically smooths over the 3 nearest neighbors twice in this paper.

[12] I inspected qq-plots for the detrended, demeaned series used in this paper. The sample quantiles map very closely onto the theoretical quantiles for a normal distribution.

- Smoothing of series was done nearest neighbor bandwidths of around .5 (i.e. roughly 50% of the data), using a local likelihood approach implemented by Loader (1999) in the S procedure locfit[13].
- Final calculation and presentation of the periodograms were done using the S function spec.pgram with 10% tapers and smoothing over the 3 nearest neighboring points using a Daniell smoother [i.e. a moving average that gives half weight to the points at the end of the smoothing window]
- The block resampling was done by me. The code for this and other aspects of the analysis of this paper will be available at `http://socrates.berkeley.edu/~jbowers`.

C Notes on the Data

The Pew Charitable Trusts and the National Science Foundation provided funds to allow Henry Brady, Robert Putnam and Andrea Campbell to buy the series of 207 surveys from Roper, and for them, plus Dorie Apollonio, Laurel Elms, and Steven Yonish to assemble the studies into a single file of 410,116 respondents.

Each survey aimed to be a representative sample of the population of the Continental United States, age 18 and up — excluding people in prisons, Army camps, nursing homes, etc... The sample proceeded by stages: In the first stage, 100 counties were selected at random within strata determined by population size within geographic region. In the second stage, cities and towns were drawn at random within counties, proportionate to population. Third, blocks (or rural routes) were drawn at random proportionate to population within urban areas or at random in rural areas. The method for selecting household within block is not as clear, the literature states: "A specified method of proceeding from the starting household was prescribed at the block level."

The sampling design included quotas for men and women over and under age 45, and quotas for employed people. Also, interviewing on half of the assigned blocks had to be conducted after 5pm on weekdays or else on Saturday or Sunday.

References

Beck, Nathaniel. 1991. "The Illusion of Cycles in International Relations." *International Studies Quarterly* 35(4):455–476.

Bloomfield, Peter. 2000. *Fourier Analysis of Time Series: An Introduction.* 2nd ed. New York, NY: John Wiley and Sons.

Brillinger, David R. 1981. *Time series : data analysis and theory.* San Francisco, CA: Holden-Day.

Brillinger, David R. 1988. "Some Statistical Methods for Random Process Data from Seismology and Neurophysiology (in The 1983 Wald Memorial Lectures)." *Annals of Statistics* 16(1):1–54.

[13]The analyses in this paper were mostly done using an open-source version of S known as R.

Chatfield, C. 1996. *The Analysis of Time Seris: An Introduction.* 5th ed. London, UK: Chapman and Hall.

Cleveland, Robert B., William S. Cleveland, Jean E. McRae and Irma Terpenning. 1990. "STL: A Seasonal-Trend Decomposition Procedure Based on Loess (with Discussion)." *Journal of Official Statistics* 6:3–73.

Cleveland, William S. 1993. *Visualizing data.* Summit, NJ: Hobart Press.

Davison, Anthony C. and David V. Hinkley. 1987. *Bootstrap Methods and Their Application.* Cambridge: Cambridge University Press.

Hansen, John Mark and Steven J. Rosenstone. 1983. Participation Outside Elections. Technical report National Election Study. A National Election Studies Pilot Study Report.
URL: *http://www.umich.edu/ nes/resources/psreport/abs/83b.htm*

Hirschman, Albert O. 1982. *Shifting Involvements Private Interest and Public Action.* Princeton, NJ: Princeton Univ Press.

Loader, Clive. 1999. *Local Regression and Likelihood.* New York, NY: Springer.

Rosenstone, Steven and John Mark Hansen. 1993. *Mobilization, Participation and Democracy in America.* MacMillan Publishing.

Shumway, Robert H. and David S. Stoffer. 2000. *Time Series Analysis and Its Applications.* Springer.

Simon, Carl P. and Lawrence Blume. 1994. *Mathematics for Economists.* New York, NY: W.W. Norton.

Stromberg, Karl Robert. 1981. *An Introduction to Classical Real Analysis.* Belmont, CA: Wadsworth International.

Verba, Sidney, Kay Lehman Schlozman and Henry Brady. 1995. *Voice and Equality: Civic Voluntarism in American Politics.* Cambridge: Harvard University Press.