DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

AGGREGATION AND DYNAMICS OF SURVEY RESPONSES: THE CASE OF PRESIDENTIAL APPROVAL

R. Michael Alvarez

Jonathan N. Katz

SOCIAL SCIENCE WORKING PAPER 1103

September 2000

Aggregation and Dynamics of Survey Responses: The Case of Presidential Approval*

R. Michael Alvarez †
California Institute of Technology

Jonathan N. Katz‡
California Institute of Technology

September 22, 2000

ABSTRACT

In this paper we critique much of the empirical literature on the important political science concept of presidential approval. Much of the recent research on presidential approval has focused on the dynamic nature of approval; arguments have raged about whether presidential approval is integrated, co-integrated, or fractionally integrated. We argue that none of these time-series concepts, imported from an econometrics literature which has fundamentally different types of data than do political scientists, can apply to the presidential approval time series. Instead, we advocate careful use of aggregated approval as a time-series cross-section, or the use of individual-level survey responses. Ultimately most of the important hypotheses political scientists wish to test regarding presidential approval involve individual voters or citizens; thus we argue that using the appropriate data unit is the best methodology.

*Previous versions of this paper were presented at the 1996 Midwest Political Science Association Meetings and at the 1997 Southern California Political Methodology Program Conference at UCLA, and we thank participants in these conferences for their comments. We are grateful for discussions with David Grether. We thank Neal Beck and Simon Jackman for answering our questions about the presidential approval series, and sharing their data. We also thank John Aldrich for making available to us the 1988 Gallup Organization polling data. Katz thanks the John M. Olin Foundation for a faculty fellowship supporting his work.

†Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125; rma@hss.caltech.edu.

‡Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125; jkatz@caltech.edu

Presidential approval has been a central concept in the study of both presidential power and public opinion. With the advent of the "new presidency" in the age of mass media politics, having high levels of approval is seen as an important political resource for presidents (Kernell 1986). Having high levels of approval is thus a central component of presidential power (Neustadt 1990), and influences electoral outcomes and legislative success (Brody 1991; Rivers and Rose 1985; Simon and Ostrom 1989).

Since the early 1970's, a long list of articles and books have examined presidential approval, almost exclusing from a time-series perspective. In general, these scholars have examined changes in the percentages of survey respondents who claim to approve of the job which the current president is doing as a function primarily of "rally events", economic conditions, and various time-related effects (Beck 1991, 1992; Brody 1991; Kernell 1978; Kiewiet and Rivers 1985; MacKuen 1983; Ostrom and Simon 1992; Ostrom and Smith 1992; Smith 1992; Williams 1992). More recent work on presidential approval has focused intensely on the dynamic properties of this important concept. Ostrom and Smith (1992), for example, proposed an error correction model for the presidential approval series based on their results which indicated that this time series is integrated, or near-integrated (at least for the Reagan years).

In this paper, we challenge this approach to studying presidential approval. Our first concern is that the dynamics attributed to the aggregate presidential approval series are often logically inconsistent and always substantively implausible. The most problematic models of the dynamics presidential approval assume that it is an integrated series. As we will show, there is no way for a bounded series, such as the approval series, to be integrated. However, even the dynamic models that are not integrated, such as the fractionally integrated and standard ARMA models, lead to substantively odd findings. For example, if we are to believe these models, the boost in approval Bush received during the Gulf War was also helping to increase Clinton's approval twenty-four months later.

This puzzling spillover effect is caused by aggregation, which is our second major concern. Most studies have assumed that the presidential approval series is in effect a single, long series. However, it is not clear from any theory why we should think that approval should pool across presidential administrations. In fact, when we turn to our empirical examination of the Gallup presidential series, the most commonly used series, it seems that the dynamics of approval seem to vary significantly between administrations.

Pooling across administration, however, is no the only aggregation problem that we find in the current studies of presidential approval. There is also the problem of aggregating up from the individual survey respondences. The theories we are interested in testing are usually stated at the level of the individuals, however it is only rarely the case that we can test such individual level claims with aggregate data, particularly when there is heterogeneity in the population.

The paper proceeds as follows. In the first consider the dynamics of the aggregate presidential approval, assuming this is a valid quantity of interest. In this section, we also briefly review the basics of integration. The next section, then turns to the problem of aggregating across administration. The third section, then considers the issue of aggregation bias and heterogeneity at the level of the respondent using actual individual level data from the Gallop Organization. The final section concludes.

1. THE DYNAMICS OF PRESIDENTIAL APPROVAL

Most recent studies that have examined presidential approval data, typically from the Gallup Organization, start with the implicit or explicit assumption that it is a single time series. The central focuses is the dynamic behavior of the series, particularly on its mememory properties. This typically boils down to whether or not the presidential series is “integrated” (and possibly co-integrated with some other series, such as various measures of the economy). In this section we argue that even under the maintained assumption that the approval series is a single time series, an assumption that we will question below, it is logically impossible for the approval series to be integrated. We start this section by briefly reviewing what it means for a series to be integrated. We then present an argument for why presidential approval can not be integrated. We conclude the section by examining the case where the series is not integrated, but long memoried. We show that this, too, is problematic.

1.1. *Integration: What Does It Mean?*

What does it mean for a series to be integrated or process an "unit root"? In order to understand this, we will need to understand the broader idea of a stationary distribution. Recall that we are interested in making inference about the data generating process of some random variable, in our case the monthly survey marginals on presidential approval. If we knew the distribution, $F(y)$, of our random variable we would know all we can about the process. Typically we assume that we know this distribution up to some set of unknown parameters, θ, and use maximum likelihood techniques to estimate to them (King 1989).

One of the properties that we might be interested in is whether or not $F(y)$ is stationary. Informally, we say a series is stationary if its distribution — data generating process — does not vary over time. Stationarity is of interest both to understand the underlying causal mechanism as well as to better forecast a series; failure to account for non-stationarity will lead to mis-specification and possibly mistaken inferences.

More formally we can define a series to be *strictly stationary* if the distribution of the series does not change if we look τ periods ahead or behind. Violation of strict stationarity can occur for many reasons. The simplest possibility would be some sort of structural break that fundamentally changes the data generation process. Given that we are working with survey marginals, a possible cause of such a structural break would be a change in the wording of a survey instrument. An example of how such structural break might occur — and the ways in which it would lead to problematic substantive inferences — occurred in the late 1960's and early 1970's when the National Election Studies changed the format of their issue preference questions, which set off a flurry of research first asserting that American politics had become more issue oriented during that period (Nie, Verba and Petrocik 1979; Pomper 1972). But subsequent research questioned whether American politics had really changed, or whether the changes in the National Election Studies questions produced an illusion of change (Bishop et al. 1978; Kessel 1972; Sullivan et al. 1978).

While strict stationarity is useful to think about, it is typically more then we can assume in practice. In general, we can neither test for strict stationarity nor can we estimate a fully non-stationary model, since by construction this means every observation comes from its own unique distribution. Instead studies usually focus on whether the mean and variance of the data generating process are stationary. We say that a series is *weakly stationary* or *covariance stationary* if:

$$\begin{aligned} \mathrm{E}[y_t] = \mu &\qquad \text{for all } t \\ \mathrm{E}[(y_t - \mu)(y_{t-j} - \mu)] = \gamma_j &\qquad \text{for all } t \text{ and any } j. \end{aligned}$$

In other words, a series is covariance stationary if its mean and variance do not depend on time and the covariance between any two observation is only a function of how far apart they are.

Even within the class of covariance non-stationary models, researchers have typically restricted themselves to either integrated or trend stationary series. An integrated series is one that can be made covariance stationary by taking first differences. The simplest example of such an integrated series is a *random walk with drift*:

$$y_t = \mu + y_{t-1} + \epsilon_t \tag{1}$$

where ϵ_t is a mean-zero random variable (often called a shock) with finite variance, σ^2_ϵ. This can be re-written by repeated substitution:

$$y_t = \sum_{s=0}^{t} (\mu + \epsilon_s). \tag{2}$$

It is easily shown that the variance of y_t is:

$$\begin{aligned} \mathrm{var}(y_t) &= \sum_{s=0}^{t} \sigma_\epsilon \\ &= t\sigma_\epsilon. \end{aligned}$$

Note that here we have assumed that the constant (or drift parameter) is fixed (i.e., non-stochastic) so that it does not contribute to the variance. From this we can see that the variance is a function of time and hence the distribution of y_t is non-stationary. But note more is true: the variance is *explosive*. As the series runs on for longer and longer, the variance gets larger and larger. In the limit, the series will have an infinite variance.

We can also show that the mean of the series is non-stationary if $\mu \neq 0$. Again we use the repeated substitution and then take expectations:

$$\begin{aligned} E(y_t) &= \sum_{s=0}^{t} \mu + E(\epsilon_s) \\ &= t\mu. \end{aligned}$$

Thus the mean is also explosive if the y_t's have non-zero means.

However, if we re-write Equation 1 by taking first differences, often denoted by Δy_t, we get:

$$y_t - y_{t-1} = \mu + \epsilon_t \tag{3}$$

This differenced series, which is also a random variable, is (covariance) stationary because its mean is constant, $\mathrm{E}[y_t - y_{t-1}] = \mu$, and its variance is both finite and constant, $\mathrm{var}(y_t - y_{t-1}) = \sigma_\epsilon$. Since the expected value of a stationary series is constant, it is often called a "mean reverting series". The shocks, $\epsilon_t' s$, just move the series around its mean. On average there should be as many negative shocks to offset the positive ones. That is, if we were to graph out such a series, we would just see random fluctuations around the mean value.

In contrast, such a graph of an integrated series would tend off to either plus or minus infinity never returning to any particular level. That is, a integrated series is infinitely memoried, a particular shock in period t affects the series forever bummed up or down depending on the sign of the shock. This can be seen by looking at the recursion that defines y_t given in Eq. 2. To restate this point, **a shock to an integrated time-series has permanent, and never-decreasing, effects.** This is a simple, and straightforward, implication of the fact that the time-series process of an integrated phenomenon is constantly and systematically changing over time.

Series which become stationary after taking their first differences are said to be integrated of order 1 and denoted as $I(1)$ series. It is obvious that a time-series might become stationary after taking the second or third differences, and those time-series would be called $I(2)$ or $I(3)$ series. It is possible that a series can be made stationary after fractional differencing, a topic we will return to below.

In fact, it is now widely accepted that many economic time-series are integrated. Most macroeconomic time-series which are tied to population growth are $I(1)$ time-series — examples are economic output and employment — and examination of these undifferenced time-series across the past few decades show that they are constantly increasing. This fact leads to very real substantive implications about the nature of these time-series and the effects of attempts by macroeconomic policymakers to affect changes in national output or employment.

Integration also has important consequences for the empirical models we try to fit to integrated series. First, assume that we have an $I(1)$ series like that given in equation 1. But, let's now assume that we fit a simple regression model to this time-series, of the sort which were once common in the empirical literature on presidential approval:

$$y_t = x_t\beta + \eta_t, \tag{4}$$

where x_t is some $I(0)$ (or stationary) variable, such as an event dummy for Eisenhower's heart attack. Fitting a regression model to equation 4 is clearly incorrect. First, given that y_t is $I(1)$, this implies that the disturbances violate the usual assumptions necessary to obtain the standard desirable OLS properties. Second, there is a serious inconsistency in this regression model, since we have two different types of variables — one which is $I(1)$, and which is consequently drifting over time in a positive or negative direction with an infinite variance and one which is stationary over time with constant variance. Given these two problems, it is not clear what inferences we can make about the coefficient we estimate in equation (4), other than the fact that these inferences will probably be incorrect!

This is not to say that $I(0)$ variables can not be used in models of $I(1)$ processes; they just can not be used to model the levels. The alternative is to model the differences of the $I(1)$ series, which by definition are stationary. So we can recast equation (4) as

$$\Delta y_t = x_t\beta + \varepsilon_t \tag{5}$$

where Δy_t is just $y_t - y_{t-1}$. We should note that the interpretation of β in 5 is different than in the model of the levels. The x_t now only effects the short-run movement of y_t around some trend line.

Another alternative is available if there is some other variable, say z_t, that is $I(1)$ and cointegrated with y_t. We will say that two, or more series, are cointegrated, if some linear combination of them is stationary. Formally, we say that two series are cointegrated if there exists some γ such that

$$\varepsilon_t = y_t - \gamma z_t \tag{6}$$

is stationary. Loosely speaking two variables are cointegrated if they are both $I(1)$ but they are in a long-term equilibrium (in levels) such that neither wanders very far from the other — in other words, they are trending along the same attractor.

Engle and Granger (1987) have shown that when two variables are cointegrated the model can be written in *error correction* form:

$$\Delta y_t = \Delta z_t \theta + \lambda(y_t - \gamma z_t) + x_t \beta + \varepsilon_t \tag{7}$$

where ε_t is a stationary and mean zero error term. Futher we can allow ε_t to be be independent and identically distributed or to be time dependent with some ARMA(p,q) process. What is nice about the error correction model is that it incorporates both short term dynamics, by the inclusion of Δz_t and x_t, as well as long term changes in levels by the cointegrated term $y_t - \gamma z_t$.

Ostrom and Simon (1992) fit an error-correction model to the presidential approval series and inflation during the Reagan administration. They argued that the presidential approval and inflation are co-integrated. Further, they claim approval series is integrated or "near-integrated" for the full span of their data (Truman through Reagan). This result has lead to a number of papers focusing on integration in the presidential approval series (Beck 1992; Ostrom and Smith 1992; Smith 1992; Williams 1992).

1.2. *Can Approval Be Integrated, Theoretically Speaking?*

Can the presidential approval series be integrated? Recall that if it were integrated, then we are assuming it will have increasing (decreasing) mean and an explosive variance. Of course in finite observed sample of an integrated series we will not literally see the variance and mean tend toward infinity. However, we should see clear trending. In addition, we should see that any shock to the presidential approval series should have a long-lasting effect. These are simply the consequences of the assumed properties of an integrated time-series

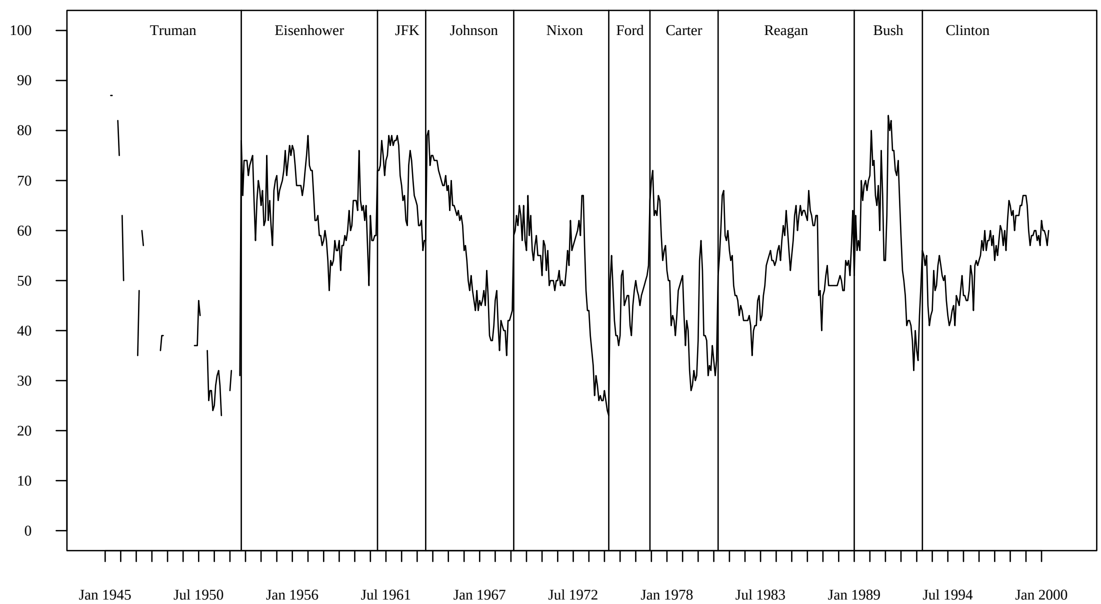

Figure 1: *Monthly Presidential Approval, 1944–2000*: This figure plots the percentage of the respondents that report that they approve of the job of the president in a given monthly survey from the Gallop Organization. In months with more then one survey, the average over the surveys is reported.

So as a first cut at this question we just plotted the monthly presidential approval series from 1944 to 1996; the full descriptive statistics for the time-series are in Table 1. As can be seen from Figure 1, there is not much of a trend even though we have about six hundred observations! The series plotted in Figure 1 seems to regress towards the mean of approximately 55, never getting close to either of the boundaries. Neither do we see any obvious increase in the variance over the course of our sample. In fact , the variance of this series seems to differ in different Administrations, as can be seen in Table 1, a point we will return to later. Third, there are signs of many shocks in the time-series of presidential approval graphed in Figure 1, but none of these shocks seem to have lasting and non-decaying impacts. For example, Nixon's popularity drops rapidly towards the end of his Administration

Table 1: *Summary Statistics of Monthly Gallop Poll Presidential Approval by Administration*

Administration.	N	Mean	SD	Lag	SD	ADF
Truman	22	43.6	17.1	.88	.07	-2.43
Eisenhower	96	65.2	7.3	.55	.07	-4.16
Kennedy	35	70.2	7.0	.83	.10	-2.45
Johnson	61	55.3	13.3	.95	.04	-2.86
Nixon	68	50.3	12.4	.96	.05	-1.89
Ford	28	46.4	4.8	.26	.14	-2.58
Carter	48	46.0	12.5	.91	.06	-2.32
Reagan	96	53.1	7.7	.85	.05	-2.18
Bush	48	60.3	14.2	.88	.07	-2.63
Clinton	39	48.3	4.5	.62	.13	-2.85
Pooled Series	563	54.7	13.3	.91	.02	-5.41

as a function of the Watergate scandal — but once Ford takes office, the effects of the Watergate shock dramatically reverse, and Ford's popularity rapidly increases. Thus at first take it does not seem at all appropriate to consider approval integrated, since the behavior of this series clearly does not show the classic signs of integration.

One's eyes can be deceived, hence the need for statistics. Our inter-ocular test need not be conclusive, however we have a deeper and more fundamental reason to be deeply suspicious of the assertion that approval is integrated: simply put, presidential approval is a bounded series, so it is not possible for the mean or variance to explode.

In order to see this somewhat more formally we need to consider exactly what happens at the boundary. There are two logical possibilities that we must consider: absorbing or reflective boundaries. With absorbing boundaris when our process has an eventual realization that hits the boundary, in our case 0 or 100, then the process will stay there forever. Such an assumption would make sense if we believed that there was some underlying latent random variable, say y_t^*, that was integrated, and hence tending off toward $\pm\infty$, but we only observed some monotonic transformation of it, such as the logistic, that kept it in bounds. Then, it is very simple to see that the distribution of the observed y_t would be stationary. For all $t > T$, y_t would be either 0 or 100. So our distribution would be degenerate, but stationary (i.e., time independent) with mean either 0 or 100 and zero variance!

The case of reflecting barriers is somewhat more difficult. Under reflective boundaries if the process ever has a realization that hits the boundary, the following realization is bounced back into the interior of the space. It can be shown that this leads stable (i.e., stationary) distribution by noting that this random walk can be expressed as a Markov chain with reflective boundaries. The case where the "states", realization of y_t, are finite then the transition properties that define the process are stationary (see Feller 1968:436–8). This can be extended, with some care, to the case of infinite (i.e., continuous) states.

Irregardless of whether we believe that the presidential approval time-series has absorbing or reflecting boundaries, it is important to note that in neither case can this time-series be thought of as integrated. This simple fact — that neither type of bounded time-series can exhibit the fundamental properties of an integrated time-series — has been largely ignored in past research.

A more recent alternative suggested by Box-Steffensmeer and Smith (1998) that seeks to circumvent the problems of integrated process for models of survey marginal is fractional integration. Recall that we made the I(1) series described in Equation 1 stationary by taking first differences. In principle there is no reason why some series might be better fit by only fractional differencing the data. We would call such a series fractionally integrated. A fractional integrated series is also long memoried, but unlike an integrated series, its memory is finite. If we were to observe the series for long enough it would eventually return to its mean. This implies that both the mean and variance would be finite and hence not be logically impossible for a bounded series, such as presidential approval. We note, however, that over long stretches such fractionally integrated series could deviate far from their means because of this long memory. That is, shocks could add up easily pushing the series out of the bounds we know must hold. Thu,s before we would suggest using such a fractionally integrated models for survey marginals, we would want to modify them to insure that they did not produce values out bounds.[1]

[1]Another nice property of the fractionally integrated series is that they can be generated by aggregating certain types of heterogeneity at the individual level (Granger 1980), hence they might be thought to solve the aggregation problem we will discuss in the last section. The conditions of Granger's (1980) aggregation theorem, however, require that the individual series being summed be AR(1) processes. It does not seem reasonable that this assumption holds for an individual response to an approval question. A more likely model is that every respondent has some propensity to support the president when asked – i.e., a random

2. POOLING PRESIDENTIAL APPROVAL ACROSS ADMINISTRATIONS

In the previous section we considered the problem of modeling aggregate presidential approval as an integrated time series. In this section we take up the first of two aggregation issues, whether it is reasonable to pool data across presidential administrations. We argue not since pooling leads to theoretical problems because of "spill-over" across administrations. Further, even without this theoretical concern, the data empirically reject homogeneity of approval across administration. We will take these up in order.

2.1. *Spill-Over*

To begin, it is best to be a bit more persist in our model of the presidential approval series. A standard and simple model to consider a simple AR(1) process:

$$y_t = \rho y_{t-1} + \mathbf{x}_t\beta + \varepsilon_t \qquad \forall t = 1, \ldots, T \tag{8}$$

where y_t is the presidential approval in month t; $\mathbf{x}_t$ is a $1 \times k$ vector of exogenous regressors, such as measures of unemployment and inflation; ε_t is some well behaved error term with zero mean; and ρ gives the correlation between y_t and y_{t-1}. We note that much of the literature has focused on the determining appropriate dynamic structure (see Beck 1991 for details. Restricting ourselves to the case of AR(1) is not crucial for our insights as we could easily allow a more general error structure, e.g. allow the errors to follow some ARMA(p,q) process, but this would not change our general insights and just complicate estimation and interpretation. We will also assume that $|\rho| < 1$, so the series is stationary (see section 1 for a detailed argument).

The question at hand is what to do about the fact that periodically a new administration comes into office? In our dataset, the presidential approval series starts in January 1944 and run through June 2000, covering ten separate Presidents. The root of problem is that ρ is really just the partial (conditional) correlation between $\{y_t, y_{t-1}\}$. We then have to ask why should Bush's approval in his last month in office (directly) effect Clinton's approval in his first month? This is implied, however, by the model as specified in Equation 8.

effect. Under such model aggregation will not lead to fractional integration.

Beck (1995) has suggested a simple way to get around this problem, at least with regard to estimation. He suggests excluding any President's first month in office (via a dummy variable). This will prevent any contamination in the estimation of the partial autocorrelation, ρ, at the break points. However, this does not really solve the problem. Equation (8) defines an infinite distributed lag model. To see this (ignoring regressors), suppose the series starts at $t = 0$ with some initial value, y_0, then:

$$
\begin{aligned}
y_1 &= \rho y_0 + \varepsilon_1 \\
y_2 &= \rho y_1 + \varepsilon_2 = \rho^2 y_0 + \rho\varepsilon_1 + \varepsilon_2 \\
&\vdots \\
y_t &= \rho^t y_0 + \rho^{t-1}\varepsilon_1 + \rho^{t-2}\varepsilon_2 + \cdots + \rho\varepsilon_{t-1} + \varepsilon_t.
\end{aligned}
$$

Note that this last equation holds generally for any starting date $t-1$, so

$$y_{t+j} = \rho^j y_{t-1} + \rho^{j-1}\varepsilon_t + \rho^{j-2}\varepsilon_{t+1} + \cdots + \rho\varepsilon_{t+j-1} + \varepsilon_{t+j} \tag{9}$$

Now consider the effect of some some shock in month t on some later period $t + j$, which is referred to as the dynamic multiplier and is defined by:

$$\frac{\partial y_{t+j}}{\partial \varepsilon_t} = \rho^{j-1}$$

This is easily seen by differentiating equation (9). So even though the first month of the President's term are excluded from the estimation of ρ, equation (8) still implies that "shocks" to previous President's approval can affect another President's approval.

So what does this mean in practice? Figure 2 plots out the dynamic multipliers for approval series assuming a ρ of 0.91, which is Beck's (1995) estimate. This figure shows the duration of the effect of a one percentage point shock in the presidential approval series for the next 48 months. As you can see from the graph, there are sizable effects even two years after the shock. Consider, for example, the hypothetical case of a 5 percent point shock to Bush's approval in February, 1991, caused by the Gulf War. The model as written implies that this shock causes Clinton's approval in February, 1993 (24 months latter) to rise by about half a percentage point! Clearly this can not be what previous scholars have meant when they have assumed models like Equation (8). We also note that this spill-over problem also exists for

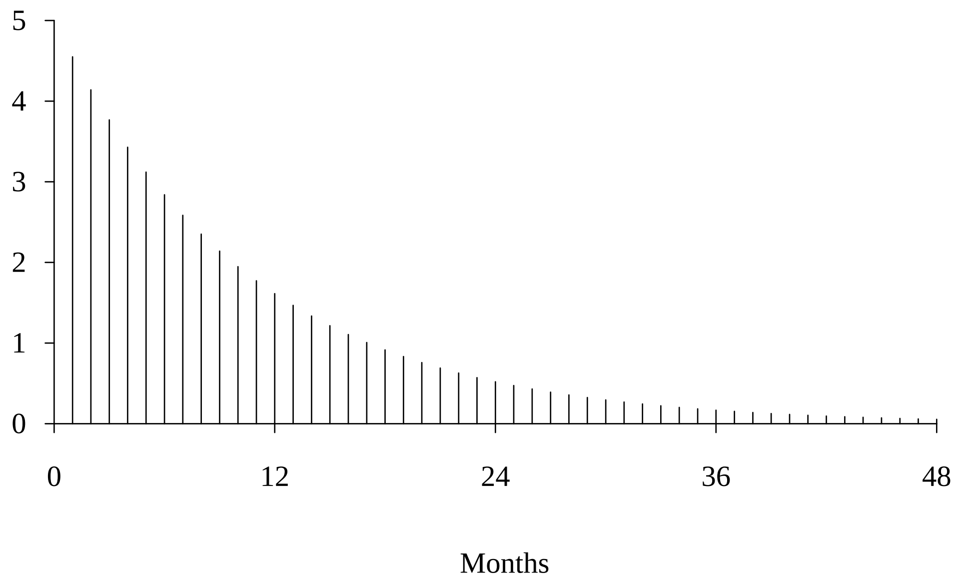

Figure 2: *Presidential Approval Dynamic Multiplier*: The figure plots out the impact of a 1 percentage point shock to an AR(1) series assuming a ρ of 0.91.

fractionally integrated models since by construction they are long memoried. Unfortunately constructing their dynamic multipliers is more difficult. However, it can be shown that any fractionally integrated series can be arbitrarily well approximated by a suitable large-order ARMA representation (Hamilton 1994:449). Using such a representation, the spill over argument presented above will hold.

2.2. *Parameter Homogeneity*

Even if we thought spill-over was not a problem, we can ask whether the data are consistent with pooling across administration. Such a investigation is not merely an estimation nuisance, but can provide valuable insight into the underlying process that generates presidential approval. Given our pooled structure, it is straightforward to ask if economic conditions effect the approval of Democratic and Republican administrations in the same fashion, or whether other factors which might influence approval positively in one Administration might work against a subsequent president's approval.

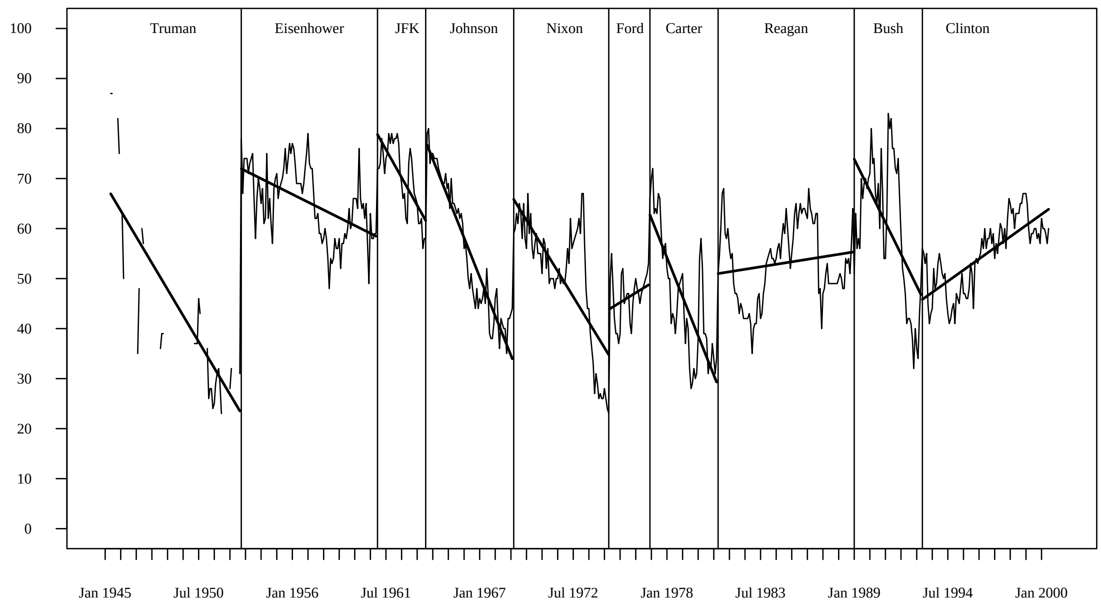

Figure 3: *Monthly Presidential Approval With Time Trend by Administration: This figure is the same as Figure 1 accept time trends are included. The trends are estiamted by regressing monthly approval on a constant and a time indicator.*

In the rest of this section we will only really consider parameter variability with regard to how to specify the dynamics in the model. That is we want to ask does the same AR(1) dynamic fit all the administrations? Our first examination of this assumption is to re-examine the plot of monthly presidential approval originally seen in Figure 1. Figure 3 is the same plot except we have now included lines that represent simple time trends that vary by president. As can be seen from the figure, there seems to be enormous variability in the overall time trends. Most striking is the fact that Reagan and Ford's approval both increase over the course of their time in office whereas all other presidents approval decline. We also note that the downward trend seems noticeably less step for Eisenhower and Clinton.

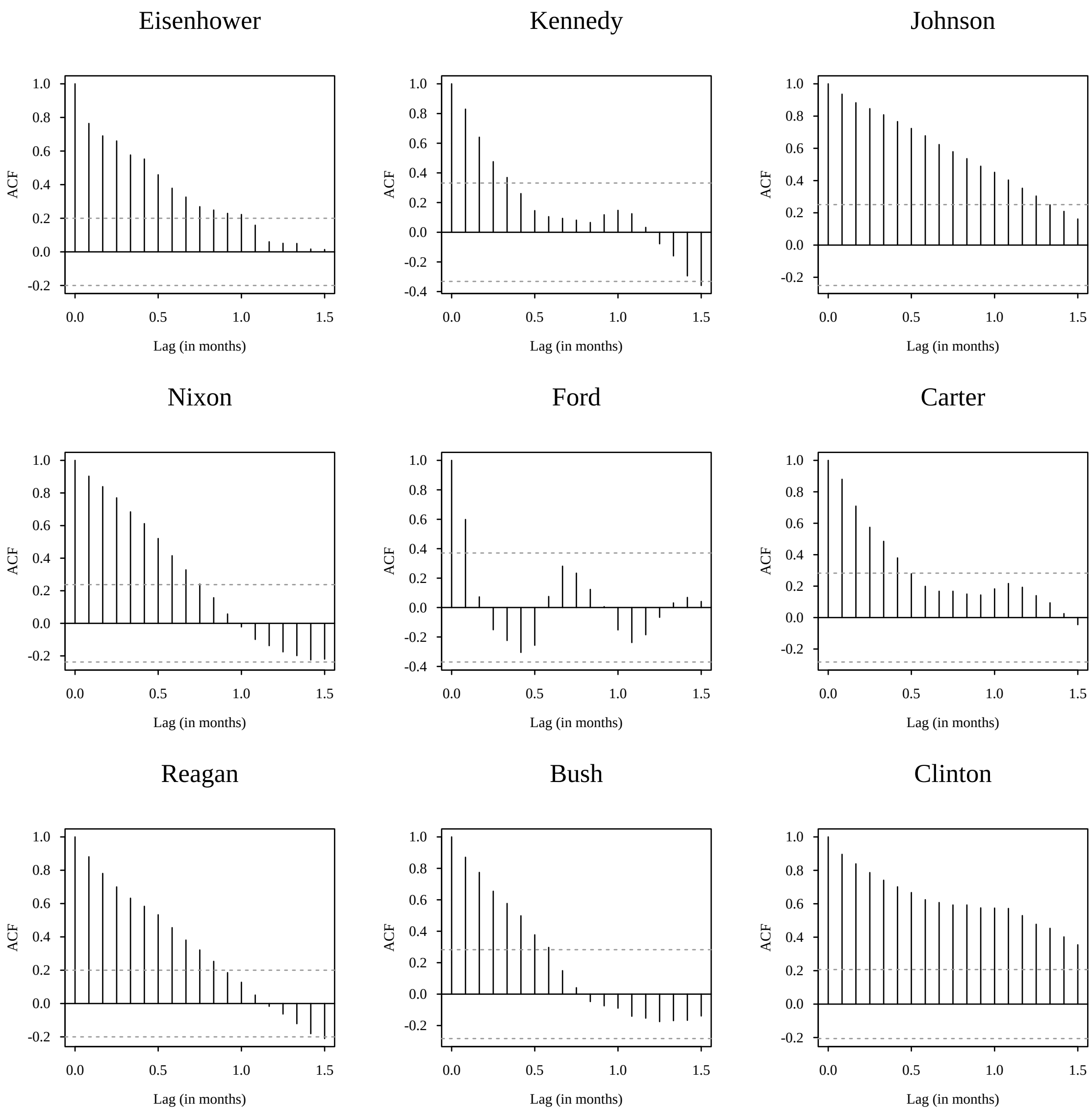

Figure 4: *Autocorrelograms for Presidential Approval By Administration*: The figure plots out the autocorrelation of the Gallop Presidential Approval series through lag 15 for each Administration. A graph for the Truman administration is not presented because of missing data. Values out side of the dashed lines represent a significant autocorrelation at the lag.

The next step in determining the homogeneity of the dynamics is to examine the correlograms (see Box and Jenkins 1964 and Granger and Newbold 1977). If the correlograms show a similar pattern, then we are in a better position to accept the assumption of common dynamics. In Figure 4 we plot the correlagrams with standard error bars[2], it is readily ap-

[2]Values outside the two error bars are statistically significant at conventional levels1

parent that there seems to quite a bit of variation in dynamics across Administrations. The oddest correlagrams is for Ford, although this seems somewhat explainable. Ford has only is very short time series, since he was only in office for two years. It is particularly conforting to note that all of Ford's fluctuations are within the standard error bars so they are not statistically significant. The other thing to note from the correlagrams is the high amount of persistence in the series for Eisenhower, Reagan, and Johnson. We could also examine the partial autocorrelograms by administration following standard time-series approaches as advocated by Box and Jenkins (1964). We would see the same pattern: wide variability in the dynamics by administration.

Instead of continuing with these fairly informal methods, we will directly test for for parameter homogeneity. A formal test can be constructed by estimating the model outlined in Equation 8. We have done so, regressing monthly presidential approval from the Gallop Organization on its lag as well as on changes in the consumer price index (CPI) and unemployment, both lagged one month.[3] In order to avoid problems of missing data, we will only used data from January 1957, the start of the Eisenhower administration, through June 2000 of the Clinton administration. Our results from this regression agree with the rest of the literature .Declines in both CPI and unemployment lead to statistically significant gains in month approval. Further, the estimated partial correlation is 0.886 with a standard error of 0.018.

In order to test for homogeneity, we will allow all of the parameters in the regression to vary by administration.[4] There is some question as what to do about the constant term. We have excluded other regressors that the some in the literature has used, such as indicators for scandal or the first month of a new administration, which we find theoretically unjustified, one might, therefore, expect to find fixed effect which are of little substantive concern. Although strictly speaking, for homogeneity to hold all parameters need to be constant. Given we are agnostic with regard to this issue, we tested allowing for fixed effects and the more stringent criteria of complete homogeneity. Regardless, both tests lead to the same conclusion, parameter homogeneity is rejected. The F-statistic for the restriction allowing for fixed

[3]The economic data come from the Bureau of Labor Statistics.

[4]This is easily done by interacting indicator variable for each administration with all regressors and the constant.

effects is 3.82. Given the degrees of freedom for the statistic of 24 and 534, we reject the null hypothesis of pooling at all conventional levels. If we further require the constant to pool, the F-test becomes 3.57 on 32 and 534 degrees of freedom, also leading us to reject the null hypothesis of parameter homogeneity at all conventional levels.

A greater feel for the variability can be found in Figure 5. This plots out estimates of ρ by administration with 95 percent confidence intervals. As you can see there is wide variability between administration, The low is from Ford with an estimate of ρ at 0.24 and a high of 0.95 for Johnson. Even if we were to throw out Ford, since it was an odd short administration, the range on the estimates of ρ run from 0.81 to 0.95, quite a bit of variability in the underlying dynamics.

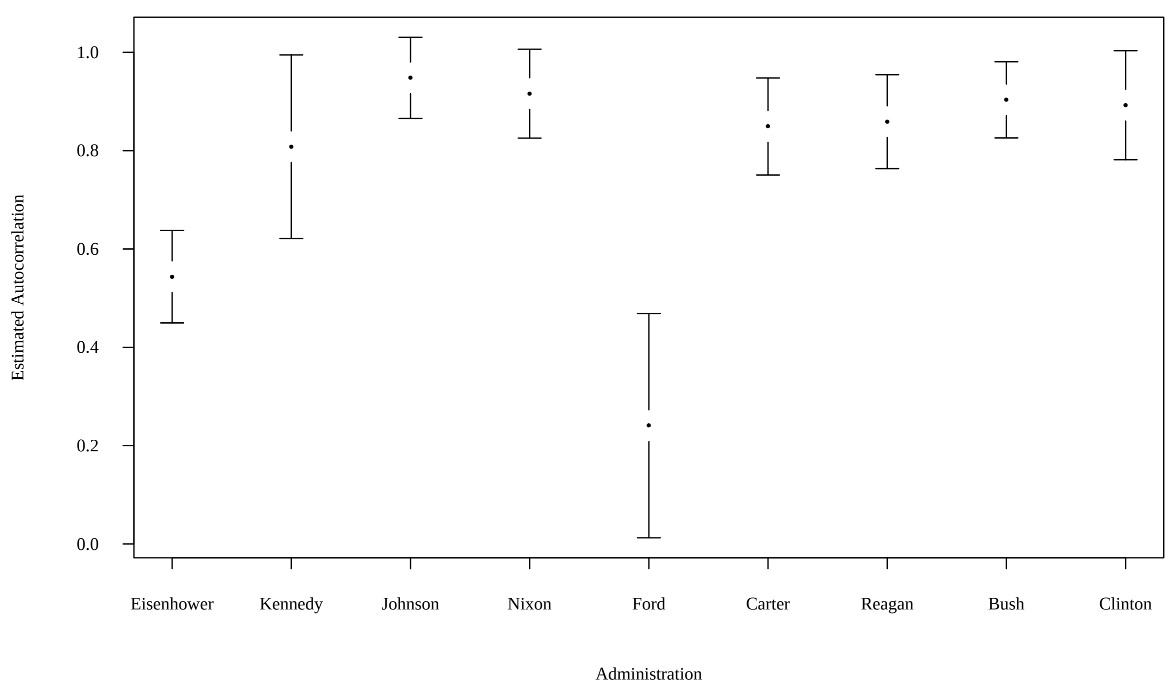

Figure 5: *Estimates of ρ with 95% Confidence Intervals by Administration*: This figure plots out the estimates of ρ, the parameter on the lagged approval in a regression that also includes a constant term and measures of changes in unemployment and inflation. All parameters are allowed to vary by administration.

The theoretical problem of spill-over along with the failure empirically to accept homogeneity casts doubts on the use of a single presidential approval series as has been standard practice in the literature. Even if these problems could be resolved, there is still one fundamental problem left to address. Recall that the approval data used in these analysis are aggregated survey marginals; thus there is some process which aggregates individual-level survey responses to aggregate survey marginals, and the model which explains variance at the aggregate survey marginal level may not be the model which determines variance at the individual-level. This is a problem of *aggregation bias*, much like that discussed in the ecological inference literature (e.g., King 1997) and in the seminal work on economic voting by Kramer (1983).

In fact, Kramer's (1983) problem, the impact of economic outcomes on presidential candidate choice, is quite similar to the presidential approval problem. Kramer begins with an individual-level behavioral model, where economic outcomes are partitioned into those that are government- or randomly-induced, and both types of economic outcomes directly determine which presidential candidate a voter chooses. Kramer then shows that a simple aggregate time-series estimate of the relationship between economic outcomes and presidential election outcomes will generally be biased: "we should expect the aggregate time-series estimate to be of the correct sign, although probably somewhat attenuated in magnitude" (Kramer 1983: 100). Kramer also points out that the aggregate time-series estimate cannot partition sociotropic behavior from self-interested pocketbook voting (Kramer 1983: 106). Thus, Kramer's work supports our basic argument throughout this paper; generally speaking, the use of aggregated time-series data for studying presidential approval has both methodological and substantive problems.

3. TURNING TO AN INDIVIDUAL-LEVEL APPROACH

Is there anything that can be gained by moving from the aggregated time-series analysis of presidential approval to an analysis of the individual-level presidential approval data? On this point, Kramer (1983) was quite pessimistic, as he asserted that individual-level cross-sectional studies of economic voting were at least as methodologically flawed as aggregate time-series studies. Yet Kramer's assertions here rest on very strong assumptions; in particular, Kramer consistently assumes that in any particular election year, the government-induced effect on voter's economic well-being across the nation is homogenous across the electorate, or that the government-induced effect is smaller than idiosyncratic effects (Kramer 1983: 100-101). In these two cases, individual-level cross-sectional analysis may indeed lead to biased estimates of economic voting or of the effect of economic perceptions on presidential approval. On this point, we do not share Kramer's pessimism, as his assumptions about the homogeneity of government-induced economic well-being effects seems quite unrealistic to us.

There are many merits to using individual-level data for studying presidential approval, in addition to avoiding the types of aggregation biases and problems we discussed in the previous sections of this paper. One obvious gain is efficiency, since we will have much more information at our disposal to explain presidential approval; however, the ultimate effect of this efficiency gain has been disputed (Beck 1991). A second obvious merit is that there are important theoretical reasons for using the individual-level presidential approval data, though, since ultimately our hypotheses which drive the specification of aggregate presidential approval models are hypotheses about individual-level behavior. Thus to test these hypotheses correctly, we should use individual-level data, not attempt to infer individual behavior from aggregate models.

However, there is on looming problem with the use of individual-level survey data for studying presidential approval — the availability of the individual-level survey responses. In our opinion, the focus on the aggregate presidential approval series in the literature has arisen largely because of the ease of aggregate data collection and analysis. The Gallup Organization and the other major commercial polling firms make their survey marginals available to the public, and these are easy and cheap to collect and analyze. Unfortunately the major commercial polling first are not as forthcoming with their individual-level data, which has inhibited their use to test hypotheses about individual behavior with the appropriate data.[5]

[5]Of course, the National Election Study has included a presidential approval question in most of their biannual surveys; also, there are a wide array of media polls from recent decades available from the ICPSR. Unfortunately, the latter polls are not necessarily systematically archived in a public access facility like the ICPSR, nor do they systematically include the same sorts of questions about presidential approval in their survey instruments, making these polls difficult to use for the systematic study of presidential approval.

We have access to the Gallup Organization's polling data for the 1988 election; we focus on the Gallup data here since it is exactly these surveys which are commonly used to construct the aggregated measures of presidential approval which we examined in the previous sections of this paper. During the span of January through October 1988, the Gallup Organization asked their standard presidential approval question, "Do you approve or disapprove of the way Ronald Reagan is handling his job as president?" in ten separate national telephone surveys. We graph each of the possible survey response marginals in Figure 6.

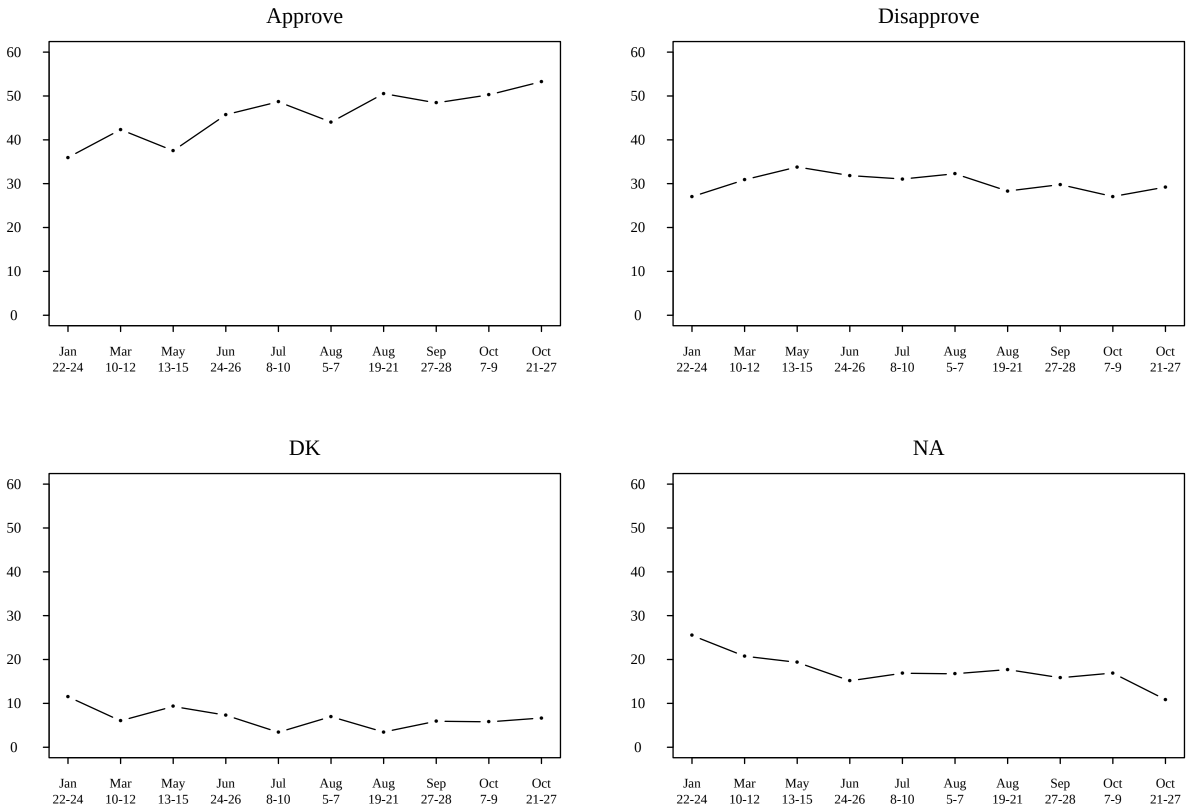

Figure 6: *Presidential Approval Responses from Gallop Monthly Surveys in 1988*: The figure plots out the marginals for the four possible responses to the approval question in the 10 Gallop Organization surveys for 1988 for which we have individual level data.

During the 1988 election the Gallup presidential approval survey marginals point to three interesting dynamics. First notice that the rate of item non-response (in the graph as "NA", or "not available") drops considerably during the presidential election campaign. This is interesting since the item non-respondents are people who were not asked the presidential approval question in the particular survey — the Gallup Organization in 1988 did not ask the presidential approval question of people who said they were not registered to vote, or were not sure about whether they were registered to vote.[6] Over the course of the election, it is clear from Figure 6 that many more respondents become registered, or could recall that they are registered, as the campaign season progressed. Thus, increasing numbers of survey respondents were posed the presidential approval question as the election year progressed.

Secondly, during the 1988 election, there is a steady increase in the percentages of registered voters who approve of Reagan's performance. This rise in Reagan approval during the election campaign could be the result of changes in the economy during 1988, but could also be due to campaign information — since one of the central components of the Bush election campaign was the continuation of Reagan's policies.

But third, notice that there is very little change in the percentages of registered voters who disapproved of Reagan's performance, or who did not care to express an opinion (the "dk" in the Figure). That the percentages of registered voters who disapproved of Reagan's performance during the 1988 election hardly changed is fascinating, since it implies that whatever factors were operating to drive Reagan's approval upward were not operating to symmetrically drive Reagan's disapproval down. If the economy were improving steadily during 1988, we would expect to see Reagan's approval go up and his disapproval go down, which we do not see in Figure 6.

What does show a downward trend during the 1988 election, though, is the percentage of non-registered voters. The rise in Reagan approval seems to be driven by survey respondents who either became registered voters, or who recalled being registered voters. Thus, the best explanation for the rise in Reagan approval in the 1988 election is campaign activation.

In general, just by examining the entire distribution of responses to the Gallup Organization presidential approval question a number of interesting patterns can easily be seen. This means that by concentrating only on the "approve of" responses, the literature on presidential approval at best may have missed important conclusions which could be found only by examination of the entire set of responses to this one survey question. At worst, this

[6] At this time we are not sure whether this is a common practice of the Gallup Organization in all of their surveys, or whether this is a particular format they use in only election years. If the latter is the case, this implies that election year and non-election year respondents to this question could be quite different.

literature may have reached some erroneous conclusions.

Obviously there is more information in these surveys than the simple marginals. To demonstrate what potentially is lost by aggregating the survey responses, we estimated binary probit models where the dependent variable is coded 1 for presidential approval, 0 for disapproval; "don't know" respondents were deleted from the analysis. Given the possibility of heterogeneity, we estimated heteroskedastic binary probit models (Alvarez and Brehm 1995, 1997; Brehm and Gronke 1995). The heteroskedastic probit model can be thought of as comprising two separate components: a model of the response (the "systemic component") and a model of the response variance (the "variance component").[7]

In the systemic component, we include variables which should account for how registered voters evaluate incumbent presidents — their assessments of the current state of the national economy, their assessments of their personal financial situation, and their partisan affiliations. The first two we measure with simple scales (better, the same, or worse), while we add two dummy variables for Democratic and Republican affiliation for the last. We expect to obtain positive coefficients on all but the Democratic affiliation variable. For the specification of the variance component, we included variables which measured how politically informed and sophisticated the voter was with education, minority status, gender, and age. We expect to see the better informed and sophisticated voters (higher education, non-minorities, males, and older voters) to have less response variance.

We give model estimates (coefficient estimates and estimated probability "first differences") from two of the 1988 Gallup polls in Table 2.[8] In the left columns we present estimates from the January 22-24, 1988 Gallup national poll, while in the right columns we give similar estimates from the August 5-7 national poll. We use these two polls since they are the national polls closest to the beginning and end of the campaign which share these variables in common; the estimates, then, may shed some light on the dynamics of presidential approval we observe in Figure 6.

In the January 22-24 Gallup poll, we see that the systemic component estimates are all

[7] Details of the heteroskedastic binary probit can be found in Alvarez and Brehm (1995, 1997).

[8] The probability "first differences" are calculated by setting a hypothetical respondent (all variables set to sample mean values); we then estimate the probability of an "approve of" response twice, first by setting the independent variable at the minimal value, second by setting the same variable at the maximal value. We report in Table 2 the difference between these two probabilities.

Table 2: Probit Estimates on the Probabilty of Approving of the President

Independent Variables	Jan 22-24 Estimates	Jan 22-24 Effects	Aug 5-7 Estimates	Aug 5-7 Effects
Systemic Component				
Constant	-2.57**		-3.52**	
	.75		.83	
Pers. Finances	.79**	.53	.63**	.40
	.22		.19	
Natl. Economy	.61**	.42	1.21**	.68
	.20		.29	
Democrats	-1.29**	-.30	-.78**	-.18
	.37		.31	
Republicans	1.52**	.26	.69**	.14
	.46		.29	
Variance Component				
Education	.30**	.26	.19*	.19
	.16		.13	
Minorities	-.22	-.20	.15	.15
	.28		.22	
Women	.43**	.55	-.12	-.14
	.15		.12	
Age	-.59*	-.44	-.21	-.18
	.42		.38	
Sample size	914		810	
Het. Test	13.6†		4.59	

Note: ** denotes p=.05 significance, and * p=.10 significance, both one-tailed tests. † denotes a χ^2 test significant at the p=.05 level with 4 degrees of freedom.

statistically significant and in the expected direction. That is, voters who thought their personal finances or the national economy had gotten better were more likely to approve of Reagan's performance. Also, Republican identifiers were more likely to approve of Reagan's performance, with Democrats less likely to like the job Reagan had been doing. The probability estimates (second column) give us the ability to make better comparative assessments about the relative effects of each variable in the systemic component. The critical conclusions are that each of the economic factors have a larger effect on the likelihood of approval than partisan affiliation; second, the effects of a voter's pocketbook perceptions has a stronger effect on the probability of approval than their perceptions of the national economy.

The variance function produces similarly interesting results. First, however, notice that we find evidence of significant heterogeneity in the January 22-24 survey data. Our test of heterogeneity is statistically significant, and three of the four terms in the variance function have statistically significant effects.[9] We find that, as expected, women and younger voters have higher response variance than males and older voters. This is consistent with the hypotheses that these voters are less informed, and as such, their responses have more variance due to their fundamental uncertainty (Alvarez 1997). However, we also find that higher educated voters have higher response variance than lower educated voters, which is inconsistent with a political information and uncertainty explanation. It is possible, thought, that higher educated voters at the beginning of this presidential campaign are more ambivalent, or more internally torn, about whether they approve of Reagan's job or not (Alvarez and Brehm 1995, 1997). In any case, we see from our estimates of the effect of each variable on the magnitude of the error variance (second column) that gender and age have much greater effects on response variance than education.[10]

In the third and fourth columns we give the results from the August 5-7 Gallup survey. The results in the systemic component look roughly similar to those from January, since again, the coefficients are all statistically significant and in the anticipated direction. However, when we turn to the probability "first differences", we see some interesting changes. First, the effects of partisan affiliation have diminished dramatically from the January results. In fact, the effects of both partisan labels have been cut almost in half. Second, the effects of both economic assessments are again stronger than partisanship. But third, in August, the effects of the national economy now swamp the effects of the voter's pocketbook perceptions.

Next, the variance function estimates in August show that there is no heterogeneity in these data by this point in the general election campaign. While the estimate of the effect of education is still positive and is greater than the associated standard error, the heteroskedasticity test indicates no heterogeneity.

[9]The heteroskedascity test is discussed in Alvarez and Brehm 1995, 1997. It is a simple χ^2 test of the restriction that the coefficients in the variance function are zero. This is done by estimating the model again, without the variance function, and by comparing the log-likelihood functions for the two models.

[10]These calculations were done much like the probability first differences; here we just estimate the different between the estimated magnitude of the error variance for high and low values of each variable in the variance function.

Thus, three important changes seem to have occurred from January to August 1988. First, the effect of partisanship on presidential approval diminished strongly. Second, the effects of perceptions of the national economy strengthen, and the importance of personal financial concerns weaken. Third, while there was significant heteroskedasticity in January, there is none in August.

It is hard to attribute these changes to anything other than the 1988 presidential campaign. Early in the presidential election, when there is little political information available, presidential approval was determined by a voter's personal financial condition, then by their assessment of the national economy, followed by their partisanship. Also, there is solid evidence of systematic error variance in January, which seems due to uncertainty and perhaps ambivalence. But by August, after eight months of primary and general election campaigning, presidential approval is now strongly determined by assessments of the national economy, followed by personal financial conditions. Partisanship now matters much less. And last, there is little evidence of systematic error variance at the end of the general election campaign.

These are all results which cannot be found by only concentrating on the aggregate-level presidential approval series. Obviously our theories about presidential approval, and what moves presidential approval, are theories about the opinions of individual citizens. To test these theories, we should not be in the business of examining aggregate-level data and making inferences about the opinions of these individual citizens. Indeed, for a long time social scientists have known that making individual-level inferences from aggregated data is highly problematic, since such "ecological" inferences often are incorrect (Achen and Shiveley 1995; King 1997). Instead, we ought to be examining the actual opinions of individuals; when we do, we see that there seems to be much more going on at the individual level than seems apparent in the aggregate presidential approval literature.

4. DISCUSSION OF SUBSTANTIVE ISSUES

In this paper we have argued that it makes little sense to think of presidential approval as an integrated time-series. The concept of integration originated in economics to account for the particular properties of many macroeconomic time-series; we have argued that the presidential approval time-series is a bounded series, and therefore cannot have infinite variance.

Instead, we have advanced two different ways to model presidential approval. There are many questions about presidential approval which involve macro-political or macro-economic phenomenon. For these questions, we believe that a panel approach to presidential approval is appropriate. But many questions about presidential approval focus on individual voters or citizens, and to answer these questions we should not rely on aggregate data. Instead we need to turn to models of the individual data to best test those theories. Indeed, when we turned to examinations of the presidential approval series either as a panel, or using the individual-level survey responses, we produced some new and interesting findings.

In the end, however, we advocate the use of individual-level data for testing hypotheses about presidential approval. Even when aggregate data is used in a methodologically appropriate manner — modeling it as a time-series cross-section and dealing with the fact that the dependent variable is a proportion — there is still the nagging aggregation bias problem which is impossible to diagnose using the aggregate approval data. It is the problematic question of aggregation bias which we argue should keep researchers wary of using the survey marginals in their research on approval.

But, presidential approval is an important concept in political science. It is integral in discussions of presidential politics, of presidential-congressional interactions, and of electoral politics. Given the conceptual importance of presidential approval in political science, we need to develop appropriate methodologies to test our theories, not continue to borrow techniques from economists.

Additionally, the cautionary tale we tell in this paper about how to model presidential approval also applies to other areas of political research. Recent years have seen a rise in the number of studies of other aggregated public opinion measures, ranging from different aspects of public opinion studied at the aggregate level (Page and Shapiro 1992) to a number of time-series studies of macro political opinions (e.g., Stimson 1998). While there is no doubt that studying public opinion at the aggregate level is interesting and important, it is clear that the methodological tools which are employed to study these phenomenon need to carefully match the data at hand, not be drawn indiscriminately from other social sciences.

REFERENCES

Achen, C. H. and W. P. Shively. 1995. *Cross-Level Inference.* Chicago, IL: University of Chicago Press.

Alvarez, R. M. 199. *Information and Elections.* Ann Arbor: Michigan University Press.

Alvarez, R. M., and J. Brehm. 1997. "Are Americans Ambivalent Towards Racial Policies?" *American Journal of Political Science*, 40: 345-374.

Alvarez, R. M., and J. Brehm. 1995. "American Ambivalence Towards Abortion Policy: Development of a Heteroskedastic Probit Model of Competing Values." *American Journal of Political Science*, 39: 1055-1082.

Beck, N. 1991. "Comparing Dynamic Specifictions: The Case of Presidential Approval." *Political Analysis* 3: 51-88.

Beck, N. and J. Katz, 1996. "Nuisance vs. Substance: Specifying and Estimating Time-Series–Cross-Section Models". *Political Analysis*, forthcoming

Brody, R. 1991. *Assessing the President.* Stanford: Stanford University Press.

Bishop, G. F., A. J. Tuchfarber, and R. W. Oldendick. "Change in the Structure of American Political Attitudes: The Nagging Question of Question Wording." *American Journal of Political Science* 22: 250-69.

Box, G. and G. Jenkins. *Time Series Analysis: Forecasting and Control*, 2nd. Edition. San Francisco: Holden-Day.

Engle, R. and C. Granger, "Co-integration and Error Correction: Representation, Estimation and Testing." *Econometrica* 35: 251-276.

Feller, W. 1968. *An Introduction to Probability Theory and Its Applications*, 3rd. Edition, Volume 1. New York: J. Wiley and Sons.

Granger, C. and P. Newbold. 1977. *Forecasting Economic Time Series.* New York: Academic Press.

Kernell, S. 1978. "Explaining Presidential Popularity." *American Political Science Review* 72: 506-22.

Kernell, S. 1986. *Going Public.* Washington, D.C.: CQ Press.

Kessel, J. H. "Comment: The Issues in Issue Voting." *American Political Science Review* 66: 459-65.

Kiewiet, D. R. and D. Rivers. 1985. "The Economic Basis of Reagan's Approval." In *Contemporary Political Economy*, edited by D. A. Hibbs and J. Fasbender, pp. 49-71. Elseview: North-Holland.

King, G. 1997. *A Solution to the Ecological Inference Problem.* Princeton, NJ: Princeton University Press.

Kramer, G. H. 1983. "The Ecological Fallacy Revisited: Aggregate- versus Individual-level Findings on Economics and Elections, and Sociotropic Voting." *American Political Science Review* 83: 92-111.

Levin, A. and C. Lin. 1993. "Unit Root Tests in Panel Data: New Results." University of California, San Diego, Discussion Paper 93-56.

MacKuen, M. 1983. "Political Drama, Economic Conditions, and the Dynamics of Presidential Popularity." *American Journal of Political Science* 27: 165-192.

Neustadt, R. 1990. *Presidential Power.* New York: Free Press.

Nickell, S. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica* 42: 579-590.

Nie, N. H., S. Verba, and J. R. Petrocik. 1979. *The Changing American Voter.* Cambridge: Harvard University Press.

Ostrom, C. W., Jr., and D. M. Simon. 1989. "The Man in The Teflon Suit." *Public Opinion Quarterly* 53: 353-387.

Ostrom, C. W., Jr., and R. M. Smith. 1992. "Error Correction, Attitude Persistence, and Executive Rewards and Punishments: A Behavioral Theory of Presidential Approval." *Political Analysis* 4: 127-184.

Page, B. I. and R. Y. Shapiro. 1992. *The Rational Public.* Chicago, IL: University of Chicago Press.

Pomper, G. M. "From Confusion to Clarity: Issues and American Voters, 1956-1968." *American Political Science Review* 70: 779-805.

Rivers, D. and N. L. Rose. 1985. "Passing the President's Program: Public Opinion and Presidential Influence in Congress." *American Journal of Political Science* 29: 183-196.

Smith, R. 1992. "Error Correction, Attractors, and Cointegration: Substantive and Methodological Issues." *Political Analysis* 4: 249-254.

Stimson, J. A. 1998. *Public Opinion in America.* Boulder, CO: Westview Press.

Sullivan, J. L., J. E. Piereson, and G. E. Marcus. "Ideological Constraint in the Mass Public: A Methodoligical Critique and Some New Findings." *American Journal of Political Science* 22: 233-49.

Williams, J. T. 1992. "What Goes Around Comes Around: Unit Root Tests and Cointegration." *Political Analysis* 4: 229-236.