Can Time Series Methods Be Used to Detect Path Dependence? *

John Freeman

July 13, 2012

That path dependence is a key feature of human systems now is well-recognized by students of politics as well as by other social scientists. Progress has been made in clarifying the concept of path dependence and related ideas like equilibrium dependence (Page 2006). Thanks to the work of historical sociologists the importance of initial conditions–"contingent events"–is clearer now as well (Mahoney 2000). But, with some notable exceptions we don't know how (if) path dependence is manifest in data.[1] The empirics in much of this genre amount to analyses of ball-urn models and to historical narratives. Studies of path dependence provide little guidance about how to interpret statistical results. For example, a well-established argument in political science is that macropartisanship is a "running tally" of political shocks (Fiorina 1981; MacKuen et al 1989). Erikson et al (1998: 904-5, 909) contend that micropartisanship–an "individual's equilibrium partisanship"–also is a random walk. But, even if this can be established statistically for macro and(or) micro data, is a random walk evidence of path dependence? How so?[2] Time series analysis expressly focuses on historical dependence–on "unpacking historical causality." It is rooted in a dynamic systems framework.[3] What does time series analysis teach us about path dependency? Do some time series methods include tests for path dependency? If so, do these tests indicate that processes like macro partisanship are, in fact, path dependent? If so, in what sense?

*Originally prepared for the NSF sponsored conference on path dependence, University of Minnesota, June 4-5, 2010. The author has benefitted greatly from discussions with John Jackson, the comments of Patrick Brandt, and the research assistance of Geoff Sheagley and Andrew Lucius. This is the fifth draft of the paper. Comments are welcome.

[1]The exceptions include Jackson and Kollman 2010, 2012 and Franzese et al 2012. These contributions are discussed further below.

[2]In fact, Green et al (1998: 886-7) used unit root tests to assess the nonstationarity of macropartisanship; these tests did not resolve the debate. Page (2006, 97, fn. 8; 98, 104) notes that different econometric methods are needed to distinguish phat from path dependency and also that time series regression models imply historical dependence. But he does not explain which econometric methods are most useful or how (if) time series regression is best used to establish path dependency. Unfortunately, recent methodological contributions to the study of path dependence like Bennett and Elman (2006) and Vergne and Durand (2010) also provide few useful insights into how to apply statistical analysis in the study of this subject.

[3]The phrase "unpacking historical causality" is from Page (2006). Page investigates path dependency both in dynamic systems and in decision theoretic frameworks. For brevity, I focus here only on the former framework.

This paper addresses these questions. It begins by reviewing the key concepts in the study of path dependence, attempting to make connections between them and concepts in time series statistics like ergodicity.[4] Some such connections are clear; others are less so. The note then shows that *linear* time series models illuminate certain kinds of outcome dependence. Moreover, familiar tools like unit root tests reveal data generating processes that embody *phat* rather than other kinds of dependence, more specifically, dependence on the *set* not *sequence* of previous outcomes. At the same time, there are concepts in the study of linear time series models that have no clear parallel in the social science path dependency literature, in particular, the idea of a (correction to) moving equilibrium in phat dependent processes (cointegration). Some *nonlinear* time series models also are evaluated. For example, threshold autoregressive (TAR) models connote outcome and also certain kinds of equilibrium (in)dependence. The notion of switches between different paths of adjustment to a moving equilibrium in phat dependent processes is suggested by nonlinear error correction models. Throughout we show what existing research implies about the nature of macropartisanship–why the defense of different models amount to contrasting arguments about whether and how macropartisanship is outcome and path dependent.

These ideas are illustrated in the second part of the paper by analyzing for nonlinearity the Green, Palmquist and Schickler (GPS) series for macropartisanship. The results of a battery of tests suggest that this series indeed may be nonlinear. A self-exciting threshold autogressive (SETAR) model is found that describes this series. The estimates and regime switching plot from this model are reported. And the GPS series is reinterpreted in relation to the ideas associated with path dependence.

1 Conceptualization

Two concepts are at the heart of the study of path dependence: outcome dependence and equilibrium dependence. As Page (2006) uses the terms, outcome dependence means that current realizations of some process, denoted by y_t, depend on previous realizations, $y_{t-s}, s = 1, \ldots, t$. Equilibrium dependence has to do with the long term, limiting distribution over possible states of the process. In general, in Page's (2006) framework, there are two possibilities: (1)outcome dependent and nonequilibrium dependent and (2) outcome dependent and equilibrium dependent. He calls both of these path dependence.[5]

There are several distinct kinds of outcome dependence. The first is dependence of the current value of a process on its initial conditions. Historical sociologists such as Mahoney (2000) and Goldstone (1998) define path dependence in terms of the impact of initial conditions on current outcomes; "[path dependent] outcomes are related stochastically to initial conditions" (*Ibid.*p. 834). Frequently this definition is extended to include the sequence of

[4]For simplicity, I focus on the likelihoodist tradition in time series analysis. I thank Jeff Gill (2011) for explaining the relative virtues of this term rather than "frequentist".

[5]Note that Jackson and Kollman (2010, 2012) argue that only the latter case should be conceived as path dependence.

the early outcomes in addition to initial conditions.[6] Page (2006) refers to phat dependent processes in which the *set* but not the sequence of early outcomes affects the current outcome. The *sequence* of past events affecting the current outcome is the other possibility. As we show below, familiar time series and regression models embody outcome dependence on initial conditions and on the set and sequence of past events.[7]

Equilibrium dependence has to do with the "limiting behavior" of a process, more specifically, whether a process fails to converge to a fixed probability distribution over possible outcomes. If a process fails to exhibit this kind of convergence it is equilibrium dependent; if it converges to a fixed probability distribution over outcomes, it is equilibrium independent. Of course, outside of experimental settings, we never observe our processes in a steady state ("lock-in").[8] The social systems we observe are constantly buffeted by endogenous, exogenous, and stochastic variables. Asymptotic theory gives us results about convergence in distribution of estimators such as the sample mean. The Central Limit Theorem and Functional Central Limit Theorems are examples. Such theorems tell us how the estimators are distributed as the number of observations grows. Also, as we will see, for some simple univariate time series processes, the distribution of the process can change in time; as history unfolds (t increases), the distribution of the variable actually changes. Neither this kind of behavior nor the asymptotic behavior of estimators is necessarily the same thing as limiting behavior of distribution over *outcomes*, however.

The use of the term ergodicity illustrates how conceptualizations employed by students of path dependence in social science and of time series statisticians can differ. Consider univariate processes. Writers like Page (2006: 95) and Vergne and Durand (2010 754) use ergodicity to define equilibrium (in)dependence. In doing so they reference Markov chains. For instance, Page (2006: 95) defines ergodicity in terms of state dependence, the possibility of writing a mapping of each history into one of N states. In a state dependent process the outcome in any period depends only on the state of the process at that time. Its state transition rule is the same in each period. A state dependent process is ergodic if through some series of states it is possible to get from one state to another. Repeated iterations of the chain produce a fixed probability vector over the possible outcomes. This constitutes equilibrium independence. If repeated iterations fail to produce such a vector–as in Page's Strong Path Dependence Process (Example 6, 2006: 102-3)–the process is equilibrium dependent.[9]

[6] Of course, students of dynamic processes–including deterministic dynamical systems, have shown that different initial conditions can produce wildly different, sometimes even chaotic, behavior. See, for instance, Tong 1990: Section 2.11.

[7] Page's Founder Process–a ball-urn process in which one of two balls is chosen, replaced, and then the other ball is removed–illustrates early outcome dependence. Page's Polya Process and Burden of History processes illustrate set(phat) and sequence outcome dependence, respectively.

[8] The term "lock in" is used by Jackson and Kollman 2012.

[9] In their recent work on the idea of revised path dependence Bednar, Page, and Toole show that revision can produce processes that converge to a *single* outcome; in other words the limiting probability mass function is a spike of value one for one outcome. Vergne and Durand (2010: 754) define path dependence in terms of nonergodic Markov chains, that is, Markov chains that are either nonirreducible, nonperiodic, or nonpositive recurrent. Technically, whether a Markov Chain is ergodic depends on the eigenvalues (eigenvectors) of

Some texts on time series use the idea of an ensemble to define ergodicity.[10] Say that for a random variable, Y_T we produced a collection of series (realizations) *each* of length 1,..., T. That is, we produce the first sequence and label it with a 1 superscript, $y_1^1, y_2^1, ...y_T^1$. We then produce a second sequence of realizations labeled with a 2 superscript, $y_1^2, y_2^2, ..., y_T^2$. If we do this I times, we would have an ensemble of I realizations. This ensemble then would be a collection of I possible histories of our process. For any y_t–a hypothetical slice across the sequences at the same time t–we could find the *ensemble average*: $\frac{1}{I}\sum_{i=1}^{I} y_t^i$. In fact, there is a formal definition of the expected value of this time slice of our process at time t, $E(Y_t)$, in terms of the probability limit of the ensemble average. The idea of the probability of the ensemble average converging in probability limit to $E(Y_t)$ connotes a kind of limiting behavior although not necessarily "equilibrium behavior" of the type associated with studies of path dependence.[11]

The corresponding *time average* for a single realization, say the first realization, can be written in formal terms as

$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t^{(1)}. \tag{2}$$

A univariate covariance stationary process is ergodic in the mean then if the expression in (2) converges in probability to the expected ensemble average $E(Y_t)$ as $T \to \infty$. Technically, a time series process will be ergodic in the mean as long as the autocovariances, γ_j, become zero valued sufficiently quickly as j increases, formally, if $\sum_{j=0}^{\infty} |\gamma_j| < \infty$. There is a similar expression for expressing the ergodicity of the second moment of the process. If Y_t is a stationary Gaussian process this condition is sufficient to ensure the ergodicity of all its moments. But it also is possible that a process is stationary and not ergodic, that is, for a given time series process, the ensemble and time averages can be different.[12]

its transition probability matrix, P. Because the rows of P add to one, P will have one eigenvalue equal to unity and one eigenvector that is the vector **1**. If the other eigenvalues are inside the unit circle, the Markov chain is ergodic. Denote the vector of erogodic probabilities for the chain as π. Then $P\pi = \pi$ and $lim_{m\to\infty} P^m = \pi 1'$. In this way, the discrete probability distribution over the possible states converges to the vector π. See, for instance, Hamilton (1994: Chapter 22).

[10] The following passage is a summary of the opening pages of Hamilton's (1994) third chapter.

[11] The expectation of the ensemble observation at t, provided it exists, can be written:

$$E(Y_t) = \int_{-\infty}^{\infty} y_t f_{Y_t}(y_t) dy_t. \tag{1}$$

This equation can be interpreted as the probability limit of the ensemble average. Using a similar expression for the ensemble variance, γ_{0t}, the jth autocovariance can be derived; it too can be interpreted as a probability limit of an ensemble average. Finally, covariance stationarity can be defined in terms of these quantities, the ensemble mean and autocovariances.

[12] Hamilton's example of a stationary but nonergodic process is the following. Let the mean $\mu^{(i)}$ for the ith realization of $[y_t^{(i)}]_{t=-\infty}^{t=\infty}$ be generated from a normal distribution with mean zero and variance $\gamma^2, N(0, \gamma^2)$. So the process is defined by the equation

$$Y_t^{(i)} = \mu^{(i)} + \epsilon_t \tag{3}$$

where ϵ_t is a Gaussian white noise process with mean zero and variance σ^2, $N(0, \sigma^2)$, that is independent of $\mu^{(i)}$. The process in (3) is covariance stationary but its time average converges to $\mu^{(i)}$ rather than to zero, The ideas of a collection of possible histories and of ensemble and time averages bear some relation to the concept outcome-equilibrium (path) dependence. In fact, Bednar et al (2012), employ the idea of a time average to define the equilibrium value of their new revised-path dependent process. But is nonergodicity in this sense the same as path dependence? The answer appears to be no. For example, the ensemble average could or could not reflect initial conditions. The same is true of the time average. There is no clear parallel in the path dependence literature to the idea that the expected ensemble and long-term time averages are unequal. To my knowledge, no student of path dependence argues that the expected value of an ensemble average of a stationary social process like macropartisanship fails to equal the long term average of the same process (fn 8). Rather, path dependence scholars emphasize the failure of a process to converge to a fixed probability vector over all possible outcomes or states(see fn. 5). So, it appears that, for univariate processes like those studied in ball-urn processes, there is a difference in the way the term ergodicity is used by social scientists who analyze path dependence and some time series statisticians.[13]

The ideas of stochastic stability and of stationary (density) distribution are more fruitful bridges to social scientists' concept of path dependence. Because of their importance, two explanations of these ideas are presented here. The first is from Tong (1990: Chapter 4). Recall that many time series models are the stochastic difference equations. In the univariate case the model has the form

$$X_{t+1} = f(X_t, \epsilon_{t+1}), \qquad t \in Z_+ \tag{4}$$

where $f : R^2 \to R, \epsilon_t$ is IID, and Z_+ is the set of positive integers. Tong explains that this expression defines a Markov chain with the state space of R. In fact, he argues that the idea of stochastic stability is *implicit* in the theory of Markov chains (*Ibid*, p. 122-3, see also p. 97). This provides a more direct connection between the conceptualization in works like Page (2006) and works in time series statistics, a clearer link between the way

the mean of Y_t.

[13]For example, Bednar et al (2012) go back and forth between discussing the expected value of their revised path dependent process and the possibility that this process converges to a singe value, by implication, a probability mass function with value 1 over one outcome. Jackson and Kollman (2012) emphasize convergence of the mean but talk about convergence in distribution of their single equation, bivariate dynamic process. Franzese et al (2012) refer to a steady state (equilibria) in which the types of and ties between actors in their network do not change. The idea of expected value (moments) nonetheless appear to inform the study of path dependence. Jackson and Kollman in their recent articles (2010, 2012) show that the familiar partial adjustment model produces outcome dependence and yields information about the long run behavior of social processes. They use the model $y_t = \rho y_{t-1} + (1 - \rho)X_1B + \mu_t$ where ρ and B are parameters, X_1 is the initial value of the exogenous variable X, and μ_t is an error term. Through recursive substitution in this model, they show how such a process, at a given time and over the long term, is or is not dependent on (1) initial conditions, (2) the value (history) of an exogenous variable and (3) the history of shocks. Jackson and Kollman also show how the expected value of such a dynamic process, depends on these same factors. Their analysis thus produces insights about the average value of y_t is a function of X_t. These results are closer to Bednar et al's idea of using the concept of long run time averages to assess the equilibrating behavior of a process. But, again, this is not the same as demonstrating (non)convergence to a fixed probability vector over all possible outcomes of a process.

path dependence theorists in the social science and time series statisticians conceive of social dynamics. Now, the "skeleton" of the time series process in (4) is the part that does not depend on the stochastic term. For instance, set ϵ_t equal to zero for all t. We then have

$$x_{t+1} = f(x_t, 0) = \phi(x_t), \qquad t \in Z_+. \tag{5}$$

The familiar theory of difference equations can be used to study the stability of the solutions (sequences) of equation (5). The idea of stochastic stability emerges from a generalization of this theory. It comes from analysis of the trajectories in a space $\mathcal{L}$ of probabilities (measures) on R defined by equation (4) for an initial probability (distribution), μ_0 for X_0. Denote by $[\mu_t(\mu_0) : t = 1, 2, \ldots]$ the trajectory in $\mathcal{L}$ starting with the initial probability $\mu_0 \in \mathcal{L}$. So $\mu_t(\mu_0)$ is the probability of X_t conditional on X_0 having distribution μ_0. The stability of $(\mu_t : t \in Z_+)$ is what is called "stochastic stability."[14] Ergodocity in this context has to do with the existence and uniqueness of a probability distribution π on R to which $\mu_t(\mu_0)$ converges as $t \to \infty$. If the rate of convergence is geometric the process is said to be "geometrically ergodic." And, if X_t is stationary, π is said to be the stationary distribution. More generally, if the skeleton of the stochastic different equation exhibits exponential stability in the large of its equilibrium points, the Markov chain defined by the corresponding equation is geometrically ergodic (*Ibid.* p. 126; Table 4.1). Tong explains (*Ibid.* Section 4.2) that, for stationary time series, the evaluation of a stationary distribution is non-trivial. Closed form solutions exist only for some special cases; numerical methods often must be used to obtain the stationary (density) distributions. However, an implicit solution is available, if the series has an ergodic Markov chain over R^n. Below we give the conditions for stochastic stability and depict the stationary distribution for some illustrative threshold autoregressive (TAR) models.[15]

A second, slightly different conceptualization of stability is offered by Granger and Teräsvirta (1993: Section 1.6). Consider the nonlinear, first order autoregressive model

$$y_t = f(y_{t-1}) + \epsilon_t \tag{7}$$

where y has the initial value y_0 and ϵ_t has a mean of zero is i.i.d. Assume that there is a solution to this equation of the form

$$y_t = y(\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_1; y_0). \tag{8}$$

[14] The first Appendix of Tong's (1990) book, which is written by the statistician K.S. Chan, gives a more technical definition. Chan's Appendix explains the deeper links to Markov chain theory.

[15] The "drift criterion" for the existence and uniqueness of a stationary distribution is explained by Tong. Say that X_t is an n dimensional stationary time series. And assume X_t an ergodic Markov chain over R^n. The implicit solution for the stationary distribution then is

$$\pi(A) = \int_{-\infty}^{\infty} P(A|x)\pi(dx) \tag{6}$$

where π is the limiting distribution of X_t (which Tong takes as the initial distribution hence π is the stationary distribution), A is a Borel set of R^n and $p(\cdot|\cdot)$ is the conditional (i.e., transition) probability.

Denote the conditional probability distribution function of y_t given y_0 as

$$Prob(y_t \leq x | y_0) = \Phi_t(x; y_0). \tag{9}$$

Now, the time series process, y_t is called "short memory" if, "after running a long time," the initial value does not affect its marginal distribution, or, formally, if $\Phi_t(x; y_0) \to \Phi_h(x)$ for $t \to h$ and h large. The process is called *stable* if

$$\lim_t \Phi_t(x; y_0) = \Phi(x; y_0), \tag{10}$$

that is, a process is stable if two values y_t, y_s have the same marginal distribution for t,s large but this marginal distribution does not depend on y_0. So, in this conceptualization, short memory is related to the idea of path independence insofar as the current marginal probability distribution does not depend on the initial condition. However, stability does allow the initial condition to affect the long term marginal probability distribution. But, like the idea of equilibrium independence in the social science literature, the marginal probability distribution is the same in the long term.[16]

Table 1 summarizes some of the properties of selected time series models relative to these ideas in the study of path dependence. Its contents are explained in the next section.

2 Linear Time Series Models

Linear time series models with constant coefficients embody all the concepts associated with outcome dependence. Their connection to the idea of equilibrium dependence is less clear. These models also illuminate an idea that apparently is not captured by the writing on path dependent dynamic systems or on historical sociology: common trends in phat outcome dependent processes (cointegration). Vector error correction models can have multiple moving equilibria (cointegating vectors) composed of phat outcome dependent processes (variables).

2.1 Univariate Linear Time Series Models

Consider the simple first order autoregressive model with constant coefficients. This model can be written

$$y_t = a_0 + a_1 y_{t-1} + \epsilon_t \tag{11}$$

where a_0, a_1 are constants, and ϵ_t is a white noise process. This is a stochastic difference equation. It can be solved in several ways. The solution is

$$y_t = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0 + \sum_{i=0}^{t-1} a_1^i \epsilon_{t-i} \tag{12}$$

.

[16]There are are various definitions of "stable equilibrium" in the time series literature but these appear to apply to the mean and deterministic component of the process. So the idea of equilibrium appears to be somewhat different than that used in the social science path dependence literature (see Granger and Teräsvirta 1993 p. 13).

This solution manifests outcome dependence insofar as the initial condition has an impact on the current value of y_t as does the the *sequence* of previous shocks. This sequence of previous shocks is *weighted* by a_1 raised to different powers of t. But does this simple linear univariate model connote. In this case, as the number of observations grows, as $t \to \infty$, we have

$$\lim y_t = \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \epsilon_{t-i} \tag{13}$$

Taking expectations of both sides of this equation, we obtain $E(y_t) = \frac{a_0}{1-a_1}$, a finite and time independent value. It is easy to show that with $|a_1| < 1$ the variance of y_t also is finite and time independent (Enders, 2010: 55-56). So the expected value of the process does not depend on initial condition, set, or order of shocks. For $|a_1| < 1$ this process therefore is outcome dependent insofar as current values depend on the initial condition and sequence of shocks, but it has a fixed expected value that does not depend on the initial condition or either the set or sequence of shocks. As regards limiting behavior, Hamilton (1994: Chapter 3) shows that, by viewing the stationary AR(1) process as a $MA(\infty)$ process, one can establish this AR(1)process is ergodic in all of its moments. From the central limit theorem, y_t is normally distributed N(0,V) where $V = \frac{\sigma_\epsilon^2}{1-a_1^2}$. Granger and Teräsvirta (1990: 12) cite this fact in explaining that when $|a_1| < 1, y_t$ is, by their definitions, both short memory and stable. The conception of macropartisanship represented in the stationary AR(1) process defended by Green et al (1998) thus implies outcome dependence but also equilibrium independence (stability).[17]

The random walk model, the simplest version of the "running tally" conception of macropartanship (Fiorina 1981, MacKuen et al. 1989), in contrast, is early outcome dependent and phat outcome dependent. It also is an equilibrium dependent process. The random walk model has a fixed expected value. It can be written:

$$y_t = y_{t-1} + \epsilon_t. \tag{14}$$

Its solution is simply

$$y_t = y_0 + \sum_{i=1}^{t} \epsilon_{t-i}. \tag{15}$$

The solution in (15) is different from the solution for the simple autoregressive model in (12). First, note that in the solution in (8) the initial condition does not disappear as t grows. In fact, $E(y_t) = y_0$. So, in this sense, the initial condition has a lasting impact on the time average of the random walk. Put another way, although the realizations in this process' ensemble can display different kinds of behavior (cf. Figure 1), the expected value is always

[17] Box-Steffensmeier and Smith (1996) argue that macropartisanship is fractionally integrated. This too implies outcome dependence in the sense that the sequence of shocks affects the current value of the series. But due to the stationarity of fractionally integrated systems, they too embody outcome but not equilibrium dependence.

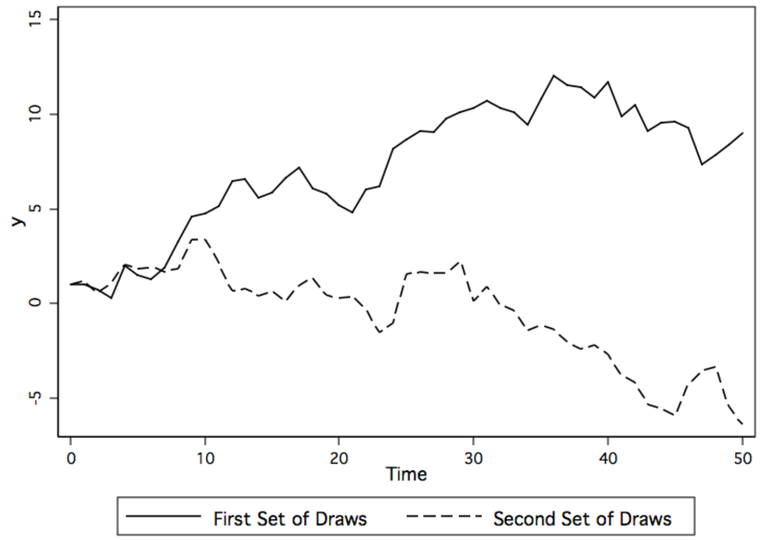

Figure 1: Two Realizations of a Random Walk Process (A two realization ensemble)

the same (y_0).[18] Second, because the shocks now are all equally weighted, their impact is more akin to phat than sequence outcome dependence. In this sense, the "running tally" idea, as operationalized in some of the work by McKuen et al (1989) and, for micropartisanship, in Erikson et al (1998), is closer to a ball-urn model of the Balancing Process type; it is not a Polya process because running tally implies phat but not equilibrium dependence.[19] It follows that tests for unit roots in univariate series, in effect, are tests of early outcome dependence and of phat outcome dependence. These tests therefore illuminate distinct kinds of data generating processes (cf. Page 2006: 97).

Finally, the distribution of the random walk process actually changes over time. Consider the case in which $y_0 = 0$ and the errors are i.i.d., distributed $N(0, \sigma^2)$. Then from (15) it follows that $y_t \sim N(0, \sigma^2 t)$. This means that at each time point, y_t has a normal distribution with the same mean but with a different variance. It is not clear that Fiorina (1981), MacKuen et al (1989) made any argument of this kind for macropartisanship, nor that Erikson et al (1998) make such an argument about individual partisanship. Either way, this appears to be closer to the kind of "limiting behavior" that Page (2006) and others associate with path dependence.

[18]The same is not true of its second moment which varies in time. Hence the random walk is a nonstationary process. See, for instance, Enders (2010: 185-6.

[19]Page's (2006:99) Balancing Process is a follows: Initially an urn contains one maroon ball and one brown ball. In any period, if a brown (resp. a maroon) ball is selected then it is put back in the urn together with an additional ball of the opposite color. This process is outcome phat dependent and it has a unique equilibrium. Again, whether a fixed expected value is synonymous with equilibrium independence is not clear.

2.2 Multivariate Linear Time Series Models

These models can be divided into strongly and weakly restricted varieties (Freeman et al 1989). Among other things, strongly restricted models are based on investigator imposed exact restrictions for exogeneity and lag length.

2.2.1 Strongly Restricted Multivariate Time Series Models

There are at least two categories here. The first is the familiar, single equation time series regression model. One of the most common models of this kind is:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \sum_{i=0}^{p} \beta_i x_{t-i} + \epsilon_t \tag{16}$$

This model is strongly restricted insofar as only one lag of the endogeneous variable is included on the right hand side of the equation, the x variable is assumed to be exogenous, and p lags of x are stipulated to be causally related to y. Page argues that "if a regression equation does not include any time lags it captures only phat dependence." He contends that retrospective voting models capture "recent path dependence" (2006: 98, 104). For Page, path dependence is embodied in the model in (16) because the sequence of the realization of the x variable (*not* the shocks) affects the values of y_t. Whether this kind of dependence also is of the "recent" type supposedly depends on the magnitudes of the β coefficients.[20]

But what does this equation tell us about equilibrium dependence? If y_t always depends on past x_t the long-term value of y_t constantly varies. So equation (16) connotes both outcome dependence and a kind of nonequilibrating behavior. The initial condition for the exogenous variable, x_0, probably is not important since its weight decreases as time increases.[21] The idea of a Polya type equation that manifests phat dependence is embodied in a simpler version of equation (16):

$$y_t = \alpha y_{t-1} + \beta x_t + \epsilon_t \tag{17}$$

Here the x variable has no lags so, in Page's framework, it connotes phat dependence. But the multipliers associated with this equation connote a form of equilibrating behavior on changes in the exogenous x variable; *one-time* increases in the values of x produce different long term values in y_t. [22] As pointed out in footnote 13 above, Jackson and Kollman (2010, 2012) already have provided an in depth explanation of how (when) these and related

[20]Page presents a useful example in which the dependent variable is vote for an incumbent congressperson and the independent variable is a measure of that individual's ideology, a measure based on the *set* not order of the congressperson's roll call votes. With no lag of this measure on the right hand side of the respective regression equation, the model captures phat dependence. Presumably, lagging this same measure means that the vote depends on the *sequence* of ideology scores and therefore comes closer to capturing path dependence. No example of a retrospective voting model that captures path dependence is provided by Page.

[21]Technically, the initial value of the dependent variable, y_0, also must be considered. But, again, if $|a_1| < 1$, this initial value should decay.

[22]For a useful study of the nature and problems of estimating equation (6), see Keele and Kelly 2005.

regression models embody path dependency.

The second type of strongly restricted model specifies a relationship between two random walks or, to be more precise, between two processes of the same order of integration.[23] This is the single equation, error correction model (ECM). It is based on the idea that even though two processes may each be nonstationary, if they are integrated of the same order, a weighted sum of them may be stationary. This weighted sum connotes a type of "equilibrium." When the weighted sum is zero, "equilibrium" is achieved; if the process is not in equilibrium, supposedly some social mechanism attempts to restore it to equilibrium or to eliminate the error (deviation from the equilibrium).[24] In the single equation case, short term changes in one integrated variable are a function of weighted lagged changes in itself, weighted lagged changes in an exogenous variable–a variable integrated of the same order, and an *error correction term*. An example of such a single equation model is

$$\Delta r_t = a_{10} + \alpha[r_{t-1} - \beta s_{t-1}] + \sum_{i=1}^{p} a_{11}(i)\Delta r_{t-i} + \sum_{i=1}^{p} a_{12}(i)\Delta s_{t-i} + \epsilon_{1t} \tag{18}$$

Here a_{10} is a constant, r_t and s_t are both assumed to be random walks (first order integrated and also cointegrated), s_t is assumed to be exogenous, and p lags of the short term changes in each variable are stipulated. The second term on the right hand side of (11) contains the cointegrating vector or equilibrium relationship; α is the rate of error correction. So when the system is in equilibrium, $r_{t-1} = \beta s_{t-1}$, there is no error correction in r_t, and this term does not produce a change in the left hand side variable. However, the other terms on the right side of (11) may affect Δr_t. The sum of the lagged past changes in r_t and(or) in s_t may be nonzero. The Granger Representation Theorem holds that when processes are cointegrated such an error correction model exists. Erikson et al (1998) posit a model of this kind to explain macropartisanship in terms of (purged) approval and consumer sentiment.

What is the connection, if any, between ECMs and path dependence? Recall from above that random walks embody dependence on the initial condition. They also embody phat outcome dependence. The expected value of a random walk–its unconditional mean–is equal to its initial condition. The distribution of a random walk changes in time. The ECM model therefore is a combination of processes. Changes in the left hand side variable depend on the initial conditions of both variables via the cointegrating term (since initial conditions are contained in each of the two integrated processes). Phat and sequence outcome dependence are present in these processes as well since each variable by itself depends on the set of past shocks it experiences, and together, their weighted sum(difference) depends on the sequence of shocks that the two processes experiences–whether and when they combine to produce zero "error". As regards limiting behavior, the ECM moves into and out of this "equilib-

[23]The focus here is on the time series statistics rationale for the ECM functional form. Many social scientists today simply posit this functional form without, for example, testing whether their variables are integrated or cointegrated.

[24]Enders (2010: 358-9) equivocates about the meaningfulness of the idea of equilibrium in this context. He provides several examples of social processes that exhibit moving equilibrium of the ECM variety but then says the term "equilibrium" is unfortunate.

rium" when this error is zero. But, once more, even if the variables combine to produce "equilibrium" (zero error), the the sum of lagged changes in *both* of them still can produce movements in the left hand side variable. Hence the ECM process is not at rest even if it is "in equilibrium."

Simply put, there does not appear to be any comparable idea in historical sociology or Page's (2006) dynamic systems framework for this model. The mix of dependencies embodied in the ECM appear to have no parallel in the literature on path dependence.

2.2.2 Weakly Restricted Multivariate Time Series Models

The most familiar formalisms of this kind are the vector autoregression (VAR) and vector error correction (VECM) models. The restrictions in these models are relatively weaker insofar as analysts let the data choose lag lengths, exogeneity assumptions are avoided, etc. [25] The general form of the VAR model with p lags, VAR(p), is:

$$y_t = AY_{t-1} + B_0 x_t + u_t \tag{19}$$

where y_t is a K x 1 vector of endogenous variables, A is a K x Kp matrix of coefficients, B_0 is a K x M matrix of coefficients, x_t is a M x 1 vector of (presumed) exogenous variables, u_t is a K x 1 vector of white noise shocks, and Y_t is a the Kp x 1 matrix denoted by $Y_t = \begin{pmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{pmatrix}$.

If the modulus of each eigenvalue of the matrix A is strictly less than one, the estimated VAR is stable.[26] Enders (2010: 295ff) explains the stability conditions for a simple, two variable model with one lag. He calls this the VAR model in standard form:

$$x_t = A_0 + A_1 x_{t-1} + \epsilon_t \tag{20}$$

where x_t is the 2 x 1 vector of variables, A_0 is a 2 x 1 vector of constants, A_1 is a 2 x 2 matrix of constant coefficients, and ϵ_t is a nx1 vector of white noise shocks. He shows that the solution of this equation can be written

$$x_t = \mu + \sum_{i=0}^{\infty} A_1^i \epsilon_{t-i} \tag{21}$$

where μ is a 2 x 1 vector of the means of the two variables. So, once again, if the relevant coefficients are less than one in absolute value, the expected value of this dynamic process is the mean of each series, μ. In this sense, the stable VAR(p) model also implies sequence outcome dependence.

The vector error correction model, VECM, is designed to analyze a system of variables, a

[25] Note the word "relatively." As equation (19) shows, even VAR models contain restrictions including, sometimes, presumed exogenous variables. Once more, the focus here is on likelihoodist models.

[26] The notation for the general version of the VAR(p) is taken from Lütkepohl(1993). STATA9 provides a test of this stability condition under the rubric varstable. Certain conditions regarding the initial conditions of the dynamic system also must be satisfied.

system which may contain *multiple* cointegrating vectors.[27] Consider the system of equations:

$$x_t = A_1 x_{t-1} + \epsilon_t \tag{22}$$

where x_t is a n x 1 vector of variables, A_1 is an n x n matrix of parameters, ϵ_t is a n x 1 vector of shocks. Subtract x_{t-1} from each side of equation 12 and define I as the n x n identity matrix. The result is:

$$\begin{aligned} x_t &= -(I - A_1)x_{t-1} + \epsilon_t \\ &= \pi x_{t-1} + \epsilon_t \end{aligned}$$

where π is the n x n matrix $-(I - A_1)$. If the rank of π is zero, the system amounts to a set of independent, first order integrated variables. In other words, we have an independent set of phat dependent processes for which the respective initial values of the variables do not decay. If the rank of π, r, greater than zero but less than n, there are r cointegrating vectors. That is, there are r moving equilibria between the phat outcome dependent variables. Once again, this system can be in "equilibrium" when all the cointegrating vectors are zero, but not at rest because of changes induced by the inclusion of the sum of lagged changes on the right side of (15).

To my knowledge, neither Erikson et al (1998) or other scholars have explored the possibility of multiple moving equilibria in the system that explains macropartisanship. [28]

3 Nonlinear Time Series Models

Nonlinear time series models describe processes which exhibit asymmetries and(or) sudden bursts in amplitude at irregular intervals. Nonlinear models also are useful for analyzing time series processes that are characterized by time irreversibility (Tong 1990: Section 1.5).[29],[30]

3.1 Univariate Nonlinear Time Series Models

The threshold autoregressive model, TAR, or self-excited threshold, SETAR, model, is one of the most widely used nonlinear time series models. It has been has been employed to a variety of physical and social processes. Tong (1990) argues that thresholds are generic concepts. He shows how the SETAR model can be used to model sunspot and animal population

[27] The following presentation of the VECM model is summarized from Enders (2010: 371ff)

[28] A study in American politics more sensitive to this possibility is Ostrom and Smith (1993). For a Bayesian approach this is more consistent with this idea see Brandt and Freeman (2009).

[29] The conditions for stochastic stability for nonlinear models of the form $y_t = f(y_{t-1}, \epsilon_t)$ are summarized by Granger and Teräsvirta (1993: 12; their original source is Lasota and Mackey 1989). A more general treatment of stochastic stability and of stationary (densities) distributions for nonlinear time series models is Tong (1990).

[30] There are a variety of nonlinear time series models including generalized autoregressive (GAR), the bilinear, and multiple forms of threshold autoregressive models such as STAR, LSTAR, and ESTAR. Still another is the markov switching time series model. For an introduction to these models see such works as Enders (2010: Chapter 7) and Granger and Teräsvirta (1993). The latter source cites Quinn (1982) for the conditions for the stability on the bilinear model. But it also notes that stability results are not always available. A more general review of nonlinear models is Tong (1990). Here I focus on a few simple examples of such time series models.

(lynx) series. Enders (2010) reviews TAR models of unemployment, models that capture the fact that when an economy is in a recession and unemployment is above a threshold the speed of recovery (job growth) might be slow whereas if unemployment is below a threshold unemployment might gravitate towards its long term equilibrium much more rapidly. The Simple TAR model is:

$$y_t = \begin{cases} a_1 y_{t-1} + \epsilon_t & \text{if } y_{t-1} > r \\ a_2 y_{t-1} + \epsilon_t & \text{if } y_{t-1} \leq r \end{cases}$$

where r is the threshold. This data generating process is a combination of two simple AR(1) processes. Which AR(1) process occurs depends on whether the previous value, y_{t-1} is above or below its threshold, r. The Simple TAR model will exhibit sequence outcome dependence but in different ways depending on which of the two regimes apply. Under certain conditions, in both cases, however, the process has the same expected value, namely, zero. So its limiting behavior is indicative of equilibrium independence. In fact, it can be shown that this model is geometrically ergodic if $a_1 < 1, a_2 < 1$ and $a_1 a_2 < 1$ (Tong 1990: 130-1).

A slightly more complicated version of the TAR model allows for each AR processes to have different constants and different errors terms. The expected values of the two AR process then are distinct as are the variances of the errors. This could be called a Basic TAR model.[31] An example of of such a model is:

$$y_t = \begin{cases} a_{10} + a_1 y_{t-1} + \epsilon_{1t} & \text{if } y_{t-1} > r \\ a_{20} + a_2 y_{t-1} + \epsilon_{2t} & \text{if } y_{t-1} \leq r \end{cases}$$

In this Basic TAR model, under certain conditions, each AR process can have a different expected value, either $\frac{a_{10}}{1-a_1}$ or $\frac{a_{20}}{1-a_2}$. So the process will exhibit two different patterns of sequence outcome dependence and, at the same time, its limiting behavior will switch between adjustment to two different long term values. The conditions for geometric ergodicity of such models have been derived by Chan et al (1985); one such condition is $a_1 < 1, a_2 < 1, a_1 a_2 < 1$. Consider, for purposes of illustration, the following model

$$y_t = \begin{cases} 1.5 - 0.9 y_{t-1} + \epsilon_t & \text{if } y_{t-1} > 0 \\ -0.4 - 0.6 y_{t-1} + \epsilon_t & \text{if } y_{t-1} \leq 0 \end{cases}$$

Tong (1990: Section 4.2.4.3) shows how a numerical method can be used to estimate the stationary density of this particular process. This density is depicted in Figure 2 above.

Denote macropartisanship by M_t. Then we might have a nonlinear a data generating process in which the switch occurs when M_{t-1}, exceeds a level such as .60. In other words, when the data generating process is in the first regime and the level of Democratic parti-

[31] My nomenclature differs somewhat from Enders (2010: 439) who uses the word "Basic" to describe a TAR model with no constants. I called this the "Simple TAR model" above. In addition, in his Introduction, Enders (2010: 429-430) uses the idea of a single, long term "attractor" for a nonlinear TAR process. Further details about this formulation are given in the Appendix in the section on testing for unit roots in a Macropartisanship series.

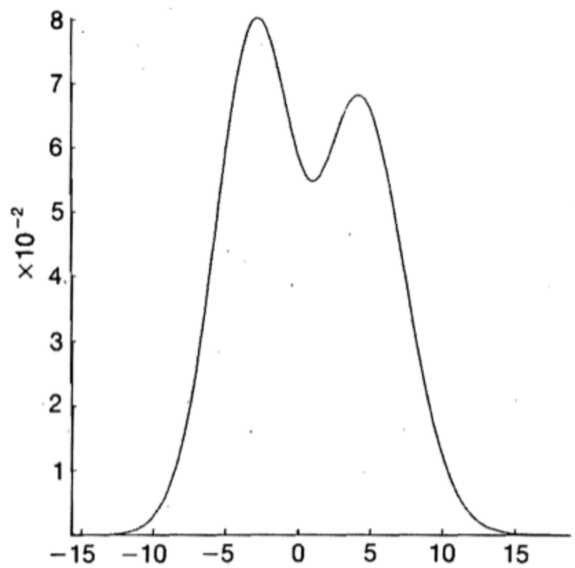

Figure 2: Stationary Density of Illustrative TAR Model. Source Tong (1990: Section 4.2.4.3)

san identification exceeds a certain threshold the American macropolity gravitates to one equilibrium. But when, because of shocks embodied in ϵ_{1t}, M_t drops below this threshold, the polity gravitates toward a different equilibrium. In each regime the macropolity would exhibit sequence outcome dependence, M_t would gravitate to a different long term fixed value. But, overall, the macropolity would be stochastically stable like the process depicted in Figure 2. With one exception (Jackman 1987) this kind of outcome dependence and limiting behavior appears not to have been explored in the literature. In fact, the data analysis below indicates that the Gallup based measure of macropartisanship studied by Green et al.(1998), Box and Smith (1996) and Erikson et al (1998) has this kind of behavior.

Some of these models allow for nonstationary behavior. One of the most simple is the Equilibrium-TAR model:

$$x_t = \begin{cases} x_{t-1} + \mu_t & \text{if } |x_{t-1}| < k \\ \rho x_{t-1} + \mu_t & \text{otherwise} \end{cases}$$

where ρ is a constant, $|\rho| < 1$, and $\mu_t \sim N(0, \sigma_\mu^2)$.

A somewhat more complex model of this kind is the Band-TAR model. Enders (2010: 446) provides an illustration. Let $s_t = r_{Lt} - r_{St}$ be the spread between long and short term interest rates. Assume this spread follows a simple AR(1) process with constant coefficients, more specifically,

$$s_t = a_0 + a_1 s_{t-1} + \epsilon_t \qquad (23)$$

where ϵ_t is the familiar white noise error process. Assume further than the AR(1) process is covariance stationary hence its expected value is $\frac{a_0}{1-a_1}$. Call this long-run value $\bar{s}$. This allows us to rewrite (1) as an adjustment process of the form

$$s_t = \bar{s} + a_1(s_{t-1} - \bar{s}) + \epsilon_t. \tag{24}$$

where, again, ϵ_t is a white noise error term. Then the Band-TAR model can be expressed in the form

$$s_t = \begin{cases} \bar{s} + a_1(s_{t-1} - \bar{s}) + \epsilon_t & \text{if } s_{t-1} > \bar{s} + c \\ s_{t-1} + \epsilon_t & \text{if } \bar{s} - c < s_{t-1} \leq \bar{s} + c \\ \bar{s} + a_2(s_{t-1} - \bar{s}) + \epsilon_t & \text{if } s_{t-1} \leq \bar{s} - c \end{cases}$$

Several points should be made about these models. First, conceptually, they are associated with the idea of transaction cost or arbitrage boundaries. Agents supposedly monitor the process and decide that once the variable exceeds certain values, the (net) benefit of intervening (incurring a transaction cost) to drive it back into the intermediate (locally non-stationary) range exceeds the cost of (foregoing the intervention) and allowing the process to be (globally) non-mean reverting (Balke and Fomby 1997). Second, analysis shows that the statistical power of conventional tests for nonstationarity depend on the parameters of these models. For instance, Pippinger and Goering (1993) demonstrated how the power of the Dickey Fuller test to detect mean revision in the Equilibrium TAR model depends on ρ, k, and σ^2_μ. The wider the interval, k, for instance, the more time the process spends in the nonstationary region. Hence, even if ρ is small, the Dickey Fuller test has low power. In this context, Pippinger and Goering conceive of "equilibrium" as the *continuum* of values in the interval [-k,k] (*Ibid.*, fn. 4). Within this range, the process is equilibrium (path) dependent. Globally, however, under certain conditions, such a process actually is stationary. What is required for global stationarity is that the process be mean-reverting in the "outer regimes." This can occur even if these regimes are random walks with drifts as long as the drift parameters "act to push the series back to the equilibrium band" (Balke and Fomby 1997: 630). Once again, locally, within this band the process is equilibrium dependent while, globally, it is stationary.[32]

3.2 Multivariate, Nonlinear Time Series Models

Jackson and Kollman (2010, 2012) analyze strongly restricted, nonlinear, multivariate time series regression models in which one variable is posited to be exogenous. They show how such models can exhibit path and near-path dependence and, concomitantly, equilibrium dependence. Interested readers are referred to their articles.

As regards weakly restricted models, the idea of "threshold cointegration" addresses the possibility that two or more series are nonstationary but share a common trend(s). Enders

[32]To illustrate this point, Balke and Fomby analyze the Returning Drift (RD) Threshold Model:

$$z_t = \begin{cases} -\mu + z_{t-1} + \epsilon_t & \text{if } z_{t-1} > \theta \\ z_{t-1} + \epsilon_t & \text{if } |z_{t-1}| \leq \theta \\ \mu + z_{t-1} + \epsilon_t & \text{if } z_{t-1} < \theta \end{cases}$$

where μ is the drift parameter and the ϵ_t are mean zero random disturbances.

desribes a model of this kind.[33] Let r_{LT}, r_{St} represent the interest rate on ten year government securities and the federal fund rate, respectively. Assume each series is I(1) and that they are cointegrated. The model captures regime shifts in terms of how changes in the interest rate spread, $s_t = r_{Lt} - r_{St}$, increasing vs. decreasing, translate into different rates of error correction. In this case, there is no error correction when $s_{t-1} = \beta$. This is a threshold model of the momentum type, M-TAR:

$$\Delta r_{Lt} = \alpha_{11} I_t[s_{t-1} - \beta] + \alpha_{12}(1 - I_t)[s_{t-1} - \beta] + A_{11}(L)\Delta r_{L,t-1} + A_{12}\Delta r_{S,t-1} + \epsilon_{1t}$$
$$\Delta r_{St} = \alpha_{21} I_t[s_{t-1} - \beta] + \alpha_{22}(1 - I_t)[s_{t-1} - \beta] + A_{21}(L)\Delta r_{L,t-1} + A_{22}\Delta r_{S,t-1} + \epsilon_{2t}$$

where the α terms are adjustment coefficients, $s_t = r_{Lt} - r_{St}$, the $[s_{t-1} - \beta]$ terms are cointegrating vectors, the $A(L)$ terms are lag operators, and the I_t variable is an indicator function defined as

$$I_t = \begin{cases} 1 & \text{if } \Delta s_{t-1} > 0 \\ 0 & \text{if } \Delta s_{t-1} \leq 0 \end{cases}$$

For this model then, the rate of adjustment to the moving equilibrium between the two phat outcome dependent processes varies depending on whether in the previous period s_t was increasing or decreasing.

Balke and Fomby (1997) is a more general treatment of threshold cointegration. They introduce the idea of a discontinuous adjustment to long-run equilibrium, a process that adjusts to long-run equlibrium at some times but not others. Again, the motivation assumes there are agents (policy makers) that sometimes find it in their interest to force to variables to trend together while in other cases they allow two processes to diverge from long term equilibrium. They explore, in the spirit of the Engle and Granger approach, in a Monte Carlo investigation, the power and size properties of five different tests for cointegration for the Equilibrium-TAR and Band-TAR models described above and the RD-TAR model described in fn. 4. They conclude standard linear methods for testing for cointegration work well in the presence of threshold cointegration. Balke and Fomby then proceed to develop a method to detect two threshold cointegration based on the concept of arranged autoregression.[34]

[33] This example is a simplified version of an example in Enders (2010: 481).

[34] So in the Balke and Fomby (1997) the models are written in terms of the *error term from the cointegrating regression.* Sometimes this error is stationary connoting long-term equilibration of the two integrated series (cointegration), and sometimes the error is nonstationary connoting a lack of long-term equilibration (and absence of cointegration). Their simple example is the model:

$$y_t + \alpha x_t = z_t, \qquad \textit{where} \quad z_t = \rho^{(i)} z_{t-1} + \epsilon_t \tag{25}$$

$$y_t + \beta x_t = B_t, \qquad \textit{where} \quad B_t = B_{t-1} + \eta_t \tag{26}$$

where the ϵ_t, η_t are white noise disturbance terms. Then the value of ρ varies depending on the magnitude of z_t:

$$\rho^{(i)} = \begin{cases} 1 & \text{if } |z_{t-1}| \leq \theta \\ \rho, \quad |\rho| < 1 & \text{if } |z_{t-1}| > \theta \end{cases}$$

Suppose we think of r_{Lt} and r_{St} as macropartisanship and presidential approval, respectively. Suppose further than these two variables, the former representing long term macropolitical disposition and the latter short term disposition, are cointegrated. Then this model suggests that the American polity switches between two regimes each with dependence on initial conditions and phat outcome dependence and also with different rates of error correction. The limiting behavior of this nonlinear ECM process again allows for the system to be "in equilibrium" but not at rest. While some years ago, Simon Jackman (1987) explored the possibility of regime switching in a time series regression of macro political variables, no one, to my knowledge, has analyzed this possibility of nonlinear error correction, let alone its explained its relation to the concept of path dependence.

4 Illustration: The Dynamics of Macropartisanship Reconsidered

...the world is more nonlinear than we think...
Nassim Nicholas Taleb, *The Black Swan* 2007: 88.

4.1 Univariate Analyses

For years, students of the U.S. macropolity have debated the nature of macropartisanship. One school argues that macropartisanship is relatively stable and slow moving; it provides "balast" for the American political system. The opposing view is that macropartisanship or, at least the microprocesses of which it is composed, is a running tally of short term political and economic shocks. In what follows we show that this controversy can not only be usefully recast in terms of a debate about path dependency, but also that the application of nonlinear time series models casts new light on the nature of American macropartisanship.

4.1.1 The Green, Palmquist, Schickler [GPS]Critique

In their article in the 1998 volume of the APSR, GPS reported the results of a series of tests on a Gallup-based, quarterly macropartisanship series for the period 1953:2-1996:4. This series is depicted Figure 3. We will call it M_t. GPS fit a stationary linear, ARMA model for M_t. They also considered the possibility that M_t was nonstationary. To this end, they performed some Dickey Fuller tests. Although the results of these Dickey Fuller tests were ambiguous, GPS concluded that for their sample M_t was stationary. This means that

When the absolute value of the first lag in the error, z_{t-1} is less than the threshold, θ, $\rho^{(i)} = 1$, and the two I(1) variables, x_t, y_t, do not revert to a long-run equilibrium. But if the first lag of this same error, is greater than θ in absolute value , $\rho^(i) = \rho$ and $|\rho| < 1$ so the two variables do move towards some equilibrium. Balke and Fomby proceed to present the most general version of this model and then study in their Monte Carlo analyses of cointegration tests, the Equilibrium TAR, Band-TAR, and RD-TAR versions of the above model.

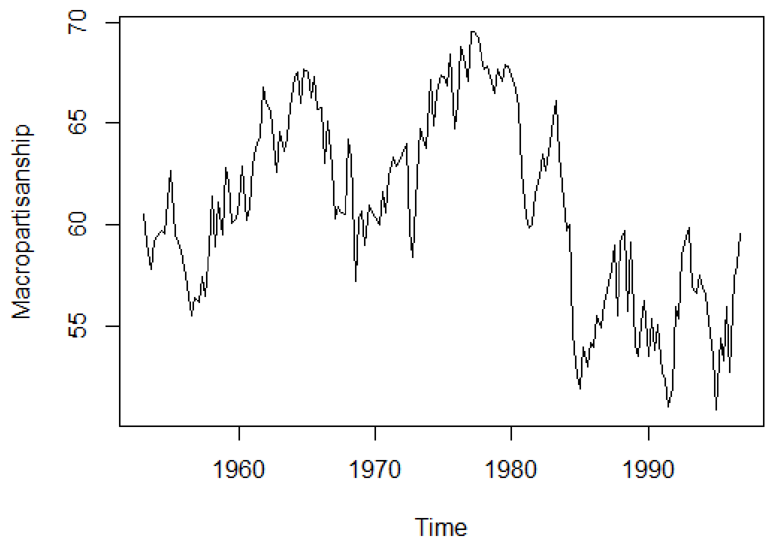

Figure 3: GPS Macropartisanship Series Based on Gallup Surveys. Source: D. Green Website, downloaded November 2011.

macropartisanship did not depend on its initial condition or on the *set* of past values of M_t, but rather only on its recent values. M_t was short-memoried and stable. Contrary to the running tally thesis, M_t was not equilibrium dependent. In this sense, M_t, argued GPS, was *path independent*.

GPS's investigation is suspect for several reasons. First, there is much evidence of temporal heterogeneity in micro partisanship. "Large changes in the state of the world [can] trigger significant systematic changes in individuals' true party utilities, leading them to change the weights given to past partisanship and to current utility assessments in updating their partisanship "(Jackson and Kollman 2011: 509). These changes can aggregate into temporal change in the coefficients in univariate models of M_t. Second, the GPS's unit root tests were not definitive. Although they made some strong arguments about why it is unlikely M_t is nonstationary, from a likelihoodist standpoint (Gill 2011), their test results did not all reject the null of a unit root. In fact, the Dickey-Fuller unit root test assumes linearity. And it is widely accepted that it and other tests for unit roots are even less powerful if the underlying process is nonlinear (Enders 2010: Section 11; Enders and Granger 1998: Section 1, Pippinger and Goering 1993). If M_t is, in fact, a nonlinear process, as explained above, it could embody a form of equilibrium dependence and hence path dependence that few students of the American macropolity have recognized.

To explore this possibility, we tested for evidence of nonlinearity in M_t and then fit a TAR model to the GPS series. Investigations of this kind usually begin descriptive analysis of the series (Tong 1990: esp. pps. 362-375). In the interest of brevity we relegate a sample of these analysis to the Appendix. Instead we follow Enders (2010: Chapter 7, Section 3) in emphasizing the results of several portmanteau tests the results of which include, implicitly or explicitly, the possibility of nonlinearity: the McCleod-Li, RESET, and BDS (delta) tests. We also implemented the test for linearity vs. the specific TAR model using a supremum F test (Enders 2010: 449-451). In this and other parts of our analysis we used a combination of RATS and R code.[35]

The first step in implementing the pretests is to estimate a linear model for the Gallup macropartisanship series. GPS found that an ARMA (1,0,1) with constant fit the data best. Using the data on Green's website, we were able to replicate GPS's 1998 estimates almost exactly. We actually found that an AR(2) model with constant had a slightly lower AIC value, however. These results are reported in the Appendix. In what follows we analyze the residuals from both linear models.[36]

Two portmanteau tests have nulls of linearity. The Regression Error Specification Test (RESET) regresses the residuals from the best fitting linear model on the regressors used in the estimating equation and *powers* of the fitted values from this equation. A F test is used to assess the joint statistical significance of the coefficients on the powers of the fitted variables.[37] The McLeod-Li test is the same as that used to detect ARCH type errors. In this case, one analyzes the sample correlation coefficients between the *squared* residuals of the best fitting linear model. A Ljung-Box Q statistic is calculated for these squared correlation coefficients. The statistic has a χ^2 distribution. Table 2 below and Figures 3 and 4 report the respective results for the residuals from both ARIMA models of M_t. In no case is there any evidence of nonlinearity.

The BDS test is a third portmanteau test. It is named after Brock, Dechert, Scheinkman, and LaBaron (1996). The BDS test analyzes the "spatial dependence" of a time series: how "close" pairs of observations m lags apart are conditional on intermediate values of the series. The distance metric is set (by the tsDyn package) to four values: .5, 1, 1.5 and 2 times the standard deviation of the series. m is called the embedding dimension.[38] There are

[35]The RATS programs are provided by Enders in his Instructor's Resource Guide. We employed the time series programs in the core R package as well as another package, tsDyn, version 0.7. This version of tsDyn appeared in 2008. Apparently it has not been updated.

[36]GPS argue the MA component of their estimated model capture measurement error (1998: fn. 6, p. 886). They make no mention of the fit of the alternative AR(2) plus constant model.

[37]As Enders (2010: 436) explains, one first estimates a linear model and obtains the fitted values of the variable of interest, say, $\hat{y}_t$. Then the following equation is estimated:

$$e_t = \delta z_t + \sum_{h=2}^{H} \alpha_h \hat{y}_t^h \qquad for \qquad H \geq 2 \tag{27}$$

where e_t represents the estimated residuals, z_t is the vector of explanatory variables in the ARIMA model including the constant. Again, the RESET,distributed F, assesses the joint statistical significance of the α_h's.

[38]This terminology, as described in tsDyn Version 7 (p. 5), is based on the following formal representation of a discrete time univariate stochastic process, $[X_t]_{t \in T}$. The "map" for this process is written:

$$X_{t+s} = f(X_t, X_{t-d}, \ldots, X_{t-(m-1)d}; \theta) + \epsilon_{t+s} \tag{28}$$

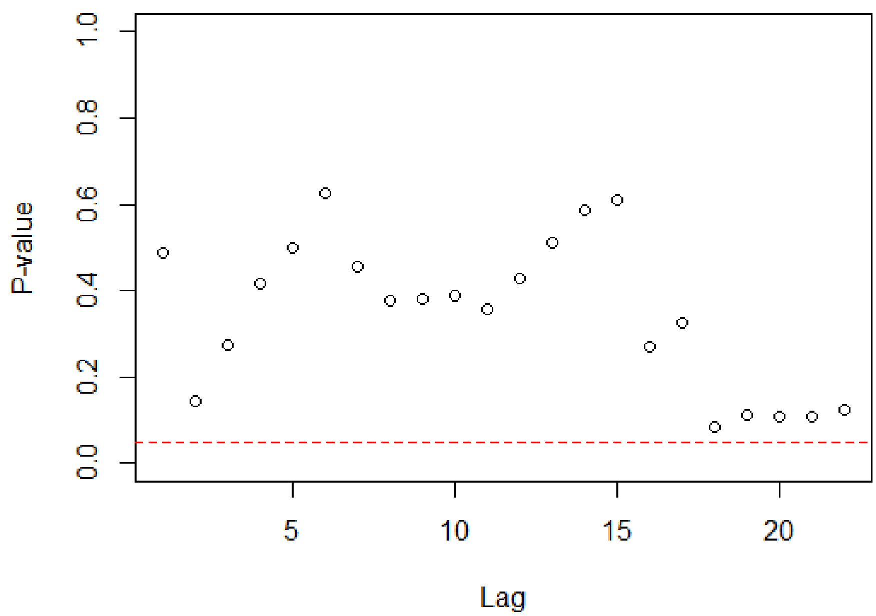

Figure 4: McLeod-Li Test P values for Different Lags for (1,0,1)+Constant Model of M_t

tests for independence and linearity. Diks and Manzan (2002: 3) apply these tests to the original (level) time series. They argue that "the advantage over testing for dependence in residuals is that [by using the raw data] the lag dependence in the time series is preserved"[39] However, Granger and Teräsvirta (1993: 91) say the test should be applied to the residuals from the best fitting linear model. If the null hypothesis is rejected "one can conclude that nonlinearity is present but its form is not determined. It can be chaos or a nonlinear stochastic process." Enders notes that rejection of the null based on the BDS test indicates various types of misspecification including but not necessarily implying nonlinearity (Enders 2010: 437).[40] The small sample performance of the BDS test is not good. Bootstrapped

where $[\epsilon_t]_{t \in T}$ is white noise and also independent of X_{t+s}, and f is a generic function from R^M to R. These models are abbreviated NLAR(m) which denotes Nonlinear AutoRegressive models of order m. The parameters m, d, s and θ are the embedding dimension, time delay, and forecasting steps, and coefficients on the lag terms of the model, respectively.

[39]Diks and Manzan (2002) develop information theoretic tests for independence and linearity based on the idea of conditional mutual information (intermediate lag values of the series).

[40] Granger and Teräsvirta (1993: 36) explain how the test has power against white noise chaotic processes as well as against a variety of nonlinear stochastic processes. They (1993: 90-91) provide a full description of the BDS test and its value in testing for chaotic dynamics. What follows is a condensation of their description. Let $X_{t,m}$ denote a set of consecutive terms from a series x_t such that $X_{t,m} = (x_t, x_{t+1}, \dots x_{t+m-1})$. A pair of vectors, $X_{t,m}$ and $X_{s,m}$, are said to be ϵ apart if the following relationship holds for each of pair of the corresponding terms:

$$|x_{t+j} - x_{s+j}| \leq \epsilon, \qquad j = 0, 1, \dots, m-1. \tag{29}$$

The correlation integral, $C_m(\epsilon)$ is the limit of T^{-2} times the number of pairs (s,t) that are close in the sense

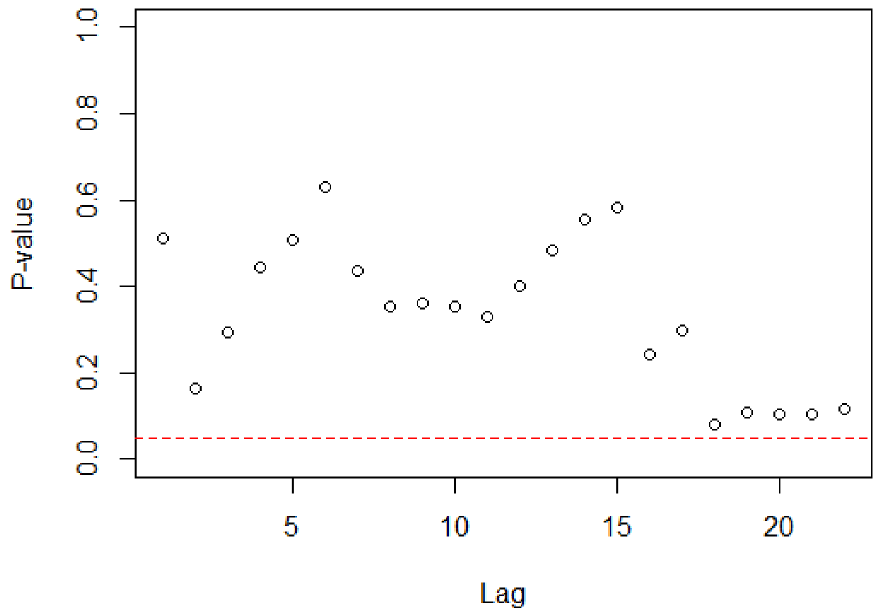

Figure 5: McLeod-Li Test P values for Different Lags for (2,0,0)+Constant Model of M_t

confidence intervals are recommended (*Ibid.*: 91, 102; Enders 2010: 437). The delta test produces bootstrapped based, one sided p-values (Manzan 2003; Diks and Manzan 2002).

Delta tests for independence and for linearity were performed on both the raw macropartisanship data and on the residuals from our two linear models. For reasons that are not clear, the tsDyn program returned an error message for the linearity tests for the residuals from the linear models.[41] Briefly, the results are mixed. The p-values for some distances (ϵ) indicate rejection of the nulls of independence and of linearity for the (raw) level series for M_t and rejection of the null of independence for the residuals of both linear models.

Enders (2010: 449-450) describes one additional test. This is the test of the null of a simple linear model against an alternative, nonlinear SETAR model of the same structure. It

(see *Ibid.* Section 3.3.1). Then the BDS statistic is:

$$S(m, \epsilon) = \hat{C_m}(\epsilon) - [\hat{C_1}(\epsilon)]^m \tag{30}$$

for some choice of m and ϵ Under the null that x_t is i.i.d. $\sqrt{T S(m, \epsilon)}$ has a normal distribution with mean zero and a variance that is a function of m and ϵ.

[41]Note that the tsDyn R package describes the delta test routine as experimental. The illustrations in the package are for delta independence and linearity tests for raw data not residuals from linear models. The illustration in the package reports rejection of the null for independence of the well known lynx data but *not* for the null of linearity of these data (*Ibid.* p. 16). At this point, the authors of tsDyn admit the results are anomalous and stress the delta test routines are experimental. Again, we can find no updated version of tsDyn since 2008.

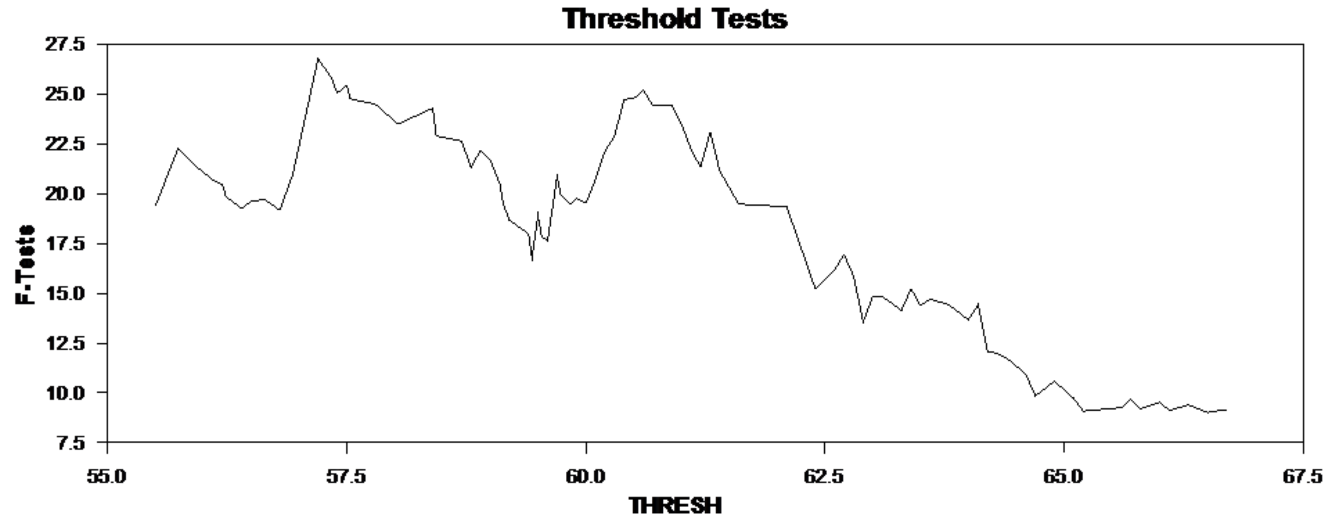

Figure 6: Maximum F Statistic for TAR(2) Model of Macropartisanship

uses Hansen's supremum F test. We implemented this test for our AR(2) model of macropartisanship. The result suggest nonlinearity. The RATS program indicate a maximum F value for the SETAR(2) model of 26.77 at a threshold of 57.2 (Figure 6). The bootstrap p value based on 5000 replications for this F statistic is .0000. This suggests that macropartisanship is governed by a nonlinear process in the period on which GPS and EMS focused.

Of course, there is no reason to assume that the SETAR model is of the AR(2) form. Therefore, using the selectSETAR routine in tsDyn we evaluated a variety of possible TAR models, 4032 to be exact.[42] On the basis of the routine's pooled-AIC criterion the best fitting model for macropartisanship was found to be:

$$M_{t+1} = \begin{cases} 1.53 + .82M_t + .07M_{t-1} + .25M_{t-2} - .18M_{t-3} & \text{if } M_{t-2} > 55.34 \\ (2.42)(.08) \quad (0.10) \qquad (0.10) \qquad\qquad (08) & \\ 22.59 + .36M_t + .67M_{t-1} - .67M_{t-2} + .23M_{t-3} & \text{if } M_{t-2} \leq 55.34 \\ (17.51)(0.21) \quad (.23) \qquad (0.30) \qquad (0.18) & \end{cases}$$

where the standard errors are in parentheses below their respective fitted coefficients. The fit statistics for this model are: residual variance 2.86, AIC 207, and MAPE 2.221%. Figure 7 is its regime switching plot.[43] As expected the plot shows that the low regime occurred more frequently in the late 1990s. More surprising are the findings that it is the value of macropartisanship lagged two quarters that precipitates the regime switch and the

[42]We set the forecast steps and regular delay parameters for the model to the defaults of 1. We then explored the fit of 3 threshold delays (1,2,3), alternative (independent) lag structures for the high and low regimes (each ranging between 1 and 4), and 84 possible thresholds corresponding the the M levels remaining after trimming the lowest and highest 15% of the values. This produced 3x4x4x84=4032 models.

[43]tsDyn appears to treat the first four observations as initial conditions and hence it does not include them in the regime switching plot. Geoff Sheagley produced Figure 7 which includes these observations.

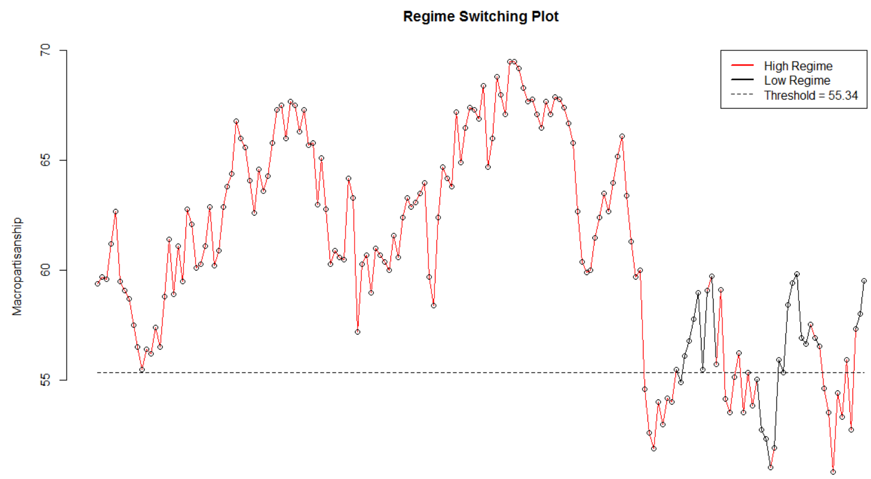

Figure 7: Regime Switching Plot for SETAR Model for Gallup based Measure of Macropartisanship

threshold has value that represents a Democratic advantage not an even split in the the partisan leanings of the electorate. As discussed in section 3.2 above, in best fitting SETAR model each regime exhibits a different kind of sequence outcome dependence; the limiting behavior of each alone approaches a different expected value. Like Figure 2, globally the stationary density of this nonlinear macropartisanship process has two humps not a single peak.[44]

5 Conclusion

Conventional time series methods give us tools to identify and analyze data generating processes that embody most of the key concepts associated with the idea of path dependency. We simply need to be clear about the nature of each model, how (if) each model embodies the impact of initial condition, the set or sequence of shocks that a data generating process experiences, and multiple equilibria. As I have shown, doing this illuminates new and potentially useful ideas about the nature of American macropolitical dynamics. It also suggests the need for tests for nonlinearity in macropartisanship and in (macropartisanship's relationship to) other theoretically important series. [45]

[44]These long term expected values, calculated from the point estimates for the coefficients at their sixth decimal places, are 56.46 and 54.83. As regards the tests for unit roots in the context of a TAR model for Macropartisanship, see Appendix 7.3

[45]The deeper challenge is to develop theories that predict mulitiple, moving equilibria in American political dynamics. See, for example, Mebane (2000).

Model	Impact of Initial Condition(s) On Outcomes	Phat Outcome Dependence	Sequence Outcome Dependence	Limiting Behavior
Univariate, Linear:				
Stationary AR	yes	no	yes	For Gaussian errors, ergodic in moments; stochastically stable
Nonstationary (Random Walk)	yes	yes	no	Fixed expected value, changing distribution in time
Multivariate Linear Models:				
Single Equation Time Series Regression (stationary errors)	no	no	yes	Depends on forcing function
Singe Equation ECM	yes	yes	yes	Fixed expected value repeated moves into/out of “equilibrium”
Stationary VAR	no	no	yes	Fixed vector of expected values
VECM (Less than full rank)	yes	yes	yes	Fixed vector of expected values; repeated moves into/out of ”equilibrium”
Nonlinear Time Series				
Threshold Autoregressive(TAR)				
Simple TAR	no	no	yes	Under certain conditions, geometrically ergodic
Basic TAR	no	no	yes	Under certain conditions, geometrically ergodic
Band TAR	yes	yes	yes	In part, changing distribution in time
Nonlinear ECM	yes	yes	yes	Fixed vector of expected values; switches between rates of error correction movements into/out of “equilibrium”
NLS Regression	no	no	yes	Depends on forcing functions

Table 1: Time Series Models, Outcome Dependency and Related Concepts.

	F-Test Statistic	(DF1,DF2)	p-value
(1,0,1)+Constant			
h=2,3	0.52	(3,170)	0.67
(2,0,0)+Constant			
h=2	0.2153	(2, 173)	.7624
h=2,3	0.4544	(3,172)	.7038

Table 2: RETEST Results for Nonlinearity of GPS Gallup-based Measure of Macropartisanship. Residuals from (1,0,1)+Constant and (2.,0,0)+Constant Models. h denotes powers to which residuals are raised in the tests.

Epsilon	2.344	4.687	7.031	9.374
Independence				
m=2	0.06	0.02	0.02	0.02
m=3	0.24	0.02	0.02	0.02
Linearity				
m=2	0.06	0.16	0.12	0.26
m=3	0.06	0.06	0.26	0.02

Table 3: Delta Test Results (p values) for Independence and for Nonlinearity of GPS Gallup-based Measure of Macropartisanship. Raw data. Epsilon is distance as calculated by formula in the text. M again is the embedding dimension for the intermediate values of the series.

(1,0,1)+Constant				
Epsilon	0.9028	1.8057	2.7085	3.36113
m=2	0.022	0.023	0.072	0.032
m=3	0.600	0.74	0.102	0.172
(2,0,0)+Constant				
Epsilon	0.9025	1.8049	2.7074	3.6098
m=2	0.040	0.060	0.090	0.060
m=3	0.58	0.12	0.10	0.24

Table 4: Delta Test Results for Independence of Residuals from Two Linear Models of GPS Gallup-based Measure of Macropartisanship. Entries are p values for respective delta statistics.

6 References

Balke, Nathan S. and Thomas B. Fomby. "Threshold Cointegration." 1997 *International Economic Review* 38: 627-643.

Bednar, Jenna, Scott E. Page, and Jameson Toole. 2012 "Revised Path Dependence." *Political Analysis* 20(2): 146-156.

Bennett, Andrew and Colin Elman. 2006. "Complex Causal Relationships and Causal Methods: The Example of Path Dependence." *Political Analysis* 14: 250-267.

Box-Steffensmeier, Janet and Renee Smith. 1996. "The Dynamics of Aggregate Partisanship." *American Political Science Review*90(3): 567-580.

Brandt, Patrick T. and John R. Freeman. 2006 "Advances in Bayesian Time Series Modeling and the Study of Politics: Theory Testing, Forecasting and Policy Analysis," *Political Analysis* 14(1): 1-36.

Brandt, Patrick T. and John R. Freeman. 2009 "Modeling Macropolitical Dynamics" *Political Analysis* 17(2):113-142.

Brock, William, Davis Dechert, Jose Scheinkman, and Blake LeBaron. 1996. "A Test for Independence Based Upon the Correlation Dimension." *Econometric Reviews* 15: 197-235.

Chan, K.S., J.D. Petruccelli, H. Tong and S.W. Woolford. 1985 "A Multiple Threshold AR(1) Model." *Journal of Applied Probability* 22: 267-279.

Diks, Cees and Sebastiano Manzan. 2002. "Tests for Serial Independence and Linearity Based on Correlation Integrals." *Studies in Nonlinear Dynamics and Econometrics* 6(2).

Enders, Walter 2010. *Applied Econometric Time Series* Third Edition. John Wiley And Sons.

Ender, Walter, and C.W. J. Granger. 1998. "Unit-root Tests and Asymmetric Adjustment with an Example Using the Term Structure of Interest Rates." *Journal of Business and Economic Statistics* 16: 304-311.

Erikson, Robert, Michael MacKuen, and James Stimson. 1998 "What Moves Macropartisanship? A Reply to Green, Palmquist, and Schickler" *American Political Science Review* 92(4): 901-912

Fiorina, Morris. 1981 *Retrospective Voting in American National Elections* New Haven, CT. Yale University Press.

Franzese, Robert Jr., Jude Hayes, and Aya Kachi. 2012 "Modeling History Dependence in Network-Behavior Coevolution" *Political Analysis* 29(2): 175-190.

Freeman, John, John Williams, and Tse-min Lin 1989. "Vector Autogression and the Study of Politics" *American Journal of Political Science* 33(4): 842-877.

Gill, Jeff. 2011. "Introduction to the Virtual Issue: Bayesian Methods in Political Science." Website.

Goldstone, Jack A. 1998 "Initial Conditions, General Laws, Path Dependence and Explanation in Historical Sociology" *American Journal of Sociology* 104(3): 829-845.

Granger, C.W. J. and T. Teräsvirta. 1993 *Modelling Nonlinear Economic Relationships* NY: Oxford University Press.

Green, Donald, Bradley Palmquist and Eric Schickler.1998 "Macropartisanship:Replication and Critique" *American Political Science Review* 92(4): 883-900.

Hamilton, James D. 1994 *Time Series Analysis* Princeton, N.J. Princeton University Press.

Hansen, Bruce E. "Inference in TAR Models." 1997 *Studies in Nonlinear Dynamics and Econometrics* 2(1): 1-14.

Jackman, Simon. 1987 "Public Opinion Over the Long Run: Presidential Approval, the Macroeconomy, and Rallies." Paper Presented at the Annual Meeting of the APSA, Washington, D.C.

Jackson, John and Kenneth Kollman. 2010. "A Formulation of Path Dependence With An Empirical Example." *Quarterly Journal of Political Science* 5(3): 257-289.

Jackson, John and Kenneth Kollman. 2011. "Connecting Micro- and Macropartisanship." *Political Analysis* 19(4): 503-518.

Jackson, John and Kenneth Kollman. 2012. "Modeling, Measuring, and Distinguishing Path Dependence, Outcome Dependence, and Outcome Independence." *Political Analysis* 20(2): 157-174.

Keele, Luke and Nathan J. Kelly 2005 "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables" *Political Analysis* 23:186-205.

Lasota, A. and M.C. Mackey. 1989. "Stochastic Perturbation of Dynamical Systems: The Weak Convergence of Measures." *Journal of Mathematical Analysis and Application* 138: 232-248.

Lütkepohl, Helmut. 1993. *Introduction to Multiple Time Series Analysis* Berlin: Springer-Verlag.

MacKuen, Michael B., Robert S. Erikson, and James A. Stimson. 1989 "Macropartisanship" *American Political Science Review* 83: 1125-1142.

Mahoney, James. 2000. "Path Dependence in Historical Sociology" *Theory and Society* 29:507-548.

Manzan, Sabastiano. 2003. "Essays in Nonlinear Economic Dynamics." Thela Thesis.

Mebane, Walter R. Jr. 2000. "Congressional Campaign Contributions, District Service, and Electoral Outcomes in the United States: Statistical Tests of a Formal Game Theoretic Model with Nonlinear Dynamics." In *Political Complexity: Nonlinear Dynamics of Politics* Diana Richards Editor. Ann Arbor: University of Michigan Press.

Ostrom, Charles W. Jr. and Renee Smith 1993 "Error Correction, Attitude Persistence, and Executive Rewards and Punishments: A Behavioral Theory of Presidential Approval." In *Political Analysis* Vol. 4 John R. Freeman Editor, Ann Arbor: University of Michigan Press.

Page, Scott. 2006. "Path Dependence" *Quarterly Journal of Political Science* 1:87-105.

Pippinger, Michael K. and Gregory E. Goering. 1993. "A Note on the Empirical Power of Unit Root Tests under Threshold Processes." *Oxford Bulletin of Economics and Statistics* 55(4): 473-481.

Quinn, B.G. 1982. "Stationarity and Invertibility of Simple Bilinear Models" *Stochastic Processes and Their Applications* 12: 225-230

Tong, Howell. 1990. *Non-Linear Time Series: A Dynamical Systems Approach* NY: Oxford University Press.

Vergne, Jean-Philippe and Rodolphe Durand. 2010. "The Missing Link Between Theory and Empirics of Path Dependence: Conceptual Clarification, Testability Issues, and Methodological Implications." *Journal of Management Studies* 47(4): 736-759.

7 Appendix

Exploratory Analysis of GPS Macropartisanship(Gallup) Series

Figure 4 is the histogram of the data. Clearly this histogram is inconsistent with the idea that the series is unimodal.[46]

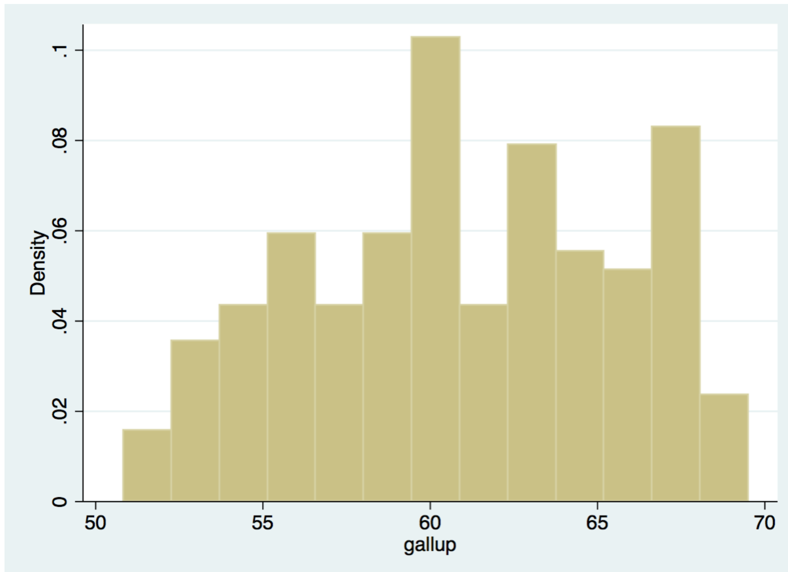

Figure 8: Histogram for Macropartisanship Data from GPS.

Figures 5-8 are the nonparametric regression lines for what the R package tsDyn calls "Autopairs graphs." Curved lines–a hump, for example– suggests that a linear model may be inappropriate (see Tong 1990: 5.2.4, 7.2.3). Again, there seems to be little evidence of such curves in the raw M_t series.[47]

7.1 Replication and fit of linear ARMA models for Gallup based estimates of U.S. Macropartisanship, 1953:2-1996:4.

Table X reports the results of the replication of GPS's results are reported in the first Table of their article (1998: 887). It shows that the RATS results produce almost identical estimates; the R results are close but not as accurate as the RATS estimates. The constant in all this output is the long-run expected value of the series as predicted by the models, not the actual constant in the linear ARMA functional forms.

In addition, using R, we fit a series of simple, linear AR models to the data. These results are in Table XX. They show at that a AR(2) model with a constant also fits the data well. In fact, according to R, the AR(2) model with a constant has a lower AIC value than the (1,0,1) model with a constant.

[46]See Tong 1990: for a discussion for statistical tests for unimodality.

[47]The nonparametric regression lines are drawn with a function called "sm.regression" from the R library.

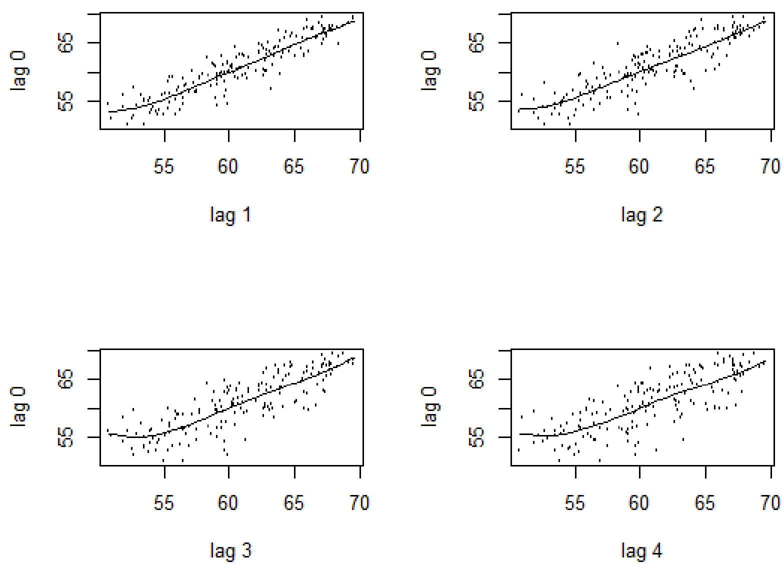

Figure 9: Autopairs Plot, Lag 1-4

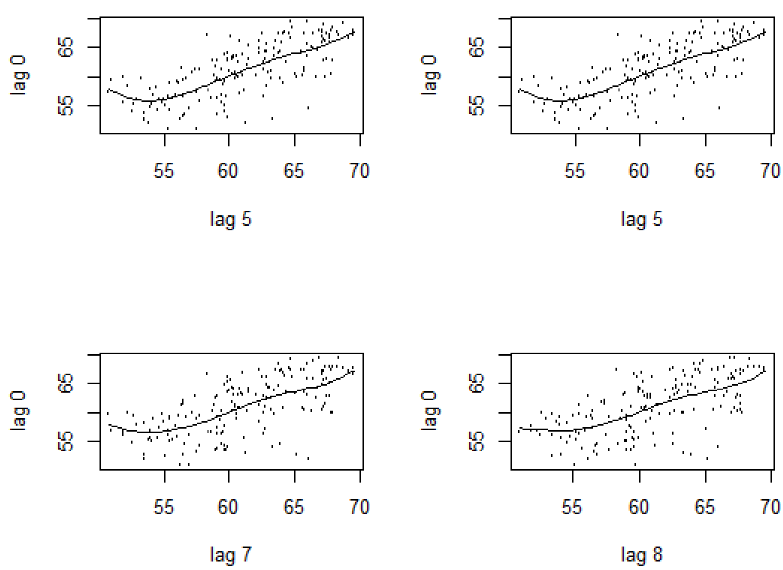

Figure 10: Autopairs Plot, Lag 5-8

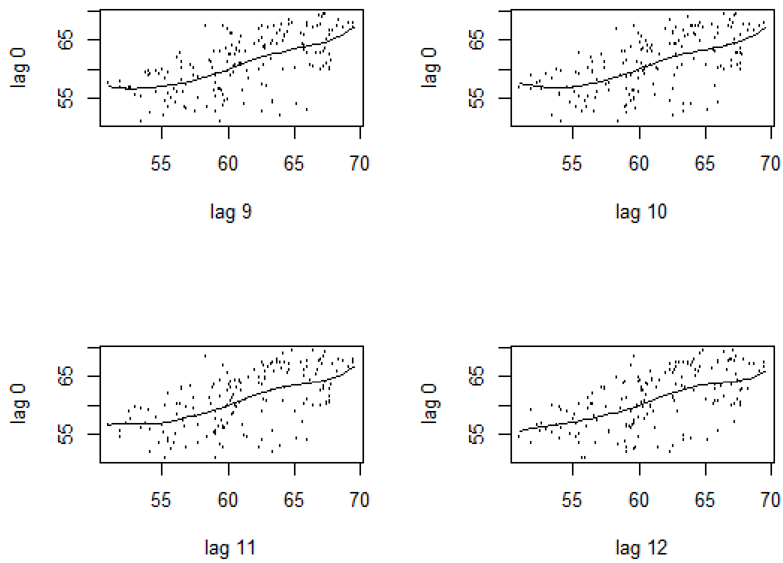

Figure 11: Autopairs Plot, Lag 9-12

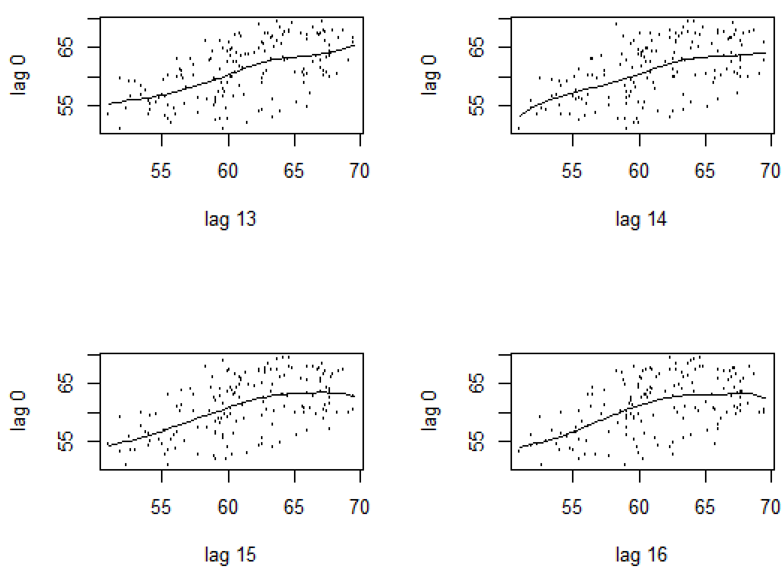

Figure 12: Autopairs Plot, Lag 13-16

	GPS Results	RATS Output	R Output
Constant	60.800	60.79	60.71
S.E.	(2.181)	(2.178)	(1.80)
AR(1) Coefficient	.949	.949	.94
S.E.	(.026)	(0.26)	(.03)
MA(1) Coefficient	-.189	-.187	-0.19
S.E.	(.081)	(.081)	(.08)
Adjt.R^2	.850	.850	
S.E. of Estimate	1.821	1.821	
N	175	175	175
Q(36-2)		44.129	44.21
Sig.Level		.115	.113
AIC			716.33

Table 5: Replication of Green, Palmquist and Schickler Estimates for ARMA (1,0,1) model with constant for Gallup based Measure of Macropartisanship. S.E. denotes standard error.

	AR(1)	AR(2)	AR(3)
Constant	60.80	60.73	60.72
S.E.	(1.54)	(1.78)	(1.79)
AR(1) Coefficient	0.91	.75	.75
S.E.	(0.03)	(0.07)	(.08)
AR(2) Coefficient		.18	.16
S.E.		(.07)	(.09)
AR(3) Coefficient			.02
S.E.			(.08)
Q(36-2)	47.51	44.32	44.18
Sig.Level	.077	.111	.093
AIC	719.97	716.19	718.09

Table 6: Estimates for Three AR Models (with Constants) for Green, Palmquist and Schickler Gallup based Measure of Macropartisanship. S.E. denotes standard error.

7.2 The Issue of Stationarity of the GPS Measure of Macropartisanship

As noted in the text, the Dickey Fuller test assumes a linear, symmetric adjustment in a time series process. If the actual data generating process (DGP) is nonlinear, the Dickey Fuller test can produce mistaken inferences (Enders 2010: 477, Pippinger and Goering 1993). Table 7 reports the results of the augmented version of this test for zero and four lags for the GPS M_t series. Only one estimate exceeds the critical value and that is at the .10 level. So, like GPS (1998:887), our results indicate that M_t is nonstationarity.[48]

[48]Green et al. (1998: 887) report an augmented DF test statistic for four lags of -2.46 so we assume that they had an intercept term only in their specification. They also report that their Phillips-Perron test statistic exceeds the critical value but only at the .10 level.

	No Intercept Or Trend Term	Intercept Term Only	Intercept and Trend Term
ADF Test Statistic (zero lags)	-.024	-2.67	-2.94
ADF Test Statistic (four lags)	-.15	-2.45	-2.79
Critical Values			
1 per cent	-2.58	-3.46	-3.99
5 per cent	-1.95	-2.88	-3.43
10 per cent	-1.62	-2.57	-3.13

Table 7: Results for Two Specifications of the Augmented Dickey Fuller Tests for GPS Measure of Macropartisanship

Enders (2010: Section 5.11) presents a test for unit roots under the alternative that the DGP is a simple TAR model. His analysis is an extension of work originally published in Enders and Granger (1993). Unfortunately, his test assumes a single variance for the time series of interest. Hence it is *not* applicable to the model we found with tsDyn. But, Enders' test is nonetheless informative. The TAR model used for Enders' test is:

$$\Delta M_t = I_t \rho_1 (M_{t-1} - \tau) + (1 - I_t)\rho_2 (M_{t-1} - \tau) + \epsilon \tag{31}$$

where τ is the threshold and I_t is the indicator function here defined as:

$$I_t = \begin{cases} 1 & \text{if } M_{t-1} \geq \tau \\ 0 & \text{if } M_{t-1} < \tau. \end{cases}$$

Once more, this is not the model we found using tsDyn for M_t because it assumes, among other things, a single variance for macropartisanship. Now, if $\rho_1 = \rho_2 = 0$ the process would be a random walk; if we reject this restriction, we infer there is an "attractor" for M_t (it is stationary as long as $-2 < \rho_1, \rho_2 < 0$. The F statistic can be used to test for such an attractor but the critical values for this particular test are nonstandard. Enders supplies a Table (2010: 494) for critical values for the test of $\rho_1 = \rho_2 = 0$ for this TAR model. Should we reject the null hypothesis, we can proceed to test for asymmetric adjustment, $\rho_1 = \rho_2$. The standard F statistic and critical values can be used for this second restriction (these critical values are an approximation of the actual critical values which can be generated by Hansen bootstrap method).[49]

We implemented the test in Enders (2010) using the RATS code supplied with his book. The estimated model for the GPS M_t series is:

[49]Enders and Granger(1993) analyze somewhat different TAR models. They also advocate a four step procedure which differs from the procedure outlined in Enders (2010: 479). In particular, in the latter one *starts* by finding the threshold of the TAR model rather than searching for this threshold *after* testing that $\rho_1 = \rho_2 = 0$. Table G in Enders (2010) does not appear in Enders and Granger (1993) apparently because the model used for the test in Enders (2010) is different from those used in the earlier article.

$$\Delta M_t = \underset{(-1.25)}{-.023} I_t(M_{t-1} - 55.34) - \underset{(-2.14)}{.360}(1 - I_t)(M_{t-1} - 55.34) - \underset{(-2.27)}{1.71}\Delta M_{t-1} + \epsilon$$

where the ΔM_{t-1} term on the right side of the equation, as in Enders subsequent illustration, is included to account for any serial correlation in the errors of the equation and the numbers under the coefficients are t statistics.[50] The estimated I_t, the indicator function is:

$$I_t = \begin{cases} 1 & \text{if } M_{t-1} \geq 55.34 \\ 0 & \text{if } M_{t-1} < 55.34 \end{cases}$$

The test for $\rho_1 = \rho_2 = 0$ yields a F(2, 171) of 2.96 which does not exceed the respective critical value in Enders' table G. This indicates that, according to the alternative of a simple TAR model with a single variance, the hypothesis of a unit root cannot be rejected.[51]

Thus we are left with the implausible conclusion that M_t has infinite variance (in sample). If M_t is nonstationary, we would like to know how (if) this fact undermines the inferences we drew from our McLeod-Li, RESET, BDS and other tests for nonlinearity. Unfortunately Enders does not address this issue; he focuses only on the shortcomings of the Dickey Fuller unit root test. Future research will be devoted to finding a unit root test that has a TAR model with multiple variances as the alternative and to understanding the implications of nonstationarity for nonlinearity testing.

[50] The Durbin Watson statistic for the fitted TAR model is 2.0

[51] The F statistic for the equality of the coefficients in the TAR model here is F(1,171) = 3.98. This value has an exact statistical significance of .048 although it is only an approximation of the level of statistical significance.