A Linear Poisson Autoregressive Model: The Poisson AR(p) Model

Patrick T. Brandt and John T. Williams
Department of Political Science and Workshop in Political Theory and Policy Analysis, Woodburn Hall 210, Indiana University, Bloomington, Indiana 47405
e-mail: pbrandt@indiana.edu
WWW: http://www.polsci.indiana.edu/jotwilli/home

Time series of event counts are common in political science and other social science applications. Presently, there are few satisfactory methods for identifying the dynamics in such data and accounting for the dynamic processes in event counts regression. We address this issue by building on earlier work for persistent event counts in the Poisson exponentially weighted moving-average model (PEWMA) of Brandt et al. (*American Journal of Political Science* 44(4):823–843, 2000). We develop an alternative model for stationary mean reverting data, the Poisson autoregressive model of order p, or PAR(p) model. Issues of identification and model selection are also considered. We then evaluate the properties of this model and present both Monte Carlo evidence and applications to illustrate.

1 Introduction

POLITICAL METHODOLOGISTS AND empirical researchers employ a wide variety of tools for analyzing discontinuous, non-normally distributed data. While these methods (built around the maximum-likelihood principle) have received wide acceptance in political science, few of these models have been developed to address time series data. This fact necessitates developing time series models for nonnormal data that may have application in political science.[1]

In an earlier paper, Brandt et al. (2000) develop a time series model for persistent event count time series. Persistent time series processes are identified by sample autocorrelation functions that have significant memory. In such data, the effects of shocks persist over many periods. Their Poisson exponentially weighted moving average model (PEWMA) captures the dynamics by estimating a time-dependent discounted average of the mean of the event

Authors' Note: Please address all correspondence to the first author. An earlier version of this work was presented at the 1998 Summer Meeting of the Society for Political Methodology, University of California, San Diego. Williams's research was supported by National Science Foundation Grant SBR-9422645. All GAUSS code, software to implement the models, and data are available with documentation at the *Political Analysis* website. The authors wish to thank Chris Achen, Ben Fordham, John Freeman, Simon Jackman, Gary King, Michael McGinnis, Walter Mebane, Stephanie Sillay, and Jim Stimson for their comments.

[1] A notable exception to this trend is work using event-history methods, as described by Box-Steffensmeier and Jones (1997).

Copyright 2001 by the Society for Political Methodology

count process. The PEWMA state–space model fits a local level mean to the event series to account for persistent dynamic processes in the data.

While the PEWMA model has significant advantages over standard event count models for persistent time series of counts, it is not suitable for cyclical and short-memoried processes that are mean reverting. In the PEWMA model the mean number of events evolves over time according to an exponentially weighted moving average. If the process that generates the event counts is mean reverting, then such a model provides a poor approximation to the data and will be inefficient.

From what we know of standard time series analysis and linear models, failing to account for some pattern of systematic dynamics leads to incorrect inferences. For Gaussian data, inferences based on static models when the data are generated by a dynamic time series process suffer from inefficiency. Such problems are also present in nonnormal data. For example, if one estimates a Poisson regression model and the process that generates the event counts is time dependent, then estimates will be inefficient (Brandt et al. 2000).

The implications of dynamic misspecification in maximum-likelihood estimators are well understood in theory. White (1994, pp. 52–59) discusses the implications of dynamic misspecification of the conditional mean in such estimators. Unless the dynamic variables omitted from the model do not Granger cause the conditional mean, the model will be misspecified. For many time series event count models, the number of events observed at time t depends on the number of events observed at some previous time period, say $t - 1$. If this is the case, failing to account for this temporal effect leads to a form of conditional mean misspecification that results in possible inefficiency.

Here we address the dynamic specification of event count time series that are mean reverting and cyclical. In Section 2 we describe the specification, estimation, and interpretation of a Poisson autoregressive model—the PAR(p)—that can be used to model stationary, mean reverting event count processes. We discuss the selection of event count time series models in Section 3. Section 4 presents Monte Carlo evidence that demonstrates the small sample properties of the PAR(p). Here we provide a gauge for assessing the effects of dynamic misspecification in dynamic event count data. We then present two examples to show the application of the PAR(p) model. In Section 5 we present an application using a classic data set from McCleary and Hay (1980). We then present a time series analysis of presidential vetoes based on Lewis and Strine (1996). We show that robust dynamic specification of event count time series models requires the same tools used in Gaussian time series analysis. Section 7 concludes by outlining a strategy for model identification and specification for event count time series data.

2 An Autoregressive Model for Event Count Data

2.1 *Previous Efforts*

A typical approach for dealing with the contagion or correlation among events in event count data is to estimate a negative binomial model or generalized event count model (GEC) (King 1989). Often, if the data are a time series, a lagged dependent variable is included.[2] An alternative approach is to ignore that the data follow an event count distribution and transform the data so that they are suitable for Gaussian ARIMA techniques.[3] Both of these

[2]This method is used by Pollins (1996), Senese (1997), and Spriggs and Wahlbeck (1995).

[3]Gaussian ARIMA techniques have been used by Brophy-Baermann and Conybeare (1994), Mansfield (1992), O'Brien (1996), Sayrs (1996), and Spriggs and Wahlbeck (1995) to account for dynamics in event count data.

approaches fail to model adequately the dynamics in the data or the distribution that gives rise to the event counts.

If an event count regression model includes a lagged dependent variable, the exponentiated coefficient on the lagged variable is no longer an autocorrelation coefficient (as it would be in the Gaussian model). Brandt et al. (2000, pp. 824–825) document why this is the case. They show that including a lagged count in the exponential function of an event count model estimates a linear exponential growth rate. This is a dynamic model with a trend, but not necessarily a cyclical or dynamic component.

Another approach to event count time series is to ignore that the event count data come from a discrete distribution and model the data as though it were from a Gaussian distribution. The Gaussian approximation is valid only for "large" event counts, where the Poisson distribution approaches normality. King (1988), shows for nondynamic event count data, that assuming normality leads to biased estimates of the regression parameters. The reason the Gaussian model should be avoided is that it ignores the basic distributional nature of discrete data.

As alternatives to these approaches, Brandt et al. (2000) review different time series models for event count data and focus on a model for nonstationary event counts.[4] They present a Poisson exponentially weighted moving average model (PEWMA) with a dynamic process that is a random walk with noise model. Such a model captures nonstationary patterns in event count time series data but is not appropriate for stationary or mean reverting data. To deal with such event count time series data, we present a model based on a simple autoregressive process. This autoregressive model, like the PEWMA model, is based on a state–space time series representation.

2.2 *An Autoregressive Model for Count Data: PAR(*p*)*

Brandt et al. (2000) develop the PEWMA state–space model based on a measurement equation, a transition equation, and a conjugate prior.[5] The assumptions of the model are that the observed or measured counts at time t are generated from a Poisson distribution with mean μ_t. The mean μ_t evolves according to a gamma-distributed transition equation. These PEWMA assumptions can be summarized by the following equations and densities.

$$\begin{aligned}
\text{Measurement equation:} \quad & \Pr(y_t \mid \mu_t) = \frac{\mu_t^{y_t} e^{-\mu_t}}{y_t!} \\
\text{Transition equation:} \quad & \mu_t = e^{r_t} \mu_{t-1} \eta_t, \qquad \text{where} \qquad \eta_t \sim \beta\left(\omega a_{t-1}, (1-\omega)\, a_{t-1}\right) \\
\text{Conjugate prior:} \quad & \Pr(\mu_{t-1}; a_{t-1}, b_{t-1}) = \frac{e^{-b_{t-1}\mu_{t-1}} \mu_{t-1}^{a_{t-1}-1} b_{t-1}^{a_{t-1}}}{\Gamma(a_{t-1})}.
\end{aligned}$$

The parameter ω captures discounting in the conditional mean function for the event counts. The term η_t is the beta-distributed random component of the state equation. Brandt et al. (2000) present the extended Kalman filter, the predictive distribution, and the log-likelihood function that characterize the model.

[4] In this earlier article, a number of time series models are discussed. Interested readers should refer to that article or Cameron and Trivedi (1998) for a review of time series models for event count data.

[5] A conjugate prior for a given distribution is that distribution that will produce a distribution in the same class of distributions when used to form a Bayesian update of the orginal distribution. See DeGroot (1970) and Gelman et al. (1995) for discussions.

An autoregressive model for event counts is constructed based on an approach similar to the PEWMA. Instead of employing the multiplicative transition equation used in the PEWMA model, we replace the transition with a linear autoregressive process. To develop this model, we first present some notation and outline what we mean by "linear autoregressive processes of order p." We then define a transition equation and set of filter equations to characterize an AR(p) model for event count data. Next we develop an extended Kalman filter for the PAR(p) and its likelihood function. In what follows, we denote the matrix of conditioning data Y_{t-1}. This includes all the observed values of the dependent and independent variables and any parameter estimates at time t. The vector of conditioning data is then $Y_{t-1} = (y_0, y_1, \ldots y_{t-1}; X_0, X_1, \ldots X_{t-1})$, where y_{t-k} are the past counts and X_{t-k} are the past covariates.

Grunwald et al. (1997a, b) discuss modeling an event count series using a linear first-order AR process. We generalize the first-order linear AR process they describe to an AR(p) process. The definition of a linear AR(p) process does not depend on the assumption of a normal distribution. To see this, let Y_{t-1} be all the prior information about the series of interest at time t. Assume that y_t is realization from a (time-homogeneous) Markov process with the conditional transition probability $\Pr(y_t \mid Y_{t-1})$ and that $E[Y_0] = \mu < \infty$. Let the conditional expectation $E[y_t \mid Y_{t-1}] = m_t$ at time t have a finite mean. Then y_t is a pth-order linear autoregressive process if

$$E[y_t \mid Y_{t-1}] = \sum_{i=1}^{p} \rho_i Y_{t-i} + \lambda \tag{1}$$

where ρ_i and λ are any real numbers. This specification of a linear AR(p) process places no restriction on the density $\Pr(y_t \mid Y_{t-1})$. The choice of this density for y_t places constraints on the admissible values of λ and ρ.

Equation (1) generates a mean stationary time series model. Grunwald et al. (1997b, Proposition 1), using iterated expectations, show that this generates a mean stationary process for an AR(1) process. We can generalize this result by finding iterated expectations for the AR(p) process in Eq. (1):

$$E[E[y_t \mid Y_{t-1}]] = E\left[\sum_{i=1}^{p} \rho_i Y_{t-i} + \lambda\right] \tag{2}$$

$$E[Y_t] = \sum_{i=1}^{p} \rho_i E[Y_{t-i}] + \lambda \tag{3}$$

Noting that (3) is a geometric series for ρ_i, then

$$\lim_{t\to\infty} E[Y_t] = \frac{\lambda}{\left(1 - \sum_{i=1}^{p} \rho_i\right)} \equiv \mu \tag{4}$$

Since $E[Y_0] = \mu$ (by definition), Eq. (1) can be written

$$E[y_t \mid Y_{t-1}] = \sum_{i=1}^{p} \rho_i Y_{t-i} + \left(1 - \sum_{i=1}^{p} \rho_i\right)\mu \tag{5}$$

This is a stationary linear AR(p) process. Note that the derivation makes no use of the distribution of y_t. The only role that the distribution of y_t plays is in defining the admissible

values of ρ_i (more on this below). The derivation requires only that finite first and second moments exist. Also, generalizing this result to an AR(p) model requires only a simple set of stationarity conditions for the admissible set of autoregressive coefficients.

We use this linear AR(p) model and the assumption that the event counts are Poisson distributed to define a new transition equation for a state–space model with an AR(p) process.

This Poisson autoregressive or PAR(p) model can be defined as follows.

1. Suppose that the observed event counts, y_t for $t = 1, 2, \ldots, T$, are drawn from a Poisson distribution conditional on m_t:

$$\Pr(y_t \mid m_t) = \frac{m_t^{y_t} e^{-m_t}}{y_t!} \tag{6}$$

   This defines the measurement equation for the observed data.
2. Assume that m_t is the conditional mean of the linear AR process of $E[y_t \mid Y_{t-1}]$ as in Eq. (5). This defines the state variable for the model. Since the Poisson distribution is the measurement equation, this state density is in the exponential family and can be characterized by its mean m_t and variance σ_t.
3. Finally, assume that the density of the state variable has a gamma-distributed conjugate prior, so

$$\Pr(m_t \mid Y_{t-1}) = \Gamma(\sigma_{t-1} m_{t-1}, \sigma_{t-1}), \qquad m_{t-1} > 0, \qquad \sigma_{t-1} > 0 \tag{7}$$

   with $m_{t-1} = E[y_t \mid Y_{t-1}]$ and $\sigma_{t-1} = Var[y_t \mid Y_{t-1}]$. The prior is constructed using the observed data. We do this by finding the conditional mean and variance of the data at time t based on the previous $t-1$ observations. The prior distribution is a gamma with mean m_{t-1} and variance m_{t-1}/σ_{t-1}.

These computations require filtering the data as in the PEWMA model using Eqs. (5), (6), and (7). Since the prior is gamma, using an extended Kalman filter, the conditional distribution at time t given $t-1$ is also gamma: $m_{t|t-1} \sim \Gamma((m_{t|t-1}\sigma_{t|t-1}), \sigma_{t|t-1})$. The reader is referred to the Appendix for a discussion of the PAR(p) filtering equations.

Since the measurement equation is Poisson, and the state equation is gamma, we can use the same updating procedure as in the PEWMA to derive the forecast density for the one-step ahead distribution. This provides an estimate of the posterior for time t:

$$\begin{aligned}
\Pr(y_t \mid Y_{t-1}) &= \int_\theta \Pr(y_t \mid \theta_t) \Pr(\theta_t \mid Y_{t-1})\, d\theta \\
&= \int_\theta \frac{\theta_t^{y_t} e^{-\theta_t}}{y_t!} \cdot \frac{e^{-\sigma_{t|t-1}\theta} \theta_t^{\sigma_{t|t-1} m_{t|t-1} - 1} \sigma_t^{\sigma_{t|t-1} m_{t|t-1}}}{\Gamma(\sigma_{t|t-1} m_{t|t-1})} \\
&= \frac{\Gamma(\sigma_{t|t-1} m_{t|t-1} + y_t)}{\Gamma(y_t + 1)\Gamma(\sigma_{t|t-1} m_{t|t-1})} (\sigma_{t|t-1})^{\sigma_{t|t-1} m_{t|t-1}} \\
&\quad \times (1 + \sigma_{t|t-1})^{-(\sigma_{t|t-1} m_{t|t-1} + y_t)}
\end{aligned} \tag{8}$$

where $\Gamma(\cdot)$ denotes the gamma function. This is a negative binomial distribution. Based on

this distribution, we can construct the log-likelihood for the PAR(p) as follows:

$$\begin{aligned}\mathcal{L}(m_{t-1}, \sigma_{t-1} \mid y_t, \ldots, y_T; Y_{t-1}) &= \ln \prod_{t=1}^{T} \Pr(y_t \mid Y_{t-1}) \\ &= \sum_{t=1}^{T} \ln \Gamma(\sigma_{t-1} m_{t-1} + y_t) - \ln \Gamma(y_t + 1) \\ &\quad - \ln \Gamma(\sigma_{t-1} m_{t-1}) + \sigma_{t-1} m_{t-1} \ln(\sigma_{t-1}) \\ &\quad - (\sigma_{t-1} m_{t-1} + y_t) \ln(1 + \sigma_{t-1}) \end{aligned} \tag{9}$$

Substituting the linear AR(1) process for m_t yields a PAR(1) model with a negative binomial predictive distribution. Thus, a linear AR(p) process can be used to generate the same distribution and a similar likelihood as the PEWMA model. Covariates can be introduced by replacing μ with $\exp(X_t \delta)$ in (5).

Using the properties of the negative binomial distribution, we can derive the forecast function for the conditional mean and variance of a PAR(p) series based on the optimized values of ρ, δ, m_t, and σ_t. The one-step ahead conditional forecast function for the PAR(p) model is

$$E[y_{t+1} \mid Y_t] = m_{t+1|t} = \sum_{i=1}^{p} \rho_i m_{t|t-1} + \left(1 - \sum_{i=1}^{p} \rho_i\right) \mu$$

The forecast variance is

$$Var[y_{t+1} \mid Y_t] = \frac{1 + \sigma_{t+1|t}}{\sigma_{t+1|t}} m_{t+1|t}$$

2.3 *Interpretation of the PAR(*p*)*

The PAR(p) model's interpretation differs from the Poisson and negative binomial models in a significant way. Consider the PAR(p) model with a covariate matrix X_t and $\mu = \exp(X_t \delta)$. The effect of a change in the regressor X_t is now given by an impact multiplier, as in a Gaussian linear autoregression model. The impact multiplier for the effect of a change in X_t on the mean number of counts at time t is determined by calculating the value of the first derivative of the mean function for this change. For the PAR(p), this derivative is

$$\frac{\partial m_t}{\partial X_t} = \frac{\partial \left(\sum_{i=1}^{p} \rho_i Y_{t-i} + \left(1 - \sum_{i=1}^{p} \rho_i\right) \exp(X_t \delta) \right)}{\partial X_t} = \left(1 - \sum_{i=1}^{p} \rho_i\right) \exp(X_t \delta) \cdot \delta \tag{10}$$

This is the instantaneous effect of a shock in X_t on the mean m_t. For the PAR(p) model, the instant effect of the change in the independent variable or impact multiplier for the number of counts at time t depends on the estimated value of the regression parameters and on the estimated values of ρ_i. This is in contrast to the standard Poisson regression model, where the instantaneous (and long-run) estimated effect would be $\exp(X_t \delta) \cdot \delta$.

This PAR(p) impact multiplier can then be used to compute the long-run multiplier for the total effect of a shock to X_t, as in Gaussian time series analysis. The long-run multiplier, which can be compared to the parameter estimates from other event count regression models, measures the effect of the shock accounting for the dynamic effects of the shock on the

conditional mean number of events. Analogous to a Gaussian time series regression model, the long-run multiplier is

$$\frac{\partial m_t/\partial X_t}{\left(1-\sum_{i=1}^{p}\rho_i\right)} = \frac{\left(1-\sum_{i=1}^{p}\rho_i\right)\exp(X_t\delta)\cdot\delta}{\left(1-\sum_{i=1}^{p}\rho_i\right)} = \exp(X_t\delta)\cdot\delta \tag{11}$$

The long-run effect for the PAR(p) model in Eq. (11) is computed with the same equation used to compute effects in standard Poisson or negative binomial regression models with an exponential link function. The implication is that standard factor change and percentage change calculations that are used to interpret instantaneous changes in event count models with exponential link functions are not valid if the true data generation process is a PAR(p). This means that we should expect differences in the estimated short-run and long-run effects of shocks to exogenous variables using the PAR(p) model. The reason for this difference is that the PAR(p) accounts for the effects to the change in the covariates and the dynamic responses to the changes in this covariates over time. We demonstrate that this is the case in our Monte Carlo analysis in Section 4 and our example in Section 5.

3 Identification and Specification: PEWMA or PAR(p)?

Since determining whether the PAR(p) or PEWMA model is appropriate for a particular problem cannot be done using a pretest, we would like to find a set of diagnostic tools to identify model order and specification. Distinguishing between the PEWMA model of Brandt et al. (2000) and the PAR(p) is important because the former is a nonstationary process and the latter is a stationary process.

This task of choosing between the models is an identification task—we must find the plausible dynamic process that generates the data to determine which model to estimate. The PEWMA model produces a persistent sample autocorrelation function (ACF). It turns out that ACF can also be used for the PAR(p) because it is based on a linear AR process. Grunwald et al. (1997b) show that the autocorrelation function for linear AR(1) models (for many conditional distributions) can be written in the same form as the standard Gaussian time series model. Standard ACF routines and standard errors can be used to diagnose general AR(1) structures for count data because the derivation of the ACF does not depend on the assumption of normality. We conjecture that the same result also applies to the AR(p) model, so the sample ACF can be used to determine the model order.

For the PAR(p) model, admissible values of the ρ_i will depend on the mean and variance of the series. A PAR(1) model with $0 < \rho < 1$ will be admissible. However, a model with $-1 < \rho < 0$ will be admissible only if the mean of the series is large enough to offset the oscillatory nature of the series. For social science data, such extreme oscillations rarely occur in the levels of the variables.

In the PAR(1) case, the stationarity conditions are the same as for the Gaussian model. Beyond this, the criteria for admissibility will again depend on the mean level. We do not try to derive exact conditions for admissibility because as we know from the Gaussian case that such a derivation does not provide us with any practical information. A *sample* of data may produce estimates that are not admissible. Inadmissible values happen because, as we know from time series analysis, historical samples always contain a large amount of information that does not replicate outside the sample. Thus, for making inferences about covariates, admissibility is not a concern.

To demonstrate that the ACF provides a reliable diagnostic for count data from the PAR model, Fig. 1 presents several sample series. We generated each series from the PAR(1)

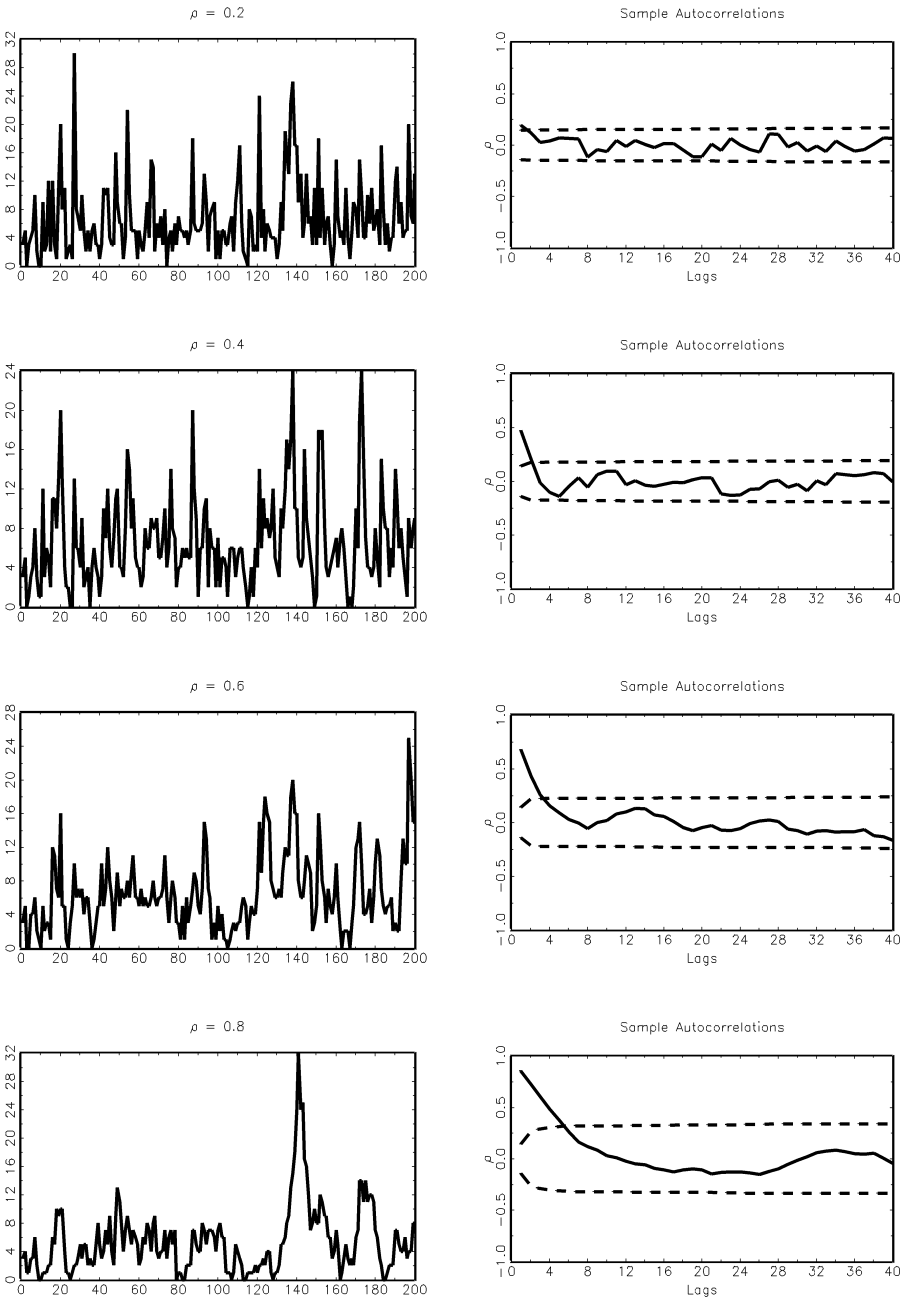

Fig. 1 Simulated PAR(1) series and ACFs.

model with different values for the autoregressive parameter ρ.[6] The first column shows the series, and the second the ACF for the series. Note how the series dynamics are accurately reflected in the sample ACFs. The values of ρ are 0.2, 0.4, 0.6, and 0.8. As the value of

[6]The DGP for these figures assumed that $m_{t|t-1} = \rho y_{t-1} + (1-\rho)\exp[\ln(5) + 0.75X_t]$, where $X_t \sim N(0,1)$ for $t = 1, 2, \ldots, 200$.

ρ approaches 1, the series become much smoother and demonstrate a longer cycle, as in Gaussian models. The estimated sample autocorrelation functions are very close to the true values used to generate the data and decline exponentially to zero after the first few lags. This is in contrast to the ACF for the PEWMA model, where the persistence lasts over many lags (see Brandt et al. 2000).

The dynamic process of time series event count data is easily identified from the ACFs. The sample ACF is not specific to the Poisson model and, in fact, can be found using the standard ACF routines in existing software packages. What should be noted is that the estimated ACF should generally be positive for PAR(1) models. This is because for small mean event counts, a negative value of ρ implies that the series may drop below zero in some finite sample.

Identification of the PAR(p) process also follows the standard Gaussian procedures. As an approximation, a PAR(p) model produces an ACF that reflects the dynamics of the event count series. For example, a PAR(2) with $\rho_1 = 0.7$ and $\rho_2 = -0.4$ will produce a smooth and dampening oscillatory pattern in the ACF. Box and Jenkins (1976) show that an ACF is enough for identification of an ARIMA model. The same holds for a PAR(p) because the autoregressive model for the latent mean is linear.

Identifying whether the observed event count data follows a persistent event count process like the PEWMA model or a stationary memoried process such as the PAR(p) is important for model specification. Failing to model the dynamics adequately will be a source of inefficiency. However, the exact implications of dynamic misspecification for the PAR(p) are unknown. We address this issue in the next section.

4 Monte Carlo Evidence

The Monte Carlo experiments we conduct to compare the various estimators for event count data (e.g., static Poisson regression and negative binomial regression) demonstrate the implications of model misspecification. We determine the implications of estimating a standard Poisson regression, negative binomial regression, or ordinary least-squares regression in the case where the true data generation process is the PAR(p).

The Monte Carlo experiments are conducted as follows. We generate 200 Monte Carlo replications from the following PAR(1) data generation process:

$$\begin{aligned} y_t &\sim \text{Poisson}\,(m_t) \quad \forall\, t \\ m_t &= \rho y_{t-1} + (1-\rho)\exp(\delta_0 + X_t\delta_1), \qquad t = 1, 2, \ldots, T \\ m_t &\sim \Gamma((\sigma_{t-1}m_{t-1}), \sigma_{t-1}) \end{aligned}$$

We varied several of the parameters in the experiment to investigate the robustness of the various estimators to PAR data. We tried three sample sizes, $T = 100,\ 200$, and 500. The matrix of independent variables in the PAR(1) data included an intercept and a T–vector of independent random variables distributed $N(0, 1)$. This regressor is fixed and nonstochastic in each experiment. The vector of regression parameters is chosen so that the intercept (mean) is 20. The value of the regression parameter for the random variable is 0.5, so $\delta = [\delta_0, \delta_1] = [\ln(20), 0.5]$. The values of ρ also vary $(0.4, 0.6, 0.8)$ to investigate the effect of the magnitude of the AR(1) dynamic.[7]

[7] We also conducted a set of Monte Carlo experiments with $\delta = [\ln(10), 0.5]$ that produce results nearly identical to those reported here. These results can be seen at the *Political Analysis* website.

We estimate the PAR(1) model as a baseline for the experiments. We also estimate a Poisson regression model, a lagged Poisson regression model, a negative binomial model, and an OLS model with a natural log of the dependent variable and a lagged endogenous dependent variable. The regression models we analyze have the following conditional means.

$$\begin{aligned}
\text{Poisson and negative binomial:} \quad & E[y_t \mid X_t] = \exp(\delta_0 + X_t\delta_1) \\
\text{Lagged Poisson:} \quad & E[y_t \mid X_t, y_{t-1}] = \exp(\delta_0 + X_t\delta_1 + \rho y_{t-1}) \\
\text{Logged-lagged OLS:} \quad & E[\ln y_t \mid X_t, y_{t-1}] = \delta_0 + X_t\delta_1 + \rho \ln(y_{t-1}) \\
\text{PAR}(p): \quad & E[y_t \mid X_t, y_{t-1}, \ldots, y_{t-p}] \\
& = \sum_{i=1}^{p} \rho_i y_{t-i} + \left(1 - \sum_{i=1}^{p} \rho_i\right) \exp(X_t\delta)
\end{aligned}$$

where $p = 1$. We do not include the PEWMA model, since it cannot account for mean reversion and cyclical patterns without the inclusion of explicit cyclical components [as in a seasonal model (see Harvey 1989, pp. 420–422)]. Since we clearly identify when to use the PEWMA based on sample ACFs, estimating this model is not an issue in the case when a PAR(p) process generates the data.[8]

For each model, we compute the maximum-likelihood estimate of the parameters. Recall that the effects of a change in the regressor(s) are estimated differently in the PAR(p) model. Thus to be able to compare the models, looking only at the coefficients is incorrect. The quantity of interest for a dynamic regression such as the PAR(p) is the long-run multiplier. To evaluate the Monte Carlo results, we compute this quantity using Eq. (11) to find the long-run or total effect of a one-standard deviation change in the single regressor for the PAR(p) model. The estimated effect of a one-standard deviation change is computed in each experiment and it is an estimate of the long-run effect of a change in the regressor on the number of events. A similar computation is then used for the other models using the formula $\exp(\delta_0 + X_t\delta) \cdot \delta_1$.

Figure 2 presents box and whisker plots of the distributions of the long-run effects of a one-standard deviation shock for each of the Monte Carlo experiments and each of the different mean specifications. Each row of graphs in the figure corresponds to a different sample size specification, while each column corresponds to a different value of ρ. For each set of estimated effects, the box and whisker plots show the median, upper, and lower quartiles of the effects, while the whiskers show the (empirical) 90% confidence interval, computed from the Monte Carlo results.

For the Poisson, lagged Poisson, logged-lagged OLS, and negative binomial regression models we see that the effect of a change in the regressor is underestimated relative to the true long-run effect implied by the PAR(1) model. The PAR(1) model is unbiased, as indicated by the horizontal line in each graph. In all cases, the empirical 90% interval is always below the true value of the PAR(1) effect for the rival estimators. When the sample size is large, the bias in the estimated long effect is more magnified. As the sample size grows, failing to account for the dynamics of the PAR(1) process leads to estimated effects whose 90% confidence intervals are further from the true effect of a one-standard deviation shock in the regressor.

[8] A negative binomial model with a lagged dependent variable is omitted because of difficulties with estimation. This model produces a singular Hessian because of the inability of the model to estimate a nonzero overdispersion term. We omit the model as infeasible given the data generation process.

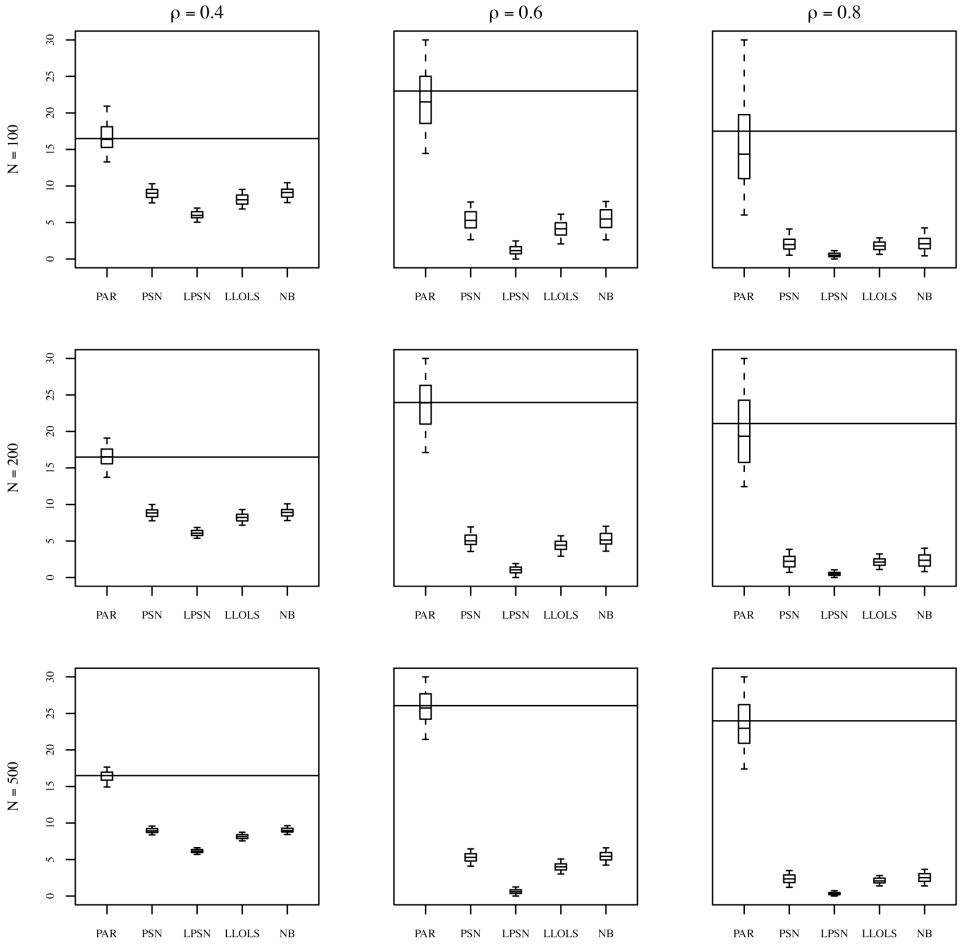

Fig. 2 Box and whisker plots for Monte Carlo estimates of effects of a shock in the exogenous regressor. These box plots illustrate the summary statistics for the Monte Carlo experiments with $\mu = 20$. The rows correspond to different sample sizes; the columns, to different values of ρ. Each box defines the median (center line), first quartile (lower edge of box), and third quartile (upper edge of the box). The end points of the whiskers denote the limits of the 90% confidence interval, computed from the Monte Carlo replications. The true effect of a 1-unit change in the regressor is indicated by the horizontal line. The estimators are PAR(1), PSN (Poisson), LPSN (lagged Poisson), LLOLS (logged-lagged OLS), and NB (negative binomial). See text for discussion.

The Monte Carlo results make intuitive sense when we consider the implications of specification uncertainty that arise when endogenous variables that are part of the data generating procedure for the event count at time t are omitted from the model. Omitting the effect of lagged observations from the model is a source of inconsistency and inefficiency (White 1994, Chap. 4). Failing to account for the dynamics in the model implies a greater uncertainty about the parameter estimates.

5 Application: Hyde Park Purse Snatchings

As an example applying the PAR(p) model, we replicate a classic intervention model presented by McCleary and Hay (1980). McCleary and Hay (1980) use data from Reed

(1978) on purse snatchings in the Hyde Park neighborhood of Chicago to demonstrate how an intervention analysis depends on the dynamic specification and treatment of outliers in a time series model. The data they use is a count of the number of purse snatchings every 28 days from January 1969 to September 1974. During the 42nd period a community crime prevention program, Operation Whistlestop, was implemented (Reed 1978). The goal of the analysis is to evaluate whether the program led to a significant reduction in purse snatchings.

Initially, McCleary and Hay use these data to show how outliers affect intervention analysis. They demonstrate that an outlier in period 66 led to an incorrect conclusion that the intervention had a significant effect because it obscured the identification of the ARIMA model. Their analysis identified an ARIMA(2,0,0) model for the purse snatching series after correcting the outlier. To evaluate the effect of Operation Whistlestop, a permanent intervention is specified. The intervention variable equals 0 for the first 41 periods and 1 thereafter. Their analysis indicates that the intervention is insignificant. As we will see, while correcting the dynamic specification for the model changes the inference, this is only partially correct since it does not account for the event count properties of the data.

To analyze the effects of the distributional and dynamic specifications for the purse snatching series, we estimate four plausible models:[9]

1. ARIMA(2,0,0) intervention model (to replicate McCleary and Hay's analysis),
2. PAR(1) intervention model,
3. PAR(2) intervention model, and
4. negative binomial regression.

The most likely models are those with an AR structure, since the ACF for the purse snatching series appears to decline quickly and is insignificant after five lags. For each of these models we use the same intervention specification employed by McCleary and Hay.

The parameter estimates for these models are presented in Table 1. The ARIMA(2,0,0) results differ slightly from those reported by McCleary and Hay. Nevertheless, we reach

Table 1 Hyde Park purse snatchings analysis (standard errors in parentheses)

Model	*ARIMA (2,0,0)*	*PAR(1)*	*PAR(2)*	*Negative binomial*
Operation Whistlestop	−2.460	−0.606	−0.597	−0.382
	(3.250)	(0.107)	(0.114)	(0.116)
Constant	14.48	2.928	2.912	2.767
	(3.000)	(0.080)	(0.084)	(0.068)
ρ_1	0.298	0.125	0.083	
	(0.095)	(0.022)	(0.057)	
ρ_2	0.350		0.116	
	(0.093)		(0.046)	
γ				0.740
				(0.252)
N	71	71	71	71
Final LLF	−227.10	−222.66	−220.47	−228.69
AIC	462.20	449.32	444.94	461.38

[9] We also estimated four other regression models: PEWMA, Poisson, Poisson with a lagged count, and negative binomial with a lagged count. The PEWMA model produced an insignificant positive intervention effect and the remaining models produced results nearly identical to the negative binomial model. The full results are available at the *Political Analysis* website.

Table 2 Long-run effects of Operation Whistlestop

Model	*Long-run effect*	*Percentage change in mean*
PAR(1)	−6.18	−41.8
PAR(2)	−6.04	−42.0
Negative binomial	−4.14	−31.7

the same conclusion: the intervention effect is not significantly different from zero. The remaining models all produce significant, negative intervention coefficients.

The PAR(1) and PAR(2) models both produce intervention coefficients of −0.6. To compare the effects of the Operation Whistlestop intervention across the estimators, we must compute the long-run multipliers. The formulas given earlier can be used for the PAR(p) models, and a similar computation can be estimated for the negative binomial model by computing the expected number of snatched purses before and after the intervention. The total effect of the Operation Whistlestop program is different in each of these models. Table 2 shows the estimated long-run effects of the intervention for the PAR(1), PAR(2), and negative binomial model. In this table, we compute the effects of the intervention for the PAR(p) model using Eq. (11). For the negative binomial model, the long-run effect is the same as the instantaneous effect, computed using $\exp(X_t\delta)\cdot\delta$. The estimated long-run effect for the PAR(p) models is a total decline of 6 purse snatchings. For the negative binomial model, the effect of the intervention is a decline of 4.14 purse snatchings. However, the interpretation of these effects is very different. For the PAR(p) models, this effect occurs over several periods. The impact multiplier for the immediate effect of the intervention is −5.41 purse snatchings for the PAR(1) model and −4.84 purse snatchings for the PAR(2). Both of these instantaneous effects are larger than the instantaneous effects of the Poisson and negative binomial models. The Poisson and negative binomial models' interpretation is that the effect of the intervention is immediate. The omission of the dynamics from the negative binomial model leads to an understatement of the magnitude of the intervention, since it fails to account for dynamics of a future drop in purse snatchings.

This difference in the static versus dynamic event count models can be seen by computing the total percentage change in the number of purse snatchings after the intervention. For Poisson, PEWMA, or negative binomial regressions, this total percentage change is calculated by taking the exponentiation of the intervention parameter, subtracting 1, and multiplying by 100. For the PAR(p), where $X_t = [x_t, z_t]$ and $\beta = [\beta_1, \beta_2]$ are the regression parameters for x_t and z_t, respectively, the estimated instantaneous percentage change in the counts for a change Δz_t can be found by

$$100\frac{m_{\triangle z} - m_z}{m_z} = 100\left[\frac{\left(1-\sum_{i=1}^{p}\rho_i\right)\exp(x_t\beta_1)\exp(z_t\beta_2)(\exp(\Delta z_t\beta_2)-1)}{\sum_{i=1}^{p}\rho_i y_{t-i} + \left(1-\sum_{i=1}^{p}\rho_i\right)\exp(x_t\beta_1 + z_t\beta_2)}\right] \tag{12}$$

The long-run percentage change is found by dividing this quantity by $(1-\sum_{i=1}^{p}\rho_i)$.

These percentage changes are shown in column 3 in Table 2. For the negative binomial model, the total change due to the intervention is a decline of 31.7% from the mean of the series. For the PAR(p) models, the total percentage change in the number of purse snatchings is 42%—a decline that is 11 points more than what we would have estimated from the negative binomial model.

This reanalysis of the Hyde Park purse snatching data demonstrates that accounting for the dynamic properties of stationary event count time series does matter for inference. Using only an ARIMA model for the purse snatching data leads to the incorrect inference that the

Operation Whistlestop program had no effect on the number of purse snatchings. Using a Poisson or negative binomial model leads to the inference that the program had a discernible effect but leads to understating the magnitude of this effect. This finding is consistent with our earlier Monte Carlo analysis showing that estimation of a Poisson or negative binomial regression may lead to incorrect estimates of the effects of regression parameters. Only the PAR(p) models produce inferences that are broadly consistent with the data: the effects of the program are large at first but do persist over several periods consistent with the AR(p) dynamics of the data.

6 Application: Presidential Vetoes

Another literature using event count time series is the study of presidential vetoes. Lewis and Strine (1996) build on earlier literature on presidential vetoes to test several hypotheses about changes in presidential leadership over the last 100 years. They point out that four theories have been proposed to explain presidential power over time.

1. Gradual secular increase in presidential power
2. Cycles of presidential power
3. Early versus modern presidents' use of power
4. Political timing of presidential power

Gradual secular change refers to the expansion of presidential powers since the founding of the nation (e.g., Corwin 1954; Rossiter 1960). Cycles of presidential power are the regimes of presidencies that cycle between the constraint and the expansion of public expectations of the presidential actor's ability to meet public expectations (e.g., Skowronek 1993). Within the major regimes (Jeffersonian, Jacksonian, Lincolnian, and Rooseveltian), one should expect cycles of expansion and consolidation in the use of presidential power. Modern time refers to the transformation of the presidency after Franklin Roosevelt. A number of scholars (e.g., Neustadt 1990; Wayne 1978) have argued that analysis of presidential actions indicates a dramatic shift in presidential activity after Roosevelt. The distinction between early and modern time implies a shift in the presidential vetoes after Roosevelt. Finally, the political timing or electoral cycles argument refers to the variation over time within a presidential administration. Light (1982) argues that presidents are more powerful early in their terms but that this power declines in second terms and as presidencies reach their later years.

Using veto data from 1890–1994, Lewis and Strine find that the early/modern theory and the cyclical theory offer the best explanations for the presidential uses of the veto. They argue that the data do not support the political timing theory. The latter conclusion is based on a Poisson regression analysis of the number of vetoes on public laws from 1890 to 1994 and an analysis that focuses explicitly on the post-1945 (post-Roosevelt) data. They find that there is a noticeable shift in the Poisson regression parameters before and after Roosevelt and that there is weak support for the effects of term in office and year in term based on their analysis.

One issue that needs to be addressed when working with data such as veto data is correctly accounting for the dynamic patterns in the data. Figure 3 shows the data series and its ACF. The ACF for the whole data series indicates that there is some serial correlation present in the data. As Lewis and Strine discuss, the dynamics of modern electoral politics has changed the pattern of presidential vetoes since FDR. Thus, ACFs are also shown for the pre- and post-FDR periods. Prior to FDR, we find that there is no significant dynamic pattern in the data. Post-FDR we find a noticeable dynamic pattern, with positive significant spikes in the ACF for the veto series.

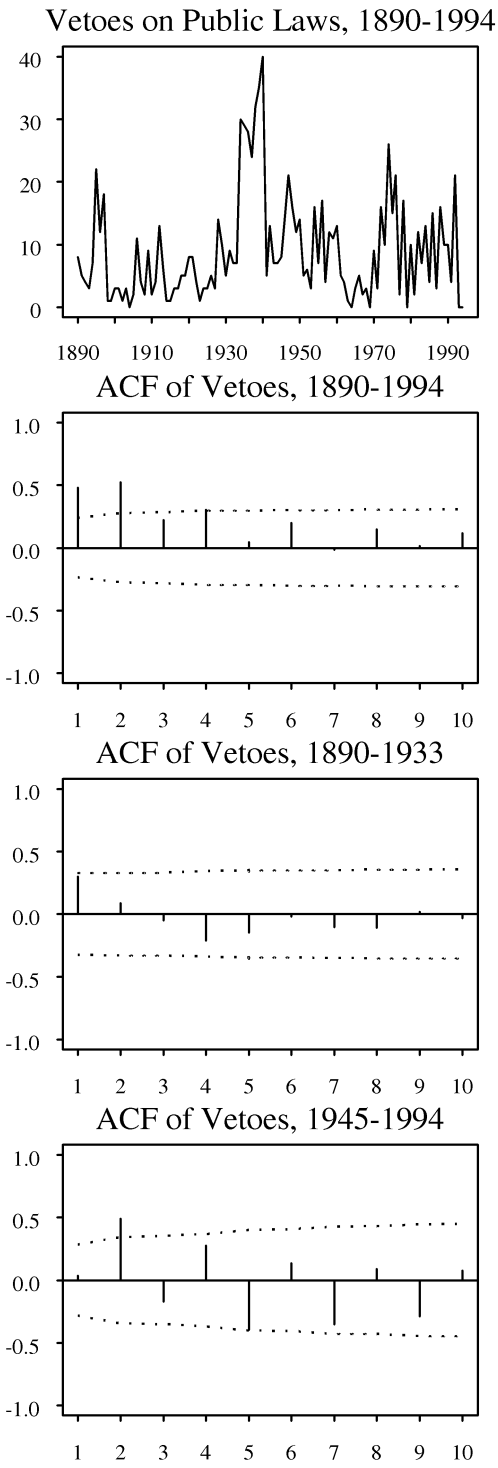

Fig. 3 Vetoes on public laws and ACFs for selected subsets, 1890–1994.

The implication of this finding is that post-FDR veto behavior may contain significant serial correlation. To see if this is the case, we estimate a dynamic regression model of veto behavior. We use the same regressors as Lewis and Strine (1996): the number of public laws per annum, the percentage of seats held in Congress, the average annual unemployment percentage, the percentage of Gallup respondents approving of the president's job performance, a dummy variable for succession presidencies, a dummy variable for whether the United States was at war, a discrete variable indicating the year in the president's term, and a dummy variable indicating the president's term in office. The unemployment rate and presidential approval are AR(1) processes, so our analysis includes lagged values of the unemployment and presidential approval regressors. The dynamic specification of the regressors is important because failing to model properly how the dynamics of the regressors enter the model can lead to an incorrect inference about the dynamic properties of the dependent variable (Hendry et al. 1984; Beck 1991).

We present the results of our analysis in Table 3. Here we present three regression models. The first is identical to the Lewis and Strine model (their Table 5) omitting the election year variable. We then estimate a Poisson regression with the lagged exogenous variables

Table 3 Poisson and PAR(1) regression analysis of vetoes, 1945–1994

Variable	*Poisson*	*Poisson*	*PAR(1)*
Constant	3.1372	4.0465	3.9356
	(0.7291)	(0.8609)	(0.7141)
Public laws	0.0036	0.0034	0.0031
	(0.0005)	(0.0006)	(0.0008)
Public laws$_{(t-1)}$		−0.0011	−0.0009
		(0.0006)	(0.0008)
Seat percentage	−0.0365	−0.0390	−0.0424
	(0.0069)	(0.0071)	(0.0080)
Unemployment	0.0714	0.0169	−0.0442
	(0.0347)	(0.0561)	(0.0614)
Unemployment$_{(t-1)}$		0.0199	0.0944
		(0.0482)	(0.0599)
Approval	−0.0136	−0.0131	−0.0191
	(0.0049)	(0.0052)	(0.0079)
Approval$_{(t-1)}$		−0.0021	0.0090
		(0.0025)	(0.0060)
Succession	0.5594	0.5139	0.5206
	(0.1314)	(0.1414)	(0.1634)
War	0.0354	−0.0669	−0.0120
	(0.1219)	(0.1395)	(0.0429)
Year in term	−0.0940	−0.1238	−0.1435
	(0.0577)	(0.0654)	(0.0802)
Term	−0.0554	−0.0051	−0.0771
	(0.1291)	(0.1415)	(0.2219)
ρ			−0.1057
			(0.0552)
N	50	50	50
Log-likelihood	−155.88	−153.53	−150.95
AIC	329.76	331.06	325.90

Table 4 Presidential veto analysis long-run multipliers

Variable	*Poisson*	*Poisson with lagged regressors*	*PAR(1)*
Public laws	0.032	0.021	0.020
Seat percentage	−0.328	−0.350	−0.368
Unemployment	0.641	0.331	0.437
Approval	−0.123	−0.137	−0.088
Succession	5.024	4.614	4.522
War	0.318	−0.601	−0.104
Year in term	−0.844	−1.112	−1.246
Term	−0.497	−0.046	−0.670

included. Finally, we estimate a PAR(1) model with the lagged exogenous variables. The results are generally identical across the models, with the number of public laws, the percentage of seats, presidential approval, succession presidency, and the year in term variable all statistically significant.

The PAR(1) results show that despite the presence of serial correlation in the ACF for the data series, modeling this serial correlation using the PAR(1) has little substantive effect on the coefficients. In fact, the PAR(1) results are basically identical to the Poisson regression estimates. The same coefficients are significantly different from zero in each of the models. Note also that a likelihood ratio test would not lead us to reject the simplest Poisson regression model.

Of additional interest are any possible differences that may emerge in the model interpretation using the PAR(1) model. Recall that the dynamic multipliers from the PAR(1) specification are used to assess the effects of one unit changes in the regressors on the mean number of events. Table 4 presents the estimated long-run multipliers for the PAR(1) models of the veto data as well as the estimated effects from the Poisson regressions. The long-run effects for the PAR(1) are computed using the formula reported in Eq. (11).

Note that for the significant regressors, there is very little change in the long-run effects across the three specifications. The long-run effect for a 1-unit change in the public law variable changes by less than 0.012 across the specifications. A one-standard deviation shock in this variable is 128 public laws, implying that under the Poisson regression specification with no lagged regressors that there would be an increase of 3.68 vetoes; in contrast, the other Poisson regression and PAR(1) models imply a 2.3-veto increase. The seat percentage variable's long run effect is identical across the specifications. The long-run effect of a shock to presidential approval does vary across the models. The Poisson regression without the lagged regressors implies that for each 1-unit increase in approval, the number of vetoes would decline by 0.123. Since the standard deviation for this variable is approximately 12 points, this implies a decline of nearly 1.5 vetoes for each one standard deviation change in approval. In contrast, the PAR(1) long-run multiplier would indicate an effect that is nearly 30% smaller.

The results of this reanalysis do not alter the main results of Lewis and Strine (1996): explanations of presidential power and veto behavior based on political timing, or the year a president is in office, appear to have weak support in the data. Even when we include a more robust dynamic specification of the event count time series regression model, we see that the substantive coefficients in the model change very little. Thus, in some cases, event count time series models such as the PAR(p) may not change the substantive interpretations of Poisson or other event count regressions.

7 Conclusion

Time series models for event count data appear to have wide application in political science. However, until the development of the PEWMA model for persistent time series data and the PAR(p) model, few tools have been available to explore the dynamic properties of event count data. These state–space models provide a simple method for estimation and inference about event count data with simple time series dynamics.

The event count time series model specification steps are as follows.

1. **Identification:** Determine the presence of dynamics using autocorrelation function plots. This provides a preliminary assessment of the order and magnitude of the dynamics.
2. **Specification and Estimation:** Estimate the model based on the dynamics found from identification.
   (a) If the dynamics are strong and persistent, estimate a PEWMA model.
   (b) If the ACF shows a dampening dynamic process, estimate a PAR(p) model.
   (c) If no dynamics are present, estimate a negative binomial, Poisson, or generalized event count model.
3. **Diagnosis and Testing:** Using standard testing procedures, evaluate the estimated models against alternative specifications. This includes testing any estimated dynamic model against the Poisson regression model, which is nested in both the PAR ($\rho = 0$) and the PEWMA ($\omega = 1$).

Following these steps allows analysts to determine the appropriate model for a particular problem. Failure to account properly for the dynamics has consequences for event count data. If the true data generation process is a PEWMA model, and an analyst estimates a Poisson or negative binomial regression, Brandt et al. (2000) show that the estimates are inefficient. If, on the other hand, the PAR(p) process generates the data, then the Poisson and negative binomial models produce estimates that are both biased and inefficient. Thus, the identification and estimation of these models are closely related and must be considered when specifying possible time series models for event count regressions.

To implement these procedures, we have written a set of GAUSS programs. These programs are available both as source code and as a set of new procedures in the COUNT program of King (1994).[10] The procedures can be used to estimate PEWMA and PAR(p) models. These procedures allow for the identification, estimation, and testing of these time series models for event count data. In addition, the procedures can be used to generate filtered estimates of the data. Our experience with these procedures indicates that they work well on a wide variety of event count data, both "real" and simulated.

A Appendix: The PAR(p) Filter

This Appendix briefly outlines the calculations needed to compute the PAR(p) filter. This filter is required to construct the predictive distribution of the event counts and the log-likelihood function. The filter is constructed based on the following assumptions.

[10]The programs and replication materials can be found at the *Political Analysis* website.

1. **Observation/Measurement Density:** The observed counts are generated by a marginal Poisson distribution, with mean (state variable) m_t:

$$\Pr(y_t \mid m_t) = \frac{m_t^{y_t} e^{-m_t}}{y_t!} \tag{A1}$$

2. **Transition Equation:** The mean or state variable of the marginal Poisson distribution evolves according to a stationary AR(p) process with autocorrelation parameters $\rho_i,\ i = 1, 2, \ldots p$,

$$m_t = \sum_{i=1}^{p} \rho_i y_{t-i} + \left(1 - \sum_{i=1}^{p} \rho_i\right) \exp(X_t \delta) \tag{A2}$$

   where X_t is a $T \times k$-matrix of regressors that are exogenous of $y_{t-i},\ \forall i = 1, 2, \ldots p$, and δ is a $k \times 1$-vector of regression parameters. Note that, unlike the PEWMA model, we assume that X_t includes an intercept.
3. **Conjugate Prior:** To complete the specification of the model, we assume that m_{t-1} is distributed gamma with location $(\sigma_{t-1} m_{t-1})$ and scale σ_{t-1}.

Using the mean and variance of the state variable is equivalent to using the gamma's location and scale as a_{t-1} and b_{t-1}, since we can alternatively assume that

$$m_{t-1} \mid Y_{t-1} \sim \Gamma(a_{t-1}, b_{t-1}) \tag{A3}$$

Since the gamma distribution is in the exponential family of distributions, the efficient estimator of the parameters a and b is equivalent to a moment estimator (i.e., the MLE estimator is a moment estimator). The optimal moment estimators would be of the form $1/(t-1)\sum_{t=1}^{t-1} m_{t-1} = a_{t-1}/b_{t-1}$ and $1/(t-1)\sum_{t=1}^{t-1} \ln(m_{t-1})$. However, since it may be the case that $\ln(m_t)$ cannot be computed (since $m_t \approx 0$), we use an alternative second moment condition, $\sum_{t=1}^{t-1} y_t^2$, which always exists. There is little loss in efficiency in this condition, and the estimator can be easily computed as a function of the observed variance of the data. Since these two statistics summarize the relevant information of a_{t-1} and b_{t-1}, we can solve and find that

$$\begin{aligned} E[m_t] &= \frac{1}{n}\sum_{t=1}^{t-1} m_t = \frac{a_{t-1}}{b_{t-1}} = m_{t|t-1} \\ Var[m_t] &= \frac{1}{n}\sum_{t=1}^{t-1} (m_t - E[m_t])^2 = \frac{a_{t-1}}{b_{t-1}^2} = \frac{m_{t|t-1}}{\sigma_{t|t-1}} \\ &\Rightarrow a_{t-1} = \sigma_{t-1} m_{t-1} \qquad \text{and} \qquad b_{t-1} = \sigma_{t-1} \end{aligned}$$

Thus, the first two moments can be used in place of directly specifying the location and scale parameters of the gamma distribution. We use these estimates to compute the filter. This also implies that the gamma distribution has mean m_{t-1} and variance m_{t-1}/σ_{t-1}.

The filter computations can be carried out using these assumptions. We proceed in three steps.

1. Combine the prior with the transition equation. We assume that m_{t-1} is conditionally gamma, as in Eq. (A3). Then from the properties of the gamma distribution and the

definition of the AR(p) process,

$$E[m_t \mid Y_{t-1}] = \sum \rho_i m_{t-i} + \left(1 - \sum \rho_i\right) \exp(X_t \delta)$$

This variable is gamma distributed, $\Gamma(m_{t|t-1}\sigma_{t|t-1}, \sigma_{t|t-1})$, based on the properties of the gamma distribution. Note that this expectation is computed based on the previous $t-1$ observations.

2. Update the value of m_t. Updating the conditional values of $m_{t|t-1}$ and $\sigma_{t|t-1}$, we find $m_t \mid Y_t$ via Bayes rule,

$$\begin{aligned}
\Pr(m_t \mid Y_t) &= \frac{\Pr(Y_t \mid m_t) \cdot \Pr(m_t)}{\int_0^\infty \Pr(Y_t \mid \theta_t) \cdot \Pr(\theta_t) d\theta_t} \\
&= \frac{\frac{e^{-m_t} m_t^{y_t}}{y_t!} \cdot \frac{e^{-\sigma_{t|t-1} m_t} m_t^{(m_{t|t-1}\sigma_{t|t-1})-1} \sigma_{t|t-1}^{(m_{t|t-1}\sigma_{t|t-1})-1}}{\Gamma(m_{t|t-1}\sigma_{t|t-1})}}{\int_0^\infty \frac{e^{-\theta_t} \theta_t^{y_1}}{y_1!} \cdot \frac{e^{-\sigma_{t|t-1} m_t} \theta_t^{(m_{t|t-1}\sigma_{t|t-1})-1} \sigma_{t|t-1}^{(m_{t|t-1}\sigma_{t|t-1})-1}}{\Gamma(m_{t|t-1}\sigma_{t|t-1})} d\theta_t} \\
&= \frac{e^{-m_t(1+\sigma_{t|t-1})} m_t^{y_t+(m_{t|t-1}\sigma_{t|t-1})-1} (1+\sigma_{t|t-1})^{y_t+(m_{t|t-1}\sigma_{t|t-1})}}{\Gamma(y_t + m_{t|t-1}\sigma_{t|t-1})}
\end{aligned}$$

which is a gamma distribution, $m_t \mid Y_t \sim \Gamma(m_{t|t-1}\sigma_{t|t-1} + y_t, \sigma_{t|t-1} + 1)$. We rewrite this $m_t \mid Y_t \sim \Gamma(m_t\sigma_t, \sigma_t)$.

3. Induction to $t+1$. Since the filter preserves the conjugate form, the model is recursive by induction. Thus, at time period t we have the recursions that generate the prior for the observation at period $t+1$.

At this point, we must specify a prior to be used to initialize the filter, i.e., the values of m_0 and σ_0. Nothing precludes setting $m_0 = \sigma_0 = 0$, since $\Pr(m = 0) = 1$ at this point of degeneracy. As in the PEWMA model [following Harvey and Fernandes (1989)] we can use the sample mean and variance of the series, as this yields the classical Kalman filter solution. Alternatively, one can adopt a Bayesian perspective and posit alternative priors for m_0 and σ_0.

References

Beck, Nathaniel. 1991. "Comparing Dynamic Specifications: The Case of Presidential Approval." *Political Analysis* 3:51–87.

Box, G. E. P., and G. M. Jenkins. 1976. *Time Series Analysis, Forecasting and Control*. San Francisco: Holden Day.

Box-Steffensmeier, Janet M., and Bradford S. Jones. 1997. "Time Is of the Essence: Event History Models in Political Science." *American Journal of Political Science* 41(4):1414–1461.

Brandt, Patrick T., John T. Williams, Benjamin O. Fordham, and Brian Pollins. 2000. "Dynamic Models for Persistent Event Count Time Series." *American Journal of Political Science* 44(4):823–843.

Brophy-Baermann, Bryan, and John A. C. Conybeare. 1994. "Retaliating Against Terrorism: Rational Expectations and the Optimality of Rule Versus Discretion." *American Journal of Political Science* 38(1):196–210.

Cameron, A. Colin, and Pravin K. Trivedi. 1998. *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

Corwin, Edward S. 1954. *The President: Office and Powers*. New York: New York University Press.

DeGroot, Morris H. 1970. *Optimal Statistical Decisions*. New York: McGraw–Hill.

Gelman, Andrew, John Carlin, Hal Stern, and Donald Rubin. 1995. *Bayesian Data Analysis*. London: Chapman and Hall.

Grunwald, Gary, Kais Hamza, and Rob Hyndman. 1997a. "Some Properties and Generalizations of Non-negative Bayesian Time Series Models." *Journal of the Royal Statistical Society, Series B* 59(3):615–626.

Grunwald, G. K., R. J. Hyndman, L. Tedesco, and R. L. Tweedie. 1997b. "A Unified View of Linear AR(1) Models." Clayton, Australia: Monash University.

Hendry, D. F., A. R. Pagan, and J. D. Sargan. 1984. "Dynamic Specification." In *Handbook of Econometrics, Vols. 2 and 3*, ed. Z. Griliches and M. D. Intriligartor. Amsterdam: North-Holland.

King, Gary. 1988. "Statistical Models for Political Science Event Counts: Bias in Conventional Procedures and Evidence for the Exponential Poisson Regression Model." *American Journal of Political Science* 32(3):838–863.

King, Gary. 1989. "Variance Specification in Event Count Models: From a Restrictive Assumptions to a Generalized Estimator." *American Journal of Political Science* 33(3):762–784.

King, Gary. 1994. "COUNT: A Program for Estimating Event Count and Duration Regressions, Versions 1988-94." http://gking.harvard.edu.

Lewis, David E., and James Michael Strine. 1996. "What Time Is It? The Use of Power in Four Different Types of Presidential Time." *Journal of Politics* 58(3):682–706.

Light, Paul. 1982. *The President's Agenda: Domestic Policy Choice from Kennedy to Carter*. Baltimore: The Johns Hopkins University Press.

Mansfield, Edward D. 1992. "The Concentration of Capabilities and the Onset of War." *Journal of Conflict Resolution* 36(2):3–24.

McCleary, R., and R. A. Hay. 1980. *Applied Time Series Analysis for the Social Sciences*. Beverly Hills: Sage.

Neustadt, Richard. 1990. *Presidential Power and the Modern Presidents. The Politics of Leadership from Roosevelt to Reagan*. New York: Free Press.

O'Brien, Sean P. 1996. "Foreign Policy Crises and the Resort to Terrorism." *Journal of Conflict Resolution* 40(2):320–335.

Pollins, Brian. 1996. "Global Political Order, Economic Change, and Armed Conflict: Coevolving Systems and the Use of Force." *American Political Science Review* 90(1):103–117.

Reed, D. 1978. "*Whistlestop: A Community Alternative for Crime Prevention*," Unpublished Ph.D. thesis. Evanston, IL: Department of Sociology, Northwestern University.

Rossiter, Clinton. 1960. *The American Presidency*. New York: Mentor.

Sayrs, Lois W. 1992. "The Effect of Provocation on Foreign Policy Response: A Test of the Matching Hypothesis." *International Interactions* 18(2):85–100.

Senese, Paul D. 1997. "Costs and Demands: International Sources of Dispute Challanges and reciprocation." *Journal of Conflict Resolution* 41(3):407–427.

Skowronek, Stephen. 1993. *The Politics Presidents Make: Leadership from John Adams to George Bush*. Cambridge, MA: Belknap Press.

Spriggs, James F., and Paul J. Wahlbeck. 1995. "Calling It Quits: Strategic Retirement on the Federal Courts of Appeals, 1893-1991." *Political Research Quarterly* 48(3):573–597.

Wayne, Stephen J. 1978. *The Legislative Presidency*. New York: Harper and Row.

White, Halbert. 1994. *Estimation, Inference and Specification Analysis*. Cambridge: Cambridge University Press.