Dynamic Modeling for Persistent Event-Count Time Series

Patrick T. Brandt Indiana University
John T. Williams Indiana University
Benjamin O. Fordham University at Albany, State University of New York
Brian Pollins Ohio State University

We present a method for estimating event-count models when the data is generated from a persistent time-series process. A Kalman filter is used to estimate a Poisson exponentially weighted moving average (PEWMA) model. The model is compared to extant methods (Poisson regression, negative binomial regression, and ARIMA models). Using Monte Carlo experiments, we demonstrate that the PEWMA provides significant improvements in efficiency. As an example, we present an analysis of Pollins (1996) models of long cycles in international relations.

Scholars use event-count models to analyze a wide variety of political science data. For example, the large sets of data on cooperative and conflictual international events often use event-count models (Huang, Kim, and Wu 1992; Sayrs 1992; Volgy and Imwalle 1995). These data have also been analyzed using time-series methods, particularly when transformed into indices of conflict and cooperation that consider the severity of each event and are no longer purely event counts (Goldstein 1991; Goldstein and Freeman 1990; Schneider, Widmer, and Ruloff 1993; Ward and Rajmaira 1992). Other event counts commonly analyzed in international relations include the number of wars (Benoit 1996; Mansfield 1992), militarized interstate disputes (Gowa 1998; Pollins 1996; Senese 1997), or other incidents of conflict and cooperation between states (Brophy-Baerman and Conybeare 1994; Eyerman and Hart 1996; Kinsella 1995; Kinsella and Tillema 1995; O'Brien 1996; Fordham 1998a, 1998b; Remmer 1998). Scholars in American politics have also employed event-count models for time-series data to analyze presidential activity (Brace and Hinckley 1993), Federal Reserve decisions (Krause 1994), federal contract awards (Mayer 1995), executive orders (Krause and Cohen 1997), and Supreme Court decisions (Caldeira and Zorn 1998). The same issues confront the smaller number of comparativists using such data (Moore, Lindstrom, and O'Regan 1996).

Patick T. Brandt is a Visiting Lecturer in Political Science, Indiana University, Woodburn Hall 210, Bloomington, IN 47405 (pbrandt@indiana.edu). John T. Williams is Professor of Political Science, Indiana University, Woodburn Hall 210, Bloomington, IN 47405 (jotwilli@indiana.edu). Benjamin O. Fordham is Assistant Professor of Political Science, University at Albany, State University of New York, 135 Western Avenue, Albany, NY 12222 (fordham@cnsvax.albany.edu). Brian Pollins is Associate Professor of Political Science, The Ohio State University, Columbus, OH 43201 (pollins.1@osu.edu).

An earlier version of this work was presented at the 1998 Summer Meetings of the Political Methodology Group, University of California, San Diego. Williams's research was supported by National Science Foundation grant SBR-9422645. Pollins's research was supported by the Mershon Center of the Ohio State University. All GAUSS code used for this analysis is available with documentation at http://www.polsci.indiana.edu/jotwilli/home/default.htm. We wish to thank Chris Achen, Neal Beck, Michael Craw, John Freeman, Simon Jackman, Gary King, David LeBlang, Michael McGinnis, Walter Mebane, Stephanie Sillay, Jim Stimson, and Pravin Trivedi for their comments and suggestions. Standard disclaimers apply.

American Journal of Political Science, Vol. 44, No. 4, October 2000, Pp. 823–843

We argue that a Poisson Exponentially Weighted Moving Average (PEWMA) model, based on Harvey and Fernandes (1989), is a useful and tractable approach to event-count time-series data that are persistent. The model assumes an underlying latent dynamic process that is continuous, and this continuous process is mapped onto the observed discrete variable. Interpretation of the dynamics is simple. The mean varies over time, and a key parameter describes the amount of variation over time in this mean. The impact of covariates is even simpler to interpret. These have the same exponential interpretation as do coefficients in the static Poisson regression model. We think that PEWMA will be a useful complement to the tool kit of political scientists who analyze event-count data.

In the next section, we review a number of models for persistent event-count time series and argue why the PEWMA is preferable for persistent event-count time series. We then outline the PEWMA model for persistent event counts.[1] The model is based on structural time-series models, as described in Harvey (1989). This structural time-series (or state space) model for event-count data is based on the representation similar to Harvey and Fernandes (1989), and Harvey (1989). We then use Monte Carlo simulations to show that this dynamic event-count model performs *much* better on efficiency grounds than do Poisson and negative binomial models. Finally, we present an application of the PEWMA model by revisiting Pollins's (1996) analysis of long-cycle theories in international relations in order to illustrate how our model can produce better inferences with persistent event-count time series.

Approaches to Modeling Time-Series Event-Count Data

The standard approach to modeling event-count data is to assume that the events are generated from a Poisson density (King 1989a). The probability of observing the count y_t is given by a Poisson distribution with mean arrival rate λ:

$$\Pr(y_t \mid \lambda) = \frac{e^{-\lambda}\lambda^{y_t}}{y_t!}.$$

[1]Elsewhere, Brandt and Williams (1998) develop a model for mean-reverting or autoregressive event count time series. This model, the Poisson autoregressive model of order p (PAR(p)) is built using a method similar to the PEWMA. The PEWMA is intended for persistent or nonmean reverting data.

Using maximum-likelihood techniques, one can estimate the Poisson mean parameter λ. When $\lambda = \exp(X_t\beta)$ the result is a Poisson regression model. The Poisson regression model assumes that events are independent so the mean and variance of the model are equivalent: $E[y_t] = V[y_t] = \lambda$. When this assumption is violated, the variance is usually larger than the mean and the events are considered *overdispersed*. When the events are overdispersed, then alternative estimators have been proposed. Among these are the negative binomial and generalized event count (GEC) estimators (King 1989b).

The two main approaches to modeling time series of event counts in political science are to use Gaussian ARIMA models, or to include a lagged-dependent event count as a regressor in the mean function of a Poisson, negative binomial, or generalized event-count regression model.

If y_t is an event count, then ARIMA models assuming normally distributed errors are flawed in four ways. First, unless the values of the observations are very large, the event-count distribution may not be accurately approximated by a normal distribution. Second, the model must produce predictions that are strictly positive to be valid. If the mean number of events is small, then the predictions may not be greater than zero—invalidating the model. Third, modeling event counts with a Gaussian distribution leads to bias and inefficiency (King 1988).

Fourth, using differencing as is common in Gaussian ARIMA modelling for nonstationary event counts is problematic. Using the first differences to deal with trend nonstationarity in count data is appealing because it is analogous to the approach used for Gaussian time series. The main problem with modeling the first differences of event counts with a Gaussian model is the validity of the distributional assumption. The difference of two *independent* Poisson variables has a known distribution. Derivations of the distribution of the difference of two independent Poisson random variables rely explicitly on this independence assumption (Skellam 1946; Strackee and van der Gon 1962). For time series of counts, this independence assumption is clearly inappropriate. In general, we do not think it is true for time series of counts, and it is never true for trending series. Thus, the lack of a well-defined distribution for the difference of a series of nonstationary event counts leads us to look for alternative models that can describe the data generation process.

An alternative approach is to retain the assumption of a Poisson or negative binomial distribution. Under this approach possible dynamics in event-count data are modelled with a lagged dependent variable in Poisson and negative binomial models. These models suffer from two possible problems. First, they fail to represent adequately the dynamics in persistent time series because these models imply that the growth rate of the process is the exponentiated coefficient on the lagged dependent variable. Such a process *may* potentially generate time-series data, but not data that are dynamic. The lagged event-count model is appropriate only for series with exponential growth rates and no dynamics.[2] This means that the lagged dependent Poisson and negative binomial models cannot be used to model stationary time series. Unless the time series of event counts has an exponential deterministic trend and no dynamics, this model has limited applicability.

Time-Series Models for Event-Count Data

While political scientists have included lags in event-count models or used ARIMA models for time series of event counts, many other models have been proposed. The literature on time-series models for event-count data contains a number of different methods for dealing with dependence in event-count time series. In this section we review these alternative approaches and defend the PEWMA as a very useful model.

In our view, there are three classes of time-series event count models.[3] The first class is integer-valued ARMA or Discrete Autoregressive Moving Average models (DARMA). The second class is based on conditional autoregressive or hidden Markov processes. These models are based on a latent variable model. The final class of models, which contains the PEWMA, is the state space or time-varying parameter models.

Integer-valued / Discrete ARMA Models

The integer-valued AR (INAR), integer-valued ARMA (INARMA), or Discrete Autoregressive Moving Average (DARMA) models discussed by Alzaid and Al-Osh (1990), Du and Li (1991), and McKenzie (1988) are based on probabilistic mixtures of different Poisson processes. For example, the integer-valued AR(p) model is based on a marginal Poisson distribution. In each period, the observed number of events is defined by a binomial or multinomial sampling procedure called "thinning." The data-generating procedure for the event-count variable y_t is of the form

$$y_t = \sum_{i=1}^{p} \alpha_i \circ y_{t-i} + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \ldots$$

where "$\circ$" is the multinomial thinning operator, and ε_t is a nonnegative integer valued variable with fixed mean and known variance.[4] This is a mixture of an event-count distribution (Poisson or negative binomial) and a multinomial distribution. Estimates of the autoregressive thinning parameters α_i can be obtained by conditional maximum likelihood. To date, we are unaware of any general implementation of this method. Squier (1996, 1997) shows that including exogenous variables in this model is difficult and produces inconsistent and inefficient estimates of the effects of exogenous variables. In addition, if covariates are included in this mixture model, it is not clear how to interpret the effects of covariates on the observed event-count time series.

Conditional Autoregressive/ Hidden Markov Models

Conditional autoregressive models are based on assuming that the conditional distribution of the events y_t, depends on a static mean function, and lagged-dependent variables or serially correlated errors. For these models, the dynamics enter either as past y_ts or through the error process. Models in this class have been discussed by Jackman (1998), Shephard (1995), Diggle, Liang, and Zeger (1994), and Zeger (1988).

As an example, consider a static regression model where the number of events y_t depends on a set of covariates and a serially correlated error term that follows an AR(1) process. Then the conditional mean of the number of events can be written as,

$$E[y_t | \varepsilon_t] = X_t \beta + u_t, \quad u_t = \rho u_{t-1} + \varepsilon_t,$$

where ε_t is i.i.d.$(0,\sigma^2)$. This model can be estimated by nonlinear least squares, or maximum-likelihood

[2] The growth rate of this lagged Poisson regression model is given by

$$\ln(\mu_t) - \ln(\mu_{t-1}) = X_t\delta - X_{t-1}\delta + \rho z_{t-1} - \rho z_{t-2}.$$

Taking expectations gives

$$E[\ln(\mu_t) - \ln(\mu_{t-1})] = \rho E[z_{t-1} - z_{t-2}]$$

If $\rho \neq 0$ and $E[z_{t-1} - z_{t-2}] \neq 0$, this model implies a nonzero growth rate for the conditional mean.

[3] See Davis, Dunsmuir, and Wang (1998a) and Cameron and Trivedi (1998, chapter 7) for alternative reviews of the literature.

[4] The thinning operation is defined as follows for the binomial case: $\alpha \circ X_n$ is the sum of X_{n-1} independent random Bernoulli variables, Y_i^{n-1} where $\Pr(Y_i^{n-1} = 1) = 1 - \Pr(Y_i^{n-1} = 0) = \alpha$. This process gives rise to the same AR(p) coefficients and autocorrelation function behavior as the standard Gaussian AR(p) model (Alzaid and Al-Osh, 1990).

methods depending on the assumptions about the distribution of ε_t.

Alternatively, Zeger (1988), and Diggle, Liang, and Zeger (1994) propose a latent-variable specification based on:

$$E[y_t|\varepsilon_t] = \exp(X_t\beta)\varepsilon_t, \quad Var[y_t|\varepsilon_t] = u_t.$$

The estimator based on this model is identified if $E(\varepsilon_t) = 1$ and $Cov(\varepsilon_t\varepsilon_{t+\tau}) = \sigma^2\rho_\varepsilon(\tau)$. This specification of the latent-variable model can be estimated using a quasi-maximum-likelihood estimator that solves a set of generalized estimating equations (see Zeger 1988 for details). The standard Poisson regression model provides consistent estimates of the regression parameters for this model, even in the presence of a dynamic latent process (Davis, Dunsmuir, and Wang, 1998a). However, parameter estimates will be inefficient when there is a significant amount of serial correlation in ε_t (Davis, Dunsmuir, and Wang 1998a, 17). Thus, inference based on these models is open to the critique that the interpretation of estimated parameters is imprecise.[5]

These examples with serially correlated errors are actually a special case, since a model with AR errors can be written as a moving average model for the dependent variable. The serial correlation in the errors imply that y_t has a moving average representation. An AR(1) error model suffers from the same limitations as a Gaussian model with AR(1) errors. Models with AR(1) error structures may be appropriate in specific contexts, but they are not general since they fit only a limited type of dynamic processes. In addition, if lagged dependent counts are included in the model and there is still serial correlation in the error process, estimates will be inconsistent, just as in OLS models (Cameron and Trivedi, 1998: 227). Thus, inclusion of lagged counts, or serially correlated errors appears to be just as difficult in event count models as in standard Gaussian linear time-series models. Such serial correlation can be analyzed using standardized residuals and standard diagnostics for serial correlation (see Cameron and Trivedi, 1998: 228–229 for details). However, if data are nonstationary, models that assume stationarity will clearly be inappropriate. The next class of models relaxes this condition and allows for models that can be either stationary with long memory or nonstationary.

State-space/Time-varying Models

The final class of models is based on a state-space representation specifying a measurement and transition equation. For event counts, the measurement equation is an event-count distribution, while the transition equation describes how the dynamics of the event count distribution parameters evolve. The state-space approach is very general because the models are specified in terms of the basic structure of the time series: trends, cycles, and time-varying stochastic components. Examples of this approach include Zeger and Qaqish (1988), Harvey (1989), Harvey and Fernandes (1989), West, Harrison, and Migon (1985), West and Harrison (1986, 1997), Chan and Ledolter (1995), Kitagawa and Gersch (1996), Durbin and Koopman (1997, 2000), Davis, Dunsmuir, and Wang (1998a, 1998b), Jorgensen, Labouriau, and Lundbye-Christensen (1996), and Jorgensen et al. (1999).

As an example, Davis, Dunsmuir, and Wang (1998a, 1998b) present a general state-space event count time-series model. In their model, the conditional mean, or state variable, for the event-count model incorporates both autoregressive (AR) and moving average (MA) components. The conditional mean of the marginal Poisson distributed event counts has the form,

$$E\left[y_t \mid Y_{t-1}, X_t, e_t\right] = \mu_t = \exp\left(X_t\beta + \sum_{i=1}^{\infty}\tau_i e_{t-i}\right),$$
$$where \;\; e_t = \frac{y_t - \mu_t}{\sqrt{\mu_t}},$$
$$and \;\; \sum_{i=1}^{\infty}\tau_i e_{t-1} = \frac{\left(1+\sum_{i=1}^{p}\gamma_i e_i\right)}{\left(1-\sum_{i=1}^{q}\phi_i e_i\right)} - 1 = \frac{\gamma(e)}{\phi(e)} - 1.$$

Thus, event counts are Poisson distributed with mean μ_t. The $\sum_{i=1}^{\infty}\tau_i e_{t-i}$ is the one-step-ahead predictor of e_t based on an ARMA(p,q) process. While this model is more general than what we present below, using higher order dynamic terms in an ARMA filter primarily only benefits the forecasting ability of the model. The state-space models of Durbin and Koopman (1996, 1997) are

[5] Jackman (1998) proposes another "latent first order autoregressive model." The latent mean for the event count y_t has the form,

$$E[y_t] = \exp(X_t\beta + \varepsilon_t)$$
$$where \quad \varepsilon_t = \rho\varepsilon_{t-1} + v_t$$
$$and \quad v_t \sim N(0,\sigma_v^2),\; t = 1,2,\ldots,T.$$

Note the similarity to Zeger's GEE model. This model is identified by a distributional assumption on the AR(1) errors and corresponds to a Gaussian regression model with AR(1) errors. Since the model's parameters cannot be simultaneously estimated, Jackman uses the EM-algorithm via a Markov Chain Monte Carlo method to compute the maximum-likelihood estimates. This model and estimation issues are also discussed in Davis, Dunsmuir, and Wang (1998a).

similar to these. Estimates can be obtained using Markov-Chain-Monte-Carlo methods or other numerical solutions. While these models appear to offer an intuitive approach that is similar to existing ARIMA models, the complexity of their estimation and high degree of parameterization make inference difficult.

The state-space models of Chan and Ledolter (1995), Kitagawa and Gersch (1996), West, Harrison, and Migon (1985), and West and Harrison (1986, 1997) are based on Bayesian dynamic linear models. While these are similar in spirit to the PEWMA we present, they are based on Bayesian rather than classical assumptions. These models typically require the use of numerical methods such as Markov-Chain-Monte-Carlo methods that are highly dependent on the specific model being estimated. Moreover, the interpretation of exogenous covariates in these models typically requires dynamic simulations. These requirements make implementing and interpreting these models costly and could put them out of the reach of most political scientists. An alternative approach discussed by Jorgensen, Labouriau, and Lundbye-Christensen (1996) and Jorgensen et al. (1999) is based on a classical Kalman filter/smoother, but is more complex to estimate and interpret than our present approach.

There are several drawbacks to these alternatives. First, there exists little available software for easily implementing these models. Second, the interpretation of these models is often complex and requires techniques beyond those currently used to evaluate event-count data. Finally, several of the models produce poorly behaved or hard to interpret estimates of the effects of exogenous covariates. This complicates inference and testing.

Our PEWMA state-space model suffers from none of these problems and has several advantages. First, estimation is simple and fast. Estimation is done using a version of the Kalman filter rather than simulation methods, which is robust and quick. For example, all of the PEWMA estimates in the six specifications in our example in Section Four take less than thirty seconds on a standard PC workstation using our software.[6] Second, unlike some of the alternative time-series event count models, the interpretation of the effect of exogenous variables does not require any new methods or simulation techniques. The standard Poisson regression model is a special case of the PEWMA, and we can interpret the PEWMA coefficients just as we would any coefficient in a standard Poisson or negative binomial regression model. Finally, the dynamics in the PEWMA model are characterized by a single parameter that takes on values between zero and one. This allows for a simple and intuitive evaluation of the model dynamics, unlike alternative approaches.

The Poisson Exponentially Weighted Moving Average Model

The Model

To model persistent event count time series, we use a structural time-series model. The intuition behind the model is that the mean number of observed events today is a weighted sum of all the past events. This mean is defined as a weighted average of the past events, plus a random shock. The random shock is included to account for unexpected changes in the mean number of events. The model discounts events in the distant past more heavily than those in the recent past. If the number of events observed at each period in time is independent of past events, the model reduces to the standard Poisson regression.

Our model has two components. The first is a function that describes how the observed number of events arises as a function of a mean number of events in the past. It is called a measurement, or system equation in a structural time-series model. The measurement equation describes the process that generates the observed data. The second function describes how important past events are for predicting the number of events in the current period. It is called the state or transition equation and describes the dynamic transition process from events in the past to events in the present.

Once we have a model in state-space form, estimation of the model's parameters proceeds using the Kalman filter. Using a recursive algorithm, the Kalman filter computes the optimal estimates of the mean and other parameters of the state equation of a state-space model at each time period using the information available up to that time period. The Kalman filter works by finding the estimates of the model parameters in the state equation that minimize the mean-squared error of the conditional mean of the series (Harvey 1989, 104–105).

To identify and estimate the model we adopt the approach of Harvey and Fernandes (1989). They employ natural conjugate densities to simplify the development of the model and numerical calculations.[7] Using natural

[6] This result is based on using our software code in Gauss 3.2 with MAXLIK 4.0.34 on a Pentium II workstation with 256 megabytes of memory.

[7] A conjugate density is a prior distribution that after being combined with a likelihood function yields a posterior distribution of the same form. A conjugate density is a natural conjugate density if it is in the class of distributions with the same functional form as

conjugate densities to describe the unobserved parameters of the model allows us to derive well-known closed-form distributions for the model parameters.

The model for time-series count data is based on a state-space form with a Poisson measurement equation and a gamma-distributed state equation. The model captures the changes in the mean of a Poisson process at time t. We assume that the count at time t, denoted y_t, follows a Poisson distribution with a conditional mean at time t denoted μ_t, with explicit dependence on t. The variable μ_t is assumed to be gamma distributed, since the gamma is the conjugate distribution for the Poisson. With these assumptions, one can then derive the resulting conditional forecast or predictive distribution for the mean. In the estimation of the mean of μ_t, one needs to account for the history of the process up to and including period $t-1$. We denote this sequence of conditioning data by the vector Y_{t-1}. This vector contains both past-observed values of counts y_t, and any independent variables X_t that have been observed up to and including period $t-1$. The vector, $Y_{t-1} = (y_0, y_1, \ldots, y_{t-1}; X_0, X_1, \ldots X_{t-1})$, is the full information set available at time t. In addition, the model also contains a hyperparameter $0 < \omega \leq 1$. This parameter allows the past observations to be discounted in making conditional forecasts of the mean of future observations.[8] Throughout the article, we use the notation $w_{t|t-1}$ to represent the value of the random variable w_t conditional on the observed-information set in the previous $t-1$ periods.

The Poisson-gamma exponentially weighted moving average model for count data (PEWMA) is built around the following three assumptions that characterize the mean and dynamics of the process:

(1) *Measurement Equation*: The observed counts at time t are drawn from a Poisson marginal distribution,

$$\Pr(y_t \mid \mu_t) = \frac{\mu_t^{y_t} e^{-\mu_t}}{y_t!}. \tag{1}$$

The parameter μ_t in this distribution is the unobserved mean-arrival rate for the count at time t. This unobserved mean μ_t is parameterized by the multiplicative equation

$$\mu_t = \mu^*_{t-1} \exp(X_t \delta), \tag{2}$$

where δ is a $K \times 1$ vector of coefficients, X_t a $1 \times K$ vector of explanatory variables (without a constant), and a time-varying component μ^*_{t-1}.[9] As in Harvey and Fernandes (1989), we assume that this separate time-varying level component μ^*_{t-1} is a multiplicative factor. This factor is estimated by the Kalman filter and accounts for the observed counts prior to time $t-1$. Thus, it is a smoothed mean of the previous observations.

(2) *Transition Equation*: The stochastic mechanism for the transition in the series from time $t-1$ to time t is a function of μ_{t-1} and μ_t. The dynamics of the mean are described by a multiplicative transition equation with the form,

$$\mu_t = e^{r_t} \mu_{t-1} \eta_t, \quad t = 1, 2, \ldots, T, \tag{3}$$

where η_t is beta distributed, $\beta(\omega a_{t-1}, (1-\omega) a_{t-1})$. The parameter ω captures the discounting of the observations in computing the mean and η_t and r_t parameterize the growth rate in period t. The beta-distributed variable η_t captures the proportional stochastic shift in the mean from time $t-1$ to time t. From the properties of the beta distribution, $E[\eta_t] = \omega$ for all t. This means that the expected stochastic shocks on average are equal to the weight of past event counts in the current event count. The parameter r_t describes the growth in the series and insures that $\mu_t > 0$.[10]

(3) *Conjugate Prior*: The prior distribution for the time varying component is a gamma distribution: $\mu^*_{t-1} \sim \Gamma(a_{t-1}, b_{t-1})$. To identify the model in equations (1–3), we specify the gamma distribution as the conjugate prior for the distribution of y_t. The gamma density f is given by

the likelihood function. For example, if y is sampled from a Poisson distribution with unknown mean λ, then the natural conjugate prior that describes λ is the gamma distribution. If y is sampled from a normal distribution with unknown mean μ and known variance σ, then the natural conjugate prior is a normal distribution (DeGroot 1970, 164,167).

[8]A hyperparameter is a parameter of a prior distribution not fixed at particular numerical values (Gelman et al. 1995, 36).

[9]Including a constant in the mean function is equivalent to specifying a deterministic trend.

[10]We caution readers who might use the Harvey and Fernandes (1989) model. Their transition equation provides an incorrect model of the stochastic process for the mean of event counts. Harvey and Fernandes adopted an alternative transition equation with the form $\mu_t = \mu_{t-1}\eta_t\omega^{-1}$. However, Shephard (1994) citing Nelson (1990) notes that if $\omega < 1$, then μ_t converges to zero as $t \to \infty$. The reason for this is that the growth rate for the Harvey and Fernandes' transition equation can be approximated by $\ln\left(\frac{\mu_t}{\mu_{t-1}}\right) = \ln(\eta_t) - \ln(\omega)$. By Jensen's inequality, this quantity will be negative on average. In their discussion of Harvey and Fernandes' model, Brockwell and Davis (1996) also note this problem, citing Grunwald, Hamza, and Hyndman (1997). We avoid this problem by adopting the transition equation suggested by Shephard's (1994) work on local scale models. So far as we know, this is the first application of Shephard's transition equation to a state-space model for event count data.

$$f(\mu; a, b) = \frac{e^{-b\mu}\mu^{a-1}b^{a}}{\Gamma(a)}, \tag{4}$$

with $a = a_{t-1}$, $b = b_{t-1}$, and $\mu = \mu^{*}_{t-1}$. These values are computed from the previous $t-1$ observations, Y_{t-1}.

Using these assumptions, we can estimate a Kalman filter using maximum-likelihood methods. The following sections describe the estimation and interpretation of the model.

Estimation

Estimation of the model requires a time-dependent series of recursions for the conditional mean $E[\mu_t | Y_{t-1}]$ and the posterior mean $E[\mu_t | Y_t]$. These recursions define a Kalman filter for the Poisson measurement density and gamma-distributed transition equation. To compute this filter, we derive the recursions for the values of a, b, and μ all evaluated at time $t-1$ (the prior), at time t given $t-1$ (the conditional or predictive), and at time t (the posterior). These estimates are then used to compute the conditional or predictive distribution of the number of events in each time period. The predictive distribution describes the probability of observing events at time t conditional on all the events observed in each of the previous $t-1$ periods. It turns out that the predictive distribution for the PEWMA model is a negative binomial distribution where the overdispersion in the observed data depends on the amount of dependence between the events in each time period. The details of these derivations are included in Appendix A. Rather than focus on the computations of these latent parameters of the model, we turn to estimation and properties.

The predictive distribution combines the information in the measurement density and the transition equation. That is, we observe events y_t conditional on the mean μ_t at time t. The value of μ_t then depends on the weighted average of the number of events observed in the past.

The joint-predictive density, conditional on Y_τ for observations $y_{\tau+1}, \ldots, y_T$ is:

$$\begin{aligned}\Pr\left(y_{\tau+1},\ldots,y_T\right) &= \prod_{t=\tau+1}^{T} \Pr\left(y_t \mid Y_{t-1}\right) \\ &= \prod_{t=\tau+1}^{T} \int_0^{\infty} \underbrace{\Pr\left(Y_t \mid \mu_t\right)}_{Measurement} \cdot \underbrace{\Pr\left(\mu_t \mid Y_{t-1}\right)}_{Transition}.\end{aligned} \tag{5}$$

The key to understanding this predictive distribution is the decomposition of the posterior probability $\Pr(y_t | Y_{t-1})$. Using the Kalman filter, we can condition the observed data on the components of the predictive distribution: the probability of the event (from a Poisson distribution) and the probability of the unobserved mean (from a gamma distribution). Note that these two components correspond to the measurement and transition equations.

Once we have the predictive distribution, we construct the log-likelihood function and estimate the model using maximum-likelihood techniques. This approach is identical to that used to derive the estimates of a Gaussian ARIMA model.

Given the conditional probability density of $y_t | Y_{t-1}$, we can construct the log-likelihood function for the unknown hyperparameter and the parameters based on (5):

$$\begin{aligned}&\ln\ L\left(\omega, \delta \mid a_{t-1}, b_{t-1}, y_t, X_t\right) \\ &\quad = \sum_{t=\tau+1}^{T} \ln\Gamma\left(y_t + \omega a_{t-1}\right) - \ln\left(y_t!\right) \\ &\quad\quad - \ln\Gamma\left(\omega a_{t-1}\right) + \omega a_{t-1} \ln\left(\omega \mathrm{b}_{t-1} \exp\left(-X_t\delta - r_t\right)\right) \\ &\quad\quad - \left(\omega a_{t-1} + y_t\right) \ln\left(1 + \omega b_{t-1} \exp\left(-X_t\delta - r_t\right)\right).\end{aligned} \tag{6}$$

Maximizing with respect to ω and δ provides an estimate of the hyperparameter and other parameters, respectively. Note that the computation of this likelihood requires an analytic solution for r_t, which characterizes the per-period growth rate from period $t-1$ to t. Appendix A presents the derivation of this quantity. Appendix B describes the forecast function for the model.

Description of the PEWMA Process

The PEWMA model does not have a simple linear structure. Thus, it is less than obvious how the dynamics of the PEWMA process evolve over time. In this section we characterize the evolution of a PEWMA time series by analyzing the properties of the transition equation (3). Our characterization shows how the mean of the process, μ_t evolves over time.

The transition equation for the PEWMA implies that the mean evolves according to a random-walk process. Recall that the transition equation is written

$$\begin{aligned}\mu_t &= e^{r_t}\mu_{t-1}\eta_t, \\ so\ \ln\left(\mu_t\right) - \ln\left(\mu_{t-1}\right) &= r_t + \ln\left(\eta_t\right),\end{aligned} \tag{7}$$

with μ_t is the conditional mean, r_t describes the per-period growth rate, and η_t are the errors. The left-hand side of this equation estimates the growth rate of the series of event counts and is zero in expectation, since we constructed $E[\mu_{t-1} | Y_{t-1}] = E[\mu_t | Y_{t-1}]$ (see Appendix A).

Thus, the expected growth rate is

$$0 = E[r_t] + E[\ln(\eta_t)], \tag{8}$$

so the transition equation implies that the mean growth rate is zero.

Even though the expected growth rate is zero, in finite samples local stochastic trends can be modeled as a PEWMA. Since in most cases trends in event counts will be local rather than global, the PEWMA will be appropriate in almost all cases where event counts are persistent.[11]

How does the mean of the PEWMA, μ_t evolve over time? We answer this question by evaluating a recursion of the natural log of the transition equation. Shephard (1994) shows that gamma-distributed transition equations of the form (3) follow a random walk in their natural logarithm. Starting at period zero, repeated substitutions show:

$$\begin{aligned}
\ln \mu_1 &= r_1 + \ln \mu_0 + \ln \eta_1, \\
\ln \mu_2 &= r_2 + r_1 + \ln \mu_0 + \ln \eta_1 + \ln \eta_2, \\
&\dots \\
\ln \mu_t &= \ln \mu_0 + \sum_{j=1}^{t} \left(\ln \eta_j - r_j \right).
\end{aligned}$$

This final equation is a random walk. This is a general result for multiplicative gamma transitions of the form in (3).

At each time period, the mean number of counts, conditional on all the past observations, is a weighted sum of all the past counts. The weights are reflected by the hyperparameter term ω, which is the discount rate for the past observations. Smaller values of ω imply less discounting, so more recent observations matter more and the series demonstrates significant persistence. Larger values imply that all past observations matter less, and the effect of history decays rapidly. When $\omega = 1$, then a constant mean describes the process. These discounted time-varying means are computed via the Kalman filter.

The parameters r_t and η_t capture the stochastic or random effects over time. These terms describe the per-period growth rate of the conditional mean μ_t. The per-period growth rate is the change in the mean number of counts in the period, $t - 1$ to t. The beta-distributed error term fulfills two functions. First, it is parameterized in terms of a_{t-1} (the location parameter of the mean) and ω to maintain conjugacy so that we can carry out later computations. Second, it describes the degree of discounting over time, since $E[\eta_t] = \omega$.

The covariates X_t enter the model contemporaneous with the level of the time series at $t - 1$, namely through the μ^*_{t-1} term. The effect of the covariates is the same as in the standard Poisson regression model. Conditional on the mean of the series at time $t - 1$, the effect of a one-unit change in X_t is given by δ. Since X_t and δ are modelled using an exponential link function, the effect of a one-unit change in X_t is a $100(\exp(\delta) - 1)$ percent change in y_t, just as in the Poisson model.

Intuitively, the PEWMA process is one where the mean changes over time. The mean at any point in time will be a weighted average of past events plus a function of the explanatory variables. The coefficient ω determines the amount of variance in the mean. If ω is one, then there is no movement in the mean, and the PEWMA is equivalent to the standard Poisson regression. As w becomes smaller, there is greater variance in the mean.

Identification

We would like to be able to identify when to use the PEWMA model. Grunwald et al. (1997) show that standard autocorrelation function (ACF) computations can be used to diagnose a linear autoregression process for event counts. This is the case for the PEWMA as well, where we would expect the ACF to display persistence over many lags.

The characteristics of the PEWMA process are demonstrated in Figure 1. Figure 1 presents a simulated PEWMA series as well as some diagnostic plots for the series. The PEWMA series in the figure was generated using the assumptions that $\mu_0 = 3$, $X_t \sim N(0,1)$, $\delta = 0.5$, and $\omega = 0.4$. The first graph in the figure shows the actual series. The second graph shows the difference, $\ln \mu_t - \ln \mu_{t-1}$. The third graph shows the autocorrelation function for the series. The final graph shows the ACF for the log difference of the mean, μ_t.

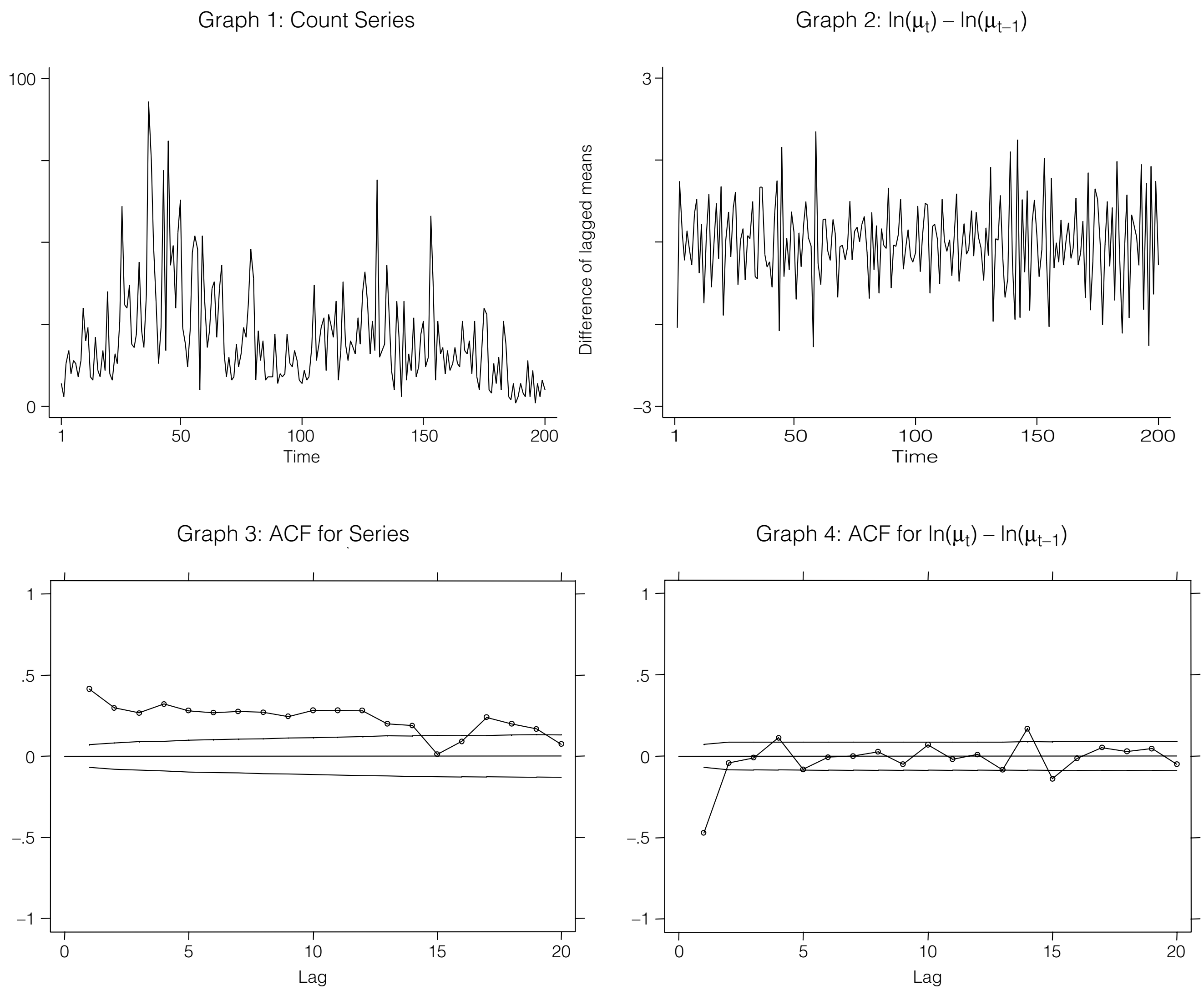
Notice that the ACF for the simulated PEWMA series shows a large degree of dependence. Once the local mean $\ln(\mu_t)$ is differenced, we see that it looks like a stationary process (Graph 2). The ACF for the differenced series has only one significant negative lag. This is consistent with the moving-average process that generates the state variable μ_t.

Monte Carlo Experiments

We motivated the PEWMA model by arguing that existing event count and ARIMA models could not ad-

[11]We conjecture that events will not be explosive because with fixed time slices, only so many events can happen. For example, only so many riots can happen in any given month. Thus, the PEWMA is much more appropriate for modeling persistent data than is a lagged-dependent regressor model, the latter requiring the number of events to increase forever.

Figure 1 Sample PEWMA Process, Differenced Log Means, and Autocorrelation Functions



equately model data that is generated from dynamic event-count process. However, it could be that existing methods such as Poisson or negative binomial regression do an adequate job in estimating the coefficients of covariates even when the dynamics are left unspecified.

To evaluate the implications of using alternative models when the data actually follow a PEWMA process, we conducted a series of Monte Carlo experiments. In these experiments we want to determine the amount of inefficiency that could arise if one were to use standard Poisson, negative binomial, or Gaussian models when in fact the true data-generation process is the PEWMA. We know *a priori* that the PEWMA model will be more efficient, as it is the true model. What is unclear is whether the efficiency losses are large when using the wrong model.

Our Monte Carlo design includes a series of twenty-four experiments, each with 200 replications. We conduct experiments based on data generated from the PEWMA model with a single fixed normally distributed covariate, $X_t \sim N(0,1)$. We vary the mean by choosing different priors for the gamma distribution that initializes the series. We chose values of a_0 and b_0 to yield three different values of $\mu_0 = 10, 20, 50$.[12] We also investigated

[12] The complete DGP is specified by initializing the process, and we allow the mean to be determined by the latent variable μ_{t-1} at time zero. Initialization is accomplished by specifying values of a_0 and b_0 so that $\mu_0 = \frac{a_o}{b_0}$. The value of μ_1 can then be computed based on the vector X_t and the initialized values, a_0, b_0, μ_0. As the filter runs through time, then new values of r_t and η_t are computed based on the realized values of y_t and the values of a_{t-1} and b_{t-1}. Note that this process is NOT equivalent to drawing a vector μ as a random walk process and then drawing a vector of Poisson distributed variables such that $y_t \sim Po(\mu_t)$, $t = 1, \ldots T$. Such a process fails to account for the past realizations of $Y_{t-1} = (y_0, y_1, \ldots y_{t-1})$.

sample sizes of $T = 50, 100, 200$. Finally, we varied the dynamics so that $\omega = 0.4, 0.6, 0.8$. The coefficient on the single covariate, δ_1 was fixed at 0.5 in all the experiments. The mean number of counts and sample sizes are reflective of data that is typically analyzed in political science applications.[13]

Once we generated the data for each replication, we estimated seven different models. The models employ different assumptions about both the dynamics and distribution of the data. As a benchmark, we estimate the PEWMA (the true model). In addition, we also estimated two Gaussian models using the natural log of the counts to assess the effects of distributional misspecification. The first Gaussian model has an AR(1) error process and is the well-known generalized least squares (GLS) model, which is an ARIMA(0,0,1) process. We estimate this GLS model using the maximum-likelihood grid search for r suggested by Hildreth and Lu (1960). The second model is an OLS model with a lagged-dependent variable. We refer to this model as logged-lagged OLS (LLOLS).[14]

The remaining four models are well-known Poisson regression and negative binomial regression models. For the Poisson regression and negative binomial regression, we posit two different mean functions:

$$\lambda = \exp\left(\delta_0 + \delta_1 X_t\right)$$
$$\textit{and}\, \lambda_t = \exp\left(\delta_0 + \delta_1 X_t + \rho y_{t-1}\right).$$

Models based on the second mean function are referred to as lagged Poisson regression or negative binomial. The negative binomial is estimated using the parameterization suggested by King (1989a).

The Monte Carlo results confirm that the PEWMA is more efficient than the other models for estimates of the single-regression parameter. This should be obvious for two reasons. First, it is the true model and should be more efficient than the rival estimators. Second, estimates from nonlinear and linear models that fail to account for some pattern of serial correlation produce standard errors that are incorrect. Thus, we are interested in two other issues. First, what is the relative efficiency of the PEWMA? Second, are the differences in the estimated covariances so large as to affect inference?

Figure 2 demonstrates the efficiency losses that arise when one uses a standard event count or Gaussian model for persistent event-count data such as that from the PEWMA model. In this figure we show the relative efficiency of the alternative models to the PEWMA model.[15] A relative efficiency value of one indicates that the two estimates are identical (i.e., the alternative estimator is as efficient as the PEWMA). Relative efficiency values less than one indicates that the alternative model is more efficient than the PEWMA. Values greater than 1 indicate that the PEWMA is more efficient. The PEWMA data generation process is the true model.

Each cell of Figure 2 shows the relative efficiency of the various estimators for given value of ω and μ_0 for each of the sample sizes. Note that for a fixed value of μ_0 and w, as the sample size increases the alternative estimators become less efficient relative to the PEWMA. As expected, the Gaussian models are grossly inefficient. Even in large samples with large mean values of counts (the case in which the Poisson model converges to normality), these models have standard deviations that are between 1.5 and 10 times larger than the PEWMA Monte Carlo standard deviations. The same can be said for the Poisson, lagged Poisson, negative binomial, and lagged negative binomial.[16]

The inclusion of a lagged-dependent variable offers little help in reducing the effect of the conditional mean misspecification for the relative standard errors. The

[13]For values of w near zero, the dynamics become so strong that the data-generating process is nonstationary. In large samples (i.e., $T > 50$) the PEWMA model will diverge to infinity for ω near 0. For small means ($\mu_t < 20$), the PEWMA series reaches an absorbing state (zero) as ω approaches 0. Our experiments do not include long series (e.g., $T = 200$, $\mu_0 = 10$) with very small mean counts, or series with large means and explosiveness. This is not a function of the transition equation, since the dynamics of the series are well defined in these cases. Rather, these experiments present a difficult data-generation problem because they are not very likely to happen in real data. In fact, few "real" datasets that we have analyzed with the PEWMA model have values of $\omega < 0.4$.

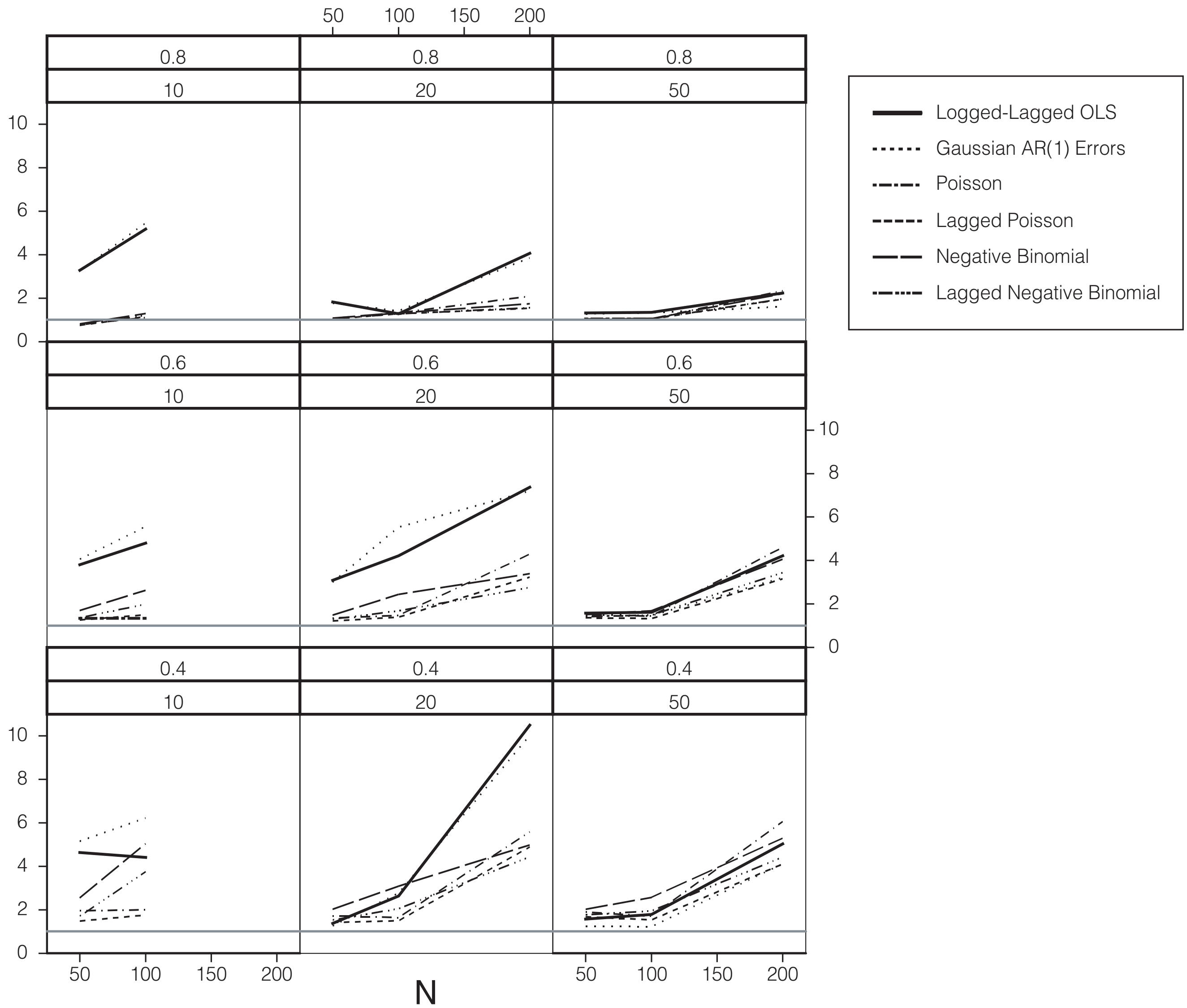
[14]For both Gaussian models, we face the problem of taking ln (y_t) when y_t is zero. To deal with these cases, we add a small positive constant (0.001) to the observed event counts. While this leads to bias and inefficiency (King 1988), it is a common practice.

[15] The relative efficiency is computed as

$$\text{Relative Efficiency} = \frac{\text{Monte Carlo standard deviation for } \delta_1}{\text{PEWMA Monte Carlo standard deviation for } \delta_1}.$$

[16]If the standard errors reported by the models suffer from serial correlation, then we would like to gauge how far from the true values the reported estimates are. One way to measure the degree to which the estimated standard errors produced by the Poisson and negative binomial models are incorrect is to compute the relative overconfidence of the standard errors (Beck and Katz 1996). The relative overconfidence is based on the ratio of the Monte Carlo standard deviations to the estimated (Huber-White) standard errors for each estimator. In samples where $T = 50$, the alternative models' reported robust standard errors are nearly identical to their Monte Carlo standard deviations. For $T = 100$, the relative overconfidence values indicate that the reported Huber-White standard errors will be *larger* than the true values (as estimated by the Monte Carlo standard deviations). Conversely, for $T = 200$, the relative overconfidence measure indicates that the Huber-White standard errors will be *smaller* than the true standard errors. Details are available from the authors upon request.

FIGURE 2 **Relative Efficiency of Various Alternative Estimators to the PEWMA**



Each line corresponds to a relative efficiency of the estimator to the PEWMA estimator. Each box in the figure shows the relative efficiency of the estimator (each line) to the PEWMA for a given value of ω and μ. The rows of the figure show the variation in the relative efficiency for the different values of μ for a given value of ω. The columns show the variation in relative efficiency for different values of ω for a given value of μ. We have indicated the relative efficiency value of 1 by a horizontal line in this plot. See text for discussion.

lagged Poisson and lagged negative binomial models are always more efficient than the standard Poisson and negative binomial regressions. These lagged count models are generally less efficient than the PEWMA. It is worth reiterating that using lagged-dependent counts in the negative binomial and Poisson regression models provides a poor method of accounting for time-series properties.[17] The implication of estimating incorrect standard errors with Poisson and negative binomial regressions should not be understated. The results demonstrate that the degree of inefficiency is large for a variety of sample sizes, means, and varying degree of dynamics.

[17] We also computed the Monte Carlo estimates for the dynamic and dispersion parameters: ω in the PEWMA, the autocorrelation parameter ρ for the ARIMA, LLOLS, the lagged Poisson, and lagged negative binomial models, and the dispersion parameter for the negative binomial models. The estimate of ω in the PEWMA models is almost always significant for the experiments. In almost every case, the hypothesis that the Poisson regression is the true model is rejected based on a t-test for omega (H_0 : ω = 1; H_A : ω < 1). However, there is a small upward bias in the small sample estimates of ω and this parameter should be interpreted carefully. Details of these results are available upon request.

Application: Pollins's Coevolving Systems Analysis

In order to demonstrate the important practical consequences of using the PEWMA rather than a lagged Poisson regression model (LP), we present a revision of Pollins's (1996) analysis of how variations in global political and economic activity affect armed conflict. We then reestimate Pollins's models using the PEWMA estimator. We show that the LP results reported in Pollins (1996) underestimate the standard errors of the parameters, leading to overly optimistic conclusions about several of the theoretical models he tests. We find that the PEWMA dominates the LP model in terms of both efficiency and the Akaike Information Criterion (AIC).

Our replication of Pollins's (1996) analysis of how cycles in global economic activity and the global political order affect armed conflict demonstrates that accounting for the temporal dependence in the data leads to different conclusions about hypothesis tests of the causes of international conflict. The problem with standard approaches such as the lagged Poisson regression is that serial dependence in the data lead to incorrect standard errors. This leads to incorrect inferences since Pollins's statistical tests are based on these erroneous standard errors.

The time dimension figures prominently in the theories tested by Pollins, including the new model of "Coevolving Systems" he develops out of his analysis. In his analysis, international relations theories of armed conflict and war are explained by these theories in part as a function of overarching systemic conditions tied to changes over time in the global economy and world order. Pollins notes five major theories that have been proposed to explain the patterns of armed conflict and war. Long Wave theories (Goldstein 1988, 1991) map long-term, repeated phases of growth and stagnation in the global economy. Wallerstein (1983) interprets these economic patterns as tied to the rise and fall of a single, dominant state, or "hegemon" in the system. Modelski and Thompson (1987, 1996) interpret growth and decline as tightly connected to a century-long process they term the Leadership Cycle, and Gilpin (1981) argues that the ascendance of hegemons like Great Britain in the Nineteenth Century and the United States in the mid-Twentieth Century brings a period of lower armed conflict and relative peace. The phenomena of interest to these scholars unfold over periods of fifty to a hundred years or more. Pollins (1996) notes the Long Wave and the Leadership Cycle and develops a model that accounts for the separate effects which they may have on the generation or suppression of conflict at any moment in time. He then tests each of the five theories against the same data on military disputes. The data consist of annual event counts of the number of Militarized Interstate Disputes assembled by the Correlates of War project. These data span the years 1816–1976, yielding 161 observations in this event-count series. Pollins estimates a lagged Poisson regression model consistent with each theory based on the following model:

$$\Pr(Y_t \mid \lambda_t) = \frac{\lambda_t^{y_t} e^{-\lambda_t}}{y_t!}$$
$$\lambda_t = \exp(\rho Y_{t-1} + X_t \delta + M_t \zeta),$$

where λ_t is the mean of the Poisson random variable, Y_{t-1} is a lagged endogenous count, X_t is a $1 \times K$ matrix of phase parameters for the different model specifications, and M_t is the number of members in the interstate system in period t.[18]

Pollins recognizes time dependence in the Dispute data series on theoretical grounds. He argues that there is a degree of inertia in peace and conflict in the international system, thus the level of conflict in any year will help predict the level of conflict in the subsequent year. Moreover, the number of disputes observed in any year should correlate with the number of states in the system, for the simple reason that opportunities for conflict will vary with system size. Pollins makes use of this information in his specification of all five models as well as in a "Sophisticated Null" model which becomes his benchmark for comparison for the five contending Long Cycle specifications (Pollins, 1996, 109–110). This null model predicts the annual number of Disputes using a lagged endogenous variable (to represent the inertia, or autocorrelation effect) and the number of states in the system that year (which, given that this number rises through history, serves as a kind of trend component in the time series).

Based on this model, Pollins estimates six different specifications for X_t based on the theories of Gilpin, Wallerstein, Goldstein, Modelski and Thompson, his own "Coevolving Systems," as well as a "Sophisticated Null" model to explain the presence of economic or leadership cycles in international disputes. The Sophisticated Null model contains only the lagged count and the number of members in the interstate system (attempting to capture autocorrelation and trend effects, respectively) and no other X_t.

The results of the Pollins's study show that each model passes statistical muster, despite their differences

[18]All variable definitions follow those used in Pollins (1996).

in approach and some conflicting claims between them. Each of the five models passes a likelihood ratio test against the "Sophisticated Null" model driven only by two trend components (lagged Militarized Interstate Disputes and the number of members in the international system), while the estimated parameters for each model are found to be consistent with most of the hypotheses generated by that particular model. Some models capture more of the variation in the annual dispute count than others, and Pollins shows that his Coevolving Systems model captures substantially more of this variation than any other contender.

Pollins's theoretical argument about inertia in the level of international conflict leads him to the lagged Poisson specification. Based on the earlier results, this correction is not advised for two reasons. First, the LP coefficient for the lagged dependent variable captures only the growth rate of the series. If an event-count time series is persistent but mean-reverting, the lagged Poisson model will have an estimated coefficient of zero for the lagged dependent variable. Since the model only captures exponential deterministic trends, it has limited applicability in most cases.[19] Second, the Monte Carlo analysis shows that simply including a lagged endogenous variable as a regressor in the standard Poisson regression model is a poor correction for the inefficiency induced by persistent serial dependence. Thus, more efficient estimates should be obtained with the PEWMA model.

Our reassessment of these five theoretical models uses three criteria: First, does the model in question fit the data better than the "Sophisticated Null" specification? Because this null model is nested within each of Pollins's five specifications, the log-likelihood ratio test may be used to answer this question. We therefore estimate a "Sophisticated Null" model using the PEWMA, including the number of states in the system as the only regressor, since the dynamics are captured by the weighting term ω. Second, are the hypotheses generated by each theoretical model supported by the PEWMA parameter estimates and standard errors? Here we will find whether the Lagged Poisson estimates presented in Pollins (1996) were overly optimistic in the inferences they produced. Third, does any one of these models perform better than all others? Because the Gilpin, Wallerstein, Modelski and Thompson, Goldstein, and Pollins models are nonnested, the log-likelihood ratio test is not appropriate for comparisons between them. Instead, we look to the Akaike Information Criterion (AIC) as the appropriate standard for comparing the relative fit and parsimony of these nonnested models.

We begin our reanalysis of Pollins's models by first assessing the time-series properties of the disputes series. Figure 3 presents a plot of the original data and the autocorrelation function for the Disputes series.

It is clear from the figure that the Disputes series is highly persistent. Thus, models with a fixed mean such as the standard Poisson regression or standard negative binomial regression will fail to account for the change in the mean over time. While an LP model may explain the trend in the data, it is highly unlikely that the growth in the number of Disputes has a deterministic growth rate as implied by this model. Rather, the more likely scenario is a stochastic trend, similar to that implied by the PEWMA.

The results of our reestimations using the PEWMA and LP are presented in Table 1.[20] We begin our comparison of Lagged Poisson and PEWMA estimates of these six models by comparing the model parameters and associated standard errors produced by these two estimators (i.e., results for the five theoretical models and the "Sophisticated Null"). As for the model parameters themselves, relatively little change is observed. Only four of the twenty-seven parameters reestimated for these models changed by more than 30 percent. Nine changed by less than 10 percent. This essential stability in the coefficients is not surprising given that the LP and PEWMA estimators are in the same linear exponential family of estimators.

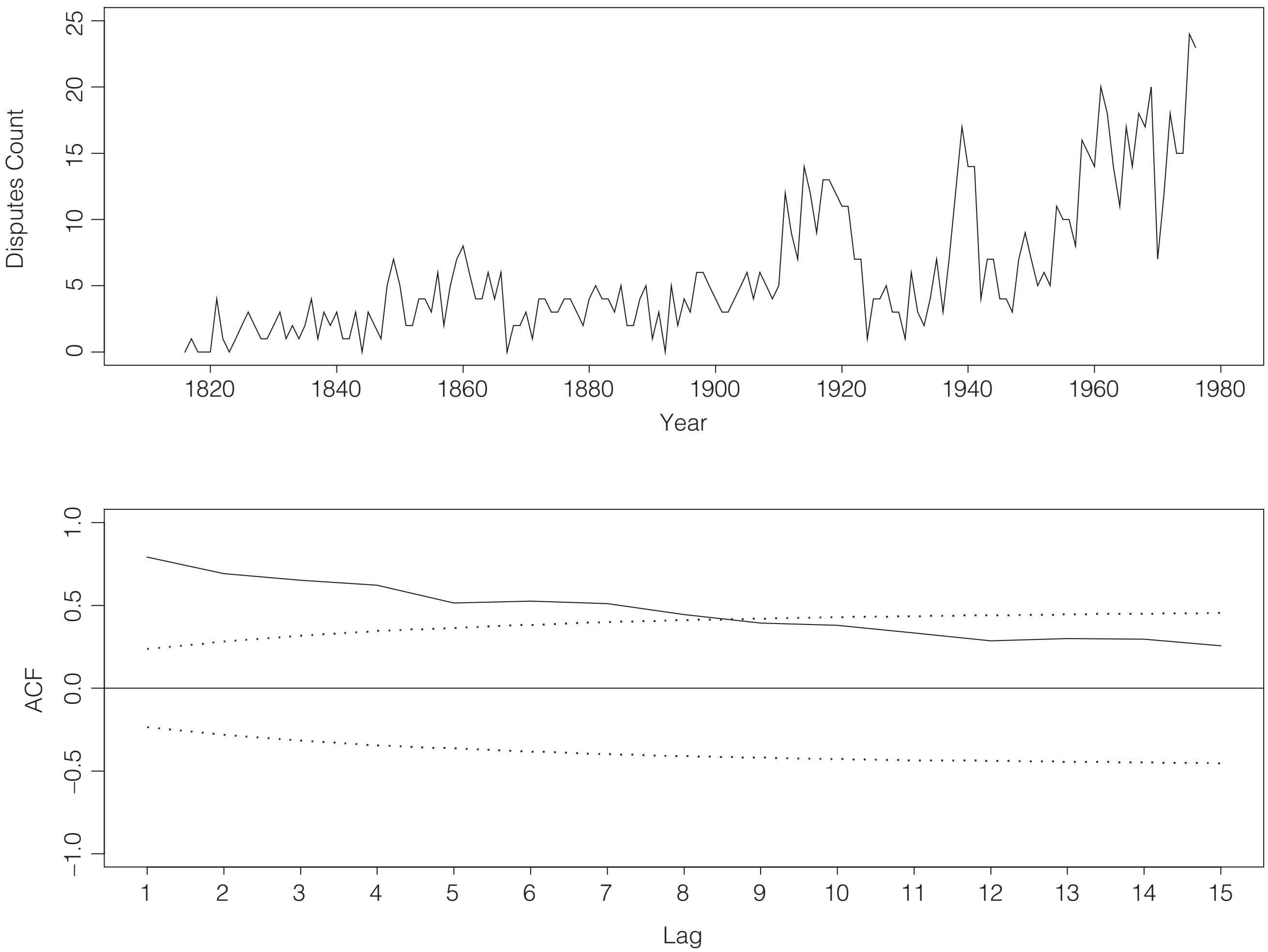
The estimated standard errors show a more distinctive pattern. Downward bias in the LP estimated standard errors is indicated, as all but two of the PEWMA standard errors are larger, and most are larger by 30 percent or more. The source of this bias is the failure to account for the time series properties of the Disputes series. This is consistent with the Monte Carlo findings that the LP estimates produce "overconfident" results (see footnote 16).

Inferences about the theories are affected; the statistical significance of individual model parameters is lower in the PEWMA results, but still strong enough in this case to pass widely accepted criterion ($p < .05$). However, the central theoretical arguments put forward in Pollins (1996) depend not on the significance of individual coefficients, but on the differences between particular coefficients. For any one of the models tested by Pollins, the

[19]For the Correlates of War Disputes data used by Pollins, the series has a growth rate of approximately 1 percent. This is consistent with his earlier estimates. Taking first differences, the data is stationary. Specification as a deterministic trend is highly unlikely, as most economic and social science series exhibit stochastic trends (Nelson and Plosser 1982).

[20]The "Lagged Poisson" results presented in Table 1 represent a successful replication of the results presented in Pollins (1996).

FIGURE 3 **Militarized Interstate Disputes Series and Autocorrelation Function, 1816–1976**



coefficient associated with the system phase (e.g., Victory, Maturity, Decline, and Ascent in the Wallerstein model) represents the effect on the expected annual number of armed Disputes within the phase. If the cycle phases are to predict the level of conflict in the international system, then the expected level of conflict (whose point-estimate is the phase parameter) should change significantly as the world moves from one period to the next (e.g., from Victory to Maturity, and so on).

The PEWMA results suggest that the claims found in Pollins (1996) regarding models other than his own are overconfident. While the LP estimates of the Wallerstein, Modelski and Thompson, and Goldstein formulations indicated reasonable evidence of cycles, most of that evidence disappears in the PEWMA estimates. In Table 1, only two of the twelve phase transition tests for these three models are now significant, rather than seven of twelve as estimated by the LP. Claims for the Pollins model are only slightly weaker than originally reported. The PEWMA estimates show that four of the six hypotheses regarding phase transition are substantiated rather than six of six as estimated by the LP. Given the PEWMA results, the Pollins model has difficulty in distinguishing between periods when the lowest levels of conflict are predicted (periods CR2 and CR3). The model's ability to distinguish between periods predicting medium and high levels of conflict remains significant.

The LP and PEWMA results yield very different answers concerning overall model adequacy. Pollins (1996) reports that each of the five models tested is shown superior to the null model, based on the log-likelihood ratio test using the LP estimates (the Wallerstein model is significant just above the 0.05 level in our replication). From the PEWMA estimates only, Pollins's "Coevolving Systems" model is shown to be superior to the null model by this same test. We also compared the PEWMA results for these models using the Akaike Information Criterion (AIC). Pollins's model exhibited the lowest

TABLE 1 Comparison of Lagged Poisson and PEWMA Results

Model and Phases	Parameters		Standard Errors		Transition Tests for Phase Shifts		Log-Likelihood Ratio Tests		AIC	
	LP	PEWMA	LP	PEWMA	LP	PEWMA	LP	PEWMA	LP	PEWMA
Null Model									**761.2**	**730.4**
Gilpin									744.5	736.1
UK	.67	.95	.08	.32						
Interregnum	1.08	1.00	.08	.17	*					
US	1.17	.87	.11	.20						
Members	.005	.008	.001	.006						
Lag	.07		.01							
ω		.67		.04						
Wallerstein									757.9	734.1
Victory	.80	1.02	.08	.16		*				
Maturity	.78	.51	.10	.14	*					
Decline	.58	.49	.13	.17	*					
Ascent	.99	.82	.12	.18						
Members	.07	0.011	.01	0.006						
Lag	.01		.002							
ω		.68		.04						
Modelski and Thompson							*		753.1	736.7
World Power	.61	.54	.10	.15						
Delegitimation	.67	.70	.11	.18	*					
Deconcentration	.87	.76	.08	.20						
Global War	.94	.88	.09	.17	*					
Members	.06	0.009	.01	0.006						
Lag	.01		.002							
ω		.68		.04						
Goldstein							*		749.1	734.2
Stagnation	.64	.65	.10	.14	*					
Rebirth	.86	1.00	.10	.13	*					
Expansion	1.07	.98	.09	.11						
War	1.10	1.06	.11	.11	*	*				
Members	.06	.01	.01	.006						
Lag	.01		.001							
ω		.70		.05						
Pollins							*	*	714.6	725.0
CR2	-.02	.42	.24	.42	*					
CR3	.56	.72	.12	.15	*					
CR4	.94	.72	.11	.12	*	*				
CR5	.75	.50	.11	.12	*	*				
CR6	1.20	1.15	.09	.12	*	*				
CR7	.85	.83	.12	.11	*	*				
CR8	1.46	1.34	.14	.19	n.a.	n.a.				
Members	.04	.01	.01	.003						
Lag	.01		.002							
ω		.83		.06						

Notes: * indicates significance at the 0.05 level or better for given test. The "Transition Test" is a t-test of successive phase coefficients, described in text and in Pollins (1996) footnote 17 as amended. The test checks for significant shift in frequency of Disputes as system moves from one phase to another. The log-likelihood ratio tests are conducted for each model against the LP and PEWMA "Sophisticated Null," respectively. The "AIC" is the Akaike Information Criterion for fit and parsimony. Lower values are superior. All of the estimated coefficients except for "CR2" are significant at the 0.05 level.

(i.e., best) AIC of all tested and was the only theoretical formulation to show itself superior to the null model.

The results of our replication of the Pollins study shed additional light on the PEWMA estimator and on the models and substantive claims examined in that earlier study. First, consistent with the Monte Carlo examination of the PEWMA estimator, we find that the model parameters tended to change only moderately, while estimated standard errors increased in most instances. This revised (and we believe more accurate) picture of the standard errors is sufficient to change several conclusions from the study we replicated.

Second, using the PEWMA we did not find all five theoretical models to be superior to the null model. Only Pollins's "Coevolving Systems" model passed the likelihood ratio test, and only this same model had a lower AIC than the null model. And once we took direct account of the dynamics in the Dispute time series (represented by the weighted moving average coefficient, ω), only the Pollins model showed multiple phase transitions as hypothesized. Our revised results underscore the call made by Pollins (1996: 116) for deeper specification of the separate contributions made by the Leadership Cycle and the Long Wave to the generation of armed conflict.

Third, we do find support for Pollins's claim that the "Coevolving Systems" model can extend these frameworks to understanding conflict patterns beyond this limited domain by taking account of independent effects of the Long Wave and the Leadership Cycle. His call for future research in this area to follow his approach is consistent with our PEWMA findings. The Pollins model was the only one among the five to show improvement over the null model (by both the log-likelihood and AIC criteria) and the only model to exhibit a number of significant changes in Dispute frequencies from one phase to the next.

In sum, the PEWMA reanalysis of the Pollins (1996) study does not resolve all debates in long-cycle research. But it does demonstrate quite clearly that proper modeling of the dynamic properties of event time series is key to model specification and theory testing. And direct estimation of these dynamic properties, as contained in the moving average parameter (ω), should be the preferred approach for those modeling event count time series.

Conclusion

The PEWMA model offers a straightforward and easily estimable (and interpretable) method for modeling event-count data with persistent dynamics. Our experience is that many event-count time series are persistent. Indeed, persistence is probably the most common feature we have found in time-series event counts. However, many series also exhibit independence, and for those cases the Poisson or negative binomial models are appropriate.

An alternative situation is when data have dynamics but are mean reverting. The proper specification for such a case is an autoregressive model (see Brandt and Williams 1998). Thus, PEWMA is not always the best model for time-series event counts, but it will often be so. Analysts of event-count data need a more complete tool bag than they now have. The PEWMA is an important addition to the tools of political scientists.

Our Monte Carlo experiments indicate significant efficiency losses occur when event-count time series data are modeled using either Poisson or negative binomial assumptions. Since most political scientists are trying to explain events with covariates, our focus has been on the loss of efficiency of coefficients on covariates. There is every reason to believe that if persistence in data is ignored, not only will estimates be inefficient, but reported confidence intervals from Poisson and negative binomial models will be much too small. This means that scientific judgments can be seriously flawed.

Also, our PEWMA model is easy to specify and estimate. We have written and made available software for estimating this model. As noted earlier, the software is as fast as standard estimation algorithms for Poisson and negative binomial regression models currently used in political science. We will be making a standalone version of this software available as part of an updated version of King's (1994) COUNT program.[21]

Finally, the PEWMA reanalysis of the Pollins study underscores the importance of properly modeling the dynamic properties of event-count time-series data. The lagged Poisson specification Pollins originally employed produced results that differed in substantively important ways from those the PEWMA model generated. Unlike Pollins, most previous research using time-series event-count data does not attempt to model the dynamics of these data at all. Since even an inadequate specification of the dynamics can produce somewhat misleading results, there is every reason to suspect that failing to consider them at all poses a grave threat to the validity of the results. The PEWMA model offers a relatively simple way to model the dynamics of persistent event-count series and thus to avoid these problems.

[21]Both the Gauss source code and the stand-alone program can be found at http://www.polsci.indiana.edu/jotwilli/home/default.htm.

Appendix A
PEWMA Filtering Algorithm and Predictive Distribution

The PEWMA uses a version of the Kalman filter to estimate a time-varying mean for event-count data. This filter can be derived using the three basic assumptions that define the PEWMA model, in equations (1)–(4).

The filter computes three quantities: the unconditional mean of the process $(E[\mu_{t-1}|Y_{t-1}])$, the conditional mean of the process $(E[\mu_t|Y_{t-1}])$, and the posterior mean of the process $(E[\mu_t|Y_t])$. Similar to Gaussian ARIMA models the Kalman filter for count data requires that $E[\mu_t|Y_{t-1}] = E[\mu_{t-1}|Y_{t-1}]$, and $Var[\mu_t|Y_{t-1}] > Var[\mu_{t-1}|Y_{t-1}]$. In words, the conditional mean in periods t and $t-1$ is equal, but the conditional variance for period t is larger. This effect is achieved in the gamma distribution by using the hyperparameter $\omega \in (0,1]$ to discount the value of past observations in the computation of $E[\mu_t|Y_{t-1}]$. The conjugacy assumption ensures that the mean of the gamma distribution is the same in each period, but that the conditional predictive variance in period t is larger than in period $t-1$. This effect is induced by multiplying μ_{t-1} by a factor less than 1 in the transition equation (2).

To derive the Kalman filter equations for the PEWMA, we compute the prior, conditional, and posterior mean. First, substituting the gamma prior (4) into the transition equation (3) yields

$$\begin{aligned}\mu_t &= \mu_{t-1}e^{r_t}\eta_t \\ &= \frac{a_{t-1}\eta_t}{b_{t-1}}\exp(X_t\delta + r_t).\end{aligned}$$

Conditional on Y_{t-1}, μ_t is gamma distributed and we write $\mu_t|Y_{t-1} \sim \Gamma(a_{t|t-1}, b_{t|t-1})$. This is the prior mean of the process at time t.

From the properties of the gamma distribution, $\mu_t|Y_{t-1}$ has a gamma distribution with parameters $a_{t|t-1}$ and $b_{t|t-1}$ such that

$$a_{t|t-1} = \omega a_{t-1} \quad \text{and} \tag{9a}$$

$$b_{t|t-1} = \omega b_{t-1}\exp(-X_t\delta - r_t),\ \omega \in (0,1]. \tag{9b}$$

Based on the properties of the gamma distribution, we can then calculate $E[\mu_t|Y_{t-1}]$ and $Var[\mu_t|Y_{t-1}]$:

$$E[\mu_t|Y_{t-1}] = \frac{a_{t|t-1}}{b_{t|t-1}} = \frac{a_{t-1}}{b_{t-1}\exp(-X_t\delta - r_t)} = E[\mu_{t-1}|Y_{t-1}] \tag{10}$$

and

$$Var[\mu_t|Y_{t-1}] = \frac{a_{t|t-1}}{b^2_{t|t-1}} = \frac{\omega a_{t-1}}{\omega^2(b_{t-1})^2\left(\exp(-X_t\delta - r_t)\right)^2}$$

$$= \omega^{-1}Var[\mu_{t-1}|Y_{t-1}] \tag{11}$$

These are the conditional mean and variance of the process and the mean and variance conditions demanded above are satisfied. When $\omega = 1$, this model is the Poisson model since the mean and variance of the model will be equal. In this case, the conditional Poisson distribution has a fixed mean, so the value of a_{t-1}/b_{t-1} is absorbed in the constant term of the Poisson regression model.

To evaluate these quantities and derive the posterior distribution, we need a formula for computing r_t. Recall from equation (7) that the transition equation can be written

$$\ln(\mu_t) - \ln(\mu_{t-1}) = r_t + \ln(\eta_t)$$

where μ_t is the conditional mean, r_t is the per period growth rate, and η_t are the errors. The left-hand side of this equation is an estimate of the growth rate of the series of event counts. The right-hand side is the growth rate, plus a stochastic component. The left-hand side of this equation is zero in expectation, since we constructed $E[\mu_{t-1}|Y_{t-1}] = E[\mu_t|Y_{t-1}]$ and the model has a zero growth rate. Therefore, since h_t follows a beta distribution, this expectation can be evaluated using an integral formula:

$$\begin{aligned}r_t &= -E[\ln(\eta_t)] \\ &= -\int_0^1 \frac{1}{B(\omega a_{t-1},(1-\omega)a_{t-1})}\eta^{\omega a_{t-1}-1}(1-\eta_t)^{a_{t-1}(1-\omega)-1}\ln(\eta_t)d\eta_t \\ &= -\frac{1}{B(\omega a_{t-1},(1-\omega)a_{t-1})}\int_0^1 \eta_t^{\omega a_{t-1}-1}(1-\eta_t)^{a_{t-1}(1-\omega)-1}\ln(\eta_t)d\eta_t \\ &= \Psi(a_{t-1}) - \Psi(\omega a_{t-1}),\end{aligned}$$

where $B(\cdot,\cdot)$ is the beta function and $\Psi(\cdot) = \frac{\partial \ln\Gamma(x)}{\partial x}$ is Euler's psi or digamma function. This expectation can be numerically evaluated using standard approximations for the digamma function (Abramowitz and Stegun, 1972).

In order to make use of the structural model, we need to be able to compute the posterior of μ_t once Y_t is available. This will allow us to construct the predictive distribution, $\Pr(y_t|Y_{t-1})$, and likelihood function, $\Pr(\omega,\delta|y_1 \ldots y_T, X_t)$. We compute the posterior using Bayes rule. Given the observed count yt we update the conditional values of $a_{t|t-1}$, $b_{t|t-1}$, $\mu_{t|t-1}$ to find $\mu_t|Y_t$ using Bayes rule:

$$\Pr(\mu_t | Y_t) = \frac{\Pr(Y_t | \mu_t) \cdot \Pr(\mu_t)}{\int_0^\infty \Pr(Y_t | \theta) \cdot \Pr(\theta) d\theta}$$

$$= \frac{e^{-\mu_t\left(1+b_{t|t-1}\right)} \mu^{y_t + a_{t|t-1} - 1} \left(1 + b_{t|t-1}\right)^{-\left(y + a_{t|t-1}\right)}}{\Gamma\left(y_t + a_{t|t-1}\right)}$$

where Γ refers to the gamma function. This is a gamma distribution, $\mu_t | Y_t \sim \Gamma(a_{t|t-1} + y_t, b_{t|t-1} + 1)$. Then, the posterior $\Pr(\mu_t | Y_t)$ is given by a gamma distribution with parameters generated by the recursions,

$$a_t = a_{t|t-1} + y_t = \omega a_{t-1} + y_t \text{ and} \quad (13a)$$

$$b_t = b_{t|t-1} + 1 = \omega b_{t-1} + \exp(X_t \delta + r_t), \quad (13b)$$

where (13b) follows from a renormalization of the scale of the gamma distribution. These recursions and the set of initial values define a Kalman filter for the count data with a Poisson measurement equation and a gamma-distributed prior. The filter is defined by the recursive system of equations (9a–13b).

Since the gamma is the natural conjugate distribution of the Poisson, the resulting predictive distribution and likelihood are in the same family of distributions as the Poisson. This predictive distribution is a Poisson with a gamma prior,

$$\Pr(y_t | Y_{t-1}) = \int_0^\infty \Pr(y_t | \mu_t) \Pr(\mu_t | Y_{t-1}) d\mu_t$$

$$= \frac{\Gamma(y_t + \omega a_{t-1})}{y_t! \Gamma(\omega a_{t-1})} \left(\omega b_{t-1} \exp(-X_t \delta - r_t)\right)^{\omega a_{t-1}} \quad (14)$$

$$\times \left(1 + \omega b_{t-1} \exp(-X_t \delta - r_t)\right)^{-(y_t + \omega a_{t-1})}$$

where a_{t-1}, b_{t-1}, and r_t are based on the filter computations. This is a negative binomial distribution.[22]

The filter parameters can then be used to compute the predictive distribution and maximum-likelihood estimates in equations (6) and (14).

[22]A proper distribution for μ_t requires that $\mu_t \neq 0$. Thus, we set $t = \tau$ where τ is the time of the first non-zero observation. In practice, we initialize the filter at $t = 0$ using the sample mean of the series for μ_0. The prior becomes diffuse as a and b approach zero. Although the gamma distribution is degenerate in this case, $a = b = 0$ implies that $\Pr(\mu = 0) = 1$. The distribution of μ_t is then properly defined at $t = \tau$ where τ is the first nonzero observation (Harvey and Fernandes 1989, 408). This can be easily implemented by removing the first t – 1 zero observations from the data.

Appendix B
PEWMA Forecasts and Predictions

The forecast function for the PEWMA model is derived given estimates of ω and δ. From the properties of the negative binomial distribution the mean and variance of the predictive distribution at time $T + 1$ are obtained from the filter parameters:

$$E(y_{T+1} | Y_T) = \frac{a_T}{b_T} = \frac{a_{T+1|T} + y_T}{b_{T+1|T} + 1} = \frac{\omega a_T + y_T}{\omega b_T + \exp(X_t \delta + r_t)}, \text{ and} \quad (15)$$

$$V(y_{T+1} | Y_T) = \frac{a_{T+1|T}(1 + b_{T+1|T})}{b_{T+1|T}^2} = \omega^{-1} V(\mu_T | Y_T) + E(\mu_T | Y_T), \quad (16)$$

where a_T and b_T are the posterior gamma distribution parameters computed by the filter.

Based on repeated substitutions of a_{t-1}, $a_{t|t-1}$, a_t, b_{t-1}, $b_{t|t-1}$, b_t (from the filter computations) the forecast function for the *one-step ahead prediction* is:

$$\bar{y}_{T+1|T} = \exp(X_{T+1}\delta + r_{T+1}) \frac{\sum_{j=0}^{T-1} \omega^j y_{T-j}}{\sum_{j=0}^{T-1} \omega^j \exp(X_{T-j}\delta + r_{T-j})} = \bar{\mu}_{T+1|T}. \quad (17)$$

This forecast function has a natural interpretation in the context of known time-series models. The forecast conditions the observation in period $T + 1$ on the realizations of past observations, as well as the covariates that enter the model multiplicatively according to (2). The forecast mean level is a discounted weighted average of past observations, where the weights decline exponentially. The forecast function produces estimates that follow an *exponentially weighted moving average* (EWMA) (Harvey 1989, 25–26). When T is large, $\bar{y}_{t+1|t}$ approaches $\omega \bar{y}_{t|t-1} + (1 - \omega) y_t$ for $t = 1, \ldots, T$. This is the limiting form of the EWMA for $\mu_{t+1|t}$.

Multi-step prediction requires integration of the conditional density function, equation (14). Thus, the multi-step predictions cannot be found by a simple additive recursion as in the Gaussian model. The ℓ-step ahead predictive distribution is

$$\Pr(y_{T+\ell} | Y_T) = \int_0^\infty \Pr(y_{T+\ell} | \mu_{T+\ell}) \cdot \Pr(\mu_{T+\ell} | Y_T) \partial \mu_{T+\ell} \quad (18)$$

Simply using the recursions (9a–13b) ignores the evolution of μ_t for observations at $T + 1$, $T + 2$, . . . , $T + \ell - 1$.

Thus, accounting for the conditional distribution of μ_t in these interim periods is accomplished by rewriting (18) as

$$\Pr(y_{T+\ell}|Y_T) = \sum_{y_{T+\ell-1}} \dots \sum_{y_{T+1}} \prod_{j=1}^{\ell} \Pr(y_{Y+j}|Y_{T+j-1}). \qquad (19)$$

In practice, equation (19) can be evaluated numerically. As Harvey and Fernandes note, closed-form solutions for $\Pr(y_{T+\ell|T})$ for $\ell > 1$ can also be evaluated numerically.

While the multi-step predictive distribution is numerically difficult, it can be shown that

$$E[y_{T+\ell}|Y_T] = \frac{a_T}{b_T}\exp(X_{T+\ell}\delta + r_{T+\ell}) \;\; \forall\, \ell \geq 1 \cdot$$

The derivation is based on an induction argument using the filter recursions and taking the conditional expectations at times $T+\ell-2$ and $T+\ell-1$ (Harvey and Fernandes 1989, 409).

Manuscript submitted September 21, 1998.
Final manuscript received April 28, 2000.

References

Abramowitz, Milton, and Irene A. Stegun. 1972. *Handbook of Mathematical Functions and Formulas, Graphs, and Mathematical Tables.* Washington: U. S. Government Printing Office.

Alzaid, A. A., and M. Al-Osh. 1990. "An Integer-Valued pth-Order Autoregressive Structure (INAR(p)) Process." *Journal of Applied Probability* 27:314–324.

Beck, Nathaniel and Jonathan N. Katz. 1996. "Nuisance v. Substance: Specifying and Estimating Time-Series-Cross-Section Models." *Political Analysis* 6:1–36.

Benoit, Kenneth. 1996. "Democracies Really Are More Pacific (In General)." *Journal of Conflict Resolution* 40:636–657.

Brace, Paul, and Barbara Hinckley. 1993. "Presidential Activities from Truman through Reagan: Timing and Impact." *Journal of Politics* 55:382–398.

Brandt, Patrick T. and John T. Williams. 1998a. "A Linear Poisson Autoregressive Model: The Poisson AR(p) Model." Unpublished manuscript. Indiana University.

Brockwell, Peter J., and Richard A. Davis. 1996. *Introduction to Time Series and Forecasting.* New York: Springer-Verlag.

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* 38:196–210.

Caldeira, Gregory, and Christopher Zorn. 1998. "Of Time and Consensual Norms in the Supreme Court." *America" Journal of Political Science* 42:874–902.

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

Chan, K. S. and Johannes Ledolter. 1995. "Monte Carlo EM Estimation for Time Series Models Involving Counts." *Journal of the American Statistical Association* 90:242–252.

Davis, Richard A., William T. M. Dunsmuir and Ying Wang. 1998a. Modelling Time Series of Count Data. In *Asymptotics, Nonparametrics and Time Series,* ed. Subir Ghosh. New York: Marcel Dekker.

Davis, Richard A., William T. M. Dunsmuir, and Ying Wang. 1998b. "On Autocorrelation in a Poisson Regression Model." Unpublished manuscript, Colorado State University.

DeCastro, Gustavo. 1952. "Note on Differences of Bernoulli and Poisson Variables." Portugaliae *Mathematica* 11:173–5.

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

Diggle, Peter J., Kung-Yee Liang, and Scott L. Zeger. 1994. *Analysis of Longitudinal Data.* Oxford: Clarendon Press.

Du, Jin-Guan, and Yuan Li. 1991. "The Integer-Valued Autoregressive (INAR(p)) Model." *Journal of Time Series Analysis* 12:129–142.

Durbin, J., and S. J. Koopman. 1997. "Time Series Analysis of Non-Gaussian Observations Based oil State Space Models." Unpublished manuscript, Department of Statistics, London School of Economics and Political Science.

Durbin, J., and S. J. Koopman. 2000. "Time Series Analysis of Non-Gaussian observations Based on State Space Models from Both Classical and Bayesian Perspectives." *Journal of the Royal Statistical Society, Series B* 62:3–56.

Eyerman, Joe, and Robert A. Hart. 1996. "An Empirical Test of the Audience Cost Proposition." *Journal of Conflict Resolution* 40:597–616.

Fordham, Benjamin O. 1998*a*. "Partisanship, Macroeconomic Policy, and the U. S. Uses of Force, 1949–94." *Journal of Conflict Resolution* 42:418–439.

Fordham, Benjamin O. 1998*b*. "The Politics of Threat Perception and the Use of Force: A Political Economy Model of U. S. Uses of Force, 1949–1994." *International Studies Quarterly* 42:567–590.

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

Gilpin, Robert. 1981. *War and Change in International Politics.* Cambridge: Cambridge University Press.

Goldstein, Joshua S. *1988. Long Cycles.* New Haven: Yale University Press.

Goldstein, Joshua S. *1991.* "Reciprocity in Superpower Relations: An Empirical Analysis." *International Studies Quarterly* 35:195–209.

Goldstein, Joshua S. and John R. Freeman. 1990. *Three-Way Street.* Chicago: University of Chicago Press.

Gowa, Joanne. 1998. "Politics at the Water's Edge: Parties, Voters, and the Use of Force Abroad." *International Organization* 52:307–24.

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

Grunwald, G. K., R. J. Hyndman, L. Tedesco, and R. L. Tweedie. 1997. A unified view of linear AR(1) models. Unpublished manuscript. Monash University.

Harvey, A. C. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge: Cambridge University Press.

Harvey, A. C., and C. Fernandes. 1989. "Time Series Models for Count or Qualitative Observations." *Journal of Business and Economic Statistics* 7:407–417.

Hildreth. C., and J. Y. Lu. 1960. Demand relations with autocorrelated disturbances. Technical report. Michigan State University Agricultural Experiment Station Technical Bulletin 276.

Huang, Chi, Woosang Kim, and Samuel Shiouh Guang Wu. 1992. "Conflicts Across the Taiwan Strait, 1951–78." *Issues and Studies* 28:35–58.

Jackman, Simon. 1998. "Analyzing Time Series Counts of International Disputes." Presented at the Annual Meeting of the Midwest Political Science Association.

Jorgensen, Bent, Rodrigo Labouriau, and Soren Lundbye-Christensen. 1996. "Linear Growth Curve Analysis Based on Exponential Dispersion Models." *Journal of the Royal Statistical Society, Series B* 58:573–592.

Jorgensen, Bent, Soren Lundbye-Christensen, Peter Xue-Kun Song, and Li Sun. 1999. "A State Space Model for Multivariate Longitudinal Count Data." *Biometrika* 86:169–181.

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* 32:838–863.

King, Gary. 1989a. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference.* Cambridge: Cambridge University Press.

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

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

Kinsella, David. 1995. "Nested Rivalries: Superpower Competition, Arms Transfers, and Regional Conflict, 1950–1990." *International Interactions* 21:109–125.

Kinsella, David, and Herbert K. Tillema. 1995. "Arms and Aggression in the Middle East: Overt Military Interventions." *Journal of Conflict Resolution* 39:306–329.

Kitagawa, Genshiro, and Will Gersch. 1996. *Smoothness Priors Analysis of Time Series.* New York: Springer-Verlag.

Krause, George. 1994. "Federal Reserve Policy Decision Making: Political and Bureaucratic Influences." *American Journal of Political Science* 38:124–144.

Krause, George, and David B. Cohen. 1997. "Presidential Use of Executive Orders, 1953–1994." *American Politics Quarterly* 25:458–481.

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

Mayer, Kenneth R. 1995. "Electoral Cycles in Federal Government Prime Contract Awards: State-Level Evidence from the 1988 and 1992 Presidential Elections." *American Journal of Political Science* 39:162–185.

McKenzie, Ed. 1988. "Some ARMA Models for Dependent Sequences of Poisson Counts." *Advances in Applied Probability* 20:822–835.

Modelski, George, and William Thompson. 1987. Testing Cobweb Models of the Long Cycle. In *Exploring Long Cycles,* ed. George Modelski. Boulder: Lynne Rienner.

Modelski, George, and William R. Thompson. 1996. *Leading Sectors and World Powers: The Coevolution of Global Economics and Politics.* Columbia: University of South Carolina Press.

Moore, Will H., Ronny Lindstrom, and Valerie O'Regan. 1996. "Land Reform, Political Violence, and the Economic Inequality-Political Conflict Nexus: A Longitudinal Analysis." *International Interactions* 21:335–363.

Nelson, C. R., and C. I. Plosser. 1982. "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications." *Journal of Monetary Economics* 10:139–162.

Nelson, Daniel B. 1990. "Stationarity and Persistence in the GARCH(1,1) Model." *Econometric Theory* 6:318–334.

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

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

Remmer, Karen L. 1998. "Does Democracy Promote Interstate Cooperation? Lessons from the Mercosur Region." *International Studies Quarterly* 42:25–52.

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

Schneider, Gerald, Thomas Widmer, and Dieter Ruloff. 1993. "Personality, Unilateralism, or Bullying: What Caused the End of the Cold War?" *International Interactions* 41:407–427.

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

Shephard, Neil. 1994. "Local Scale Models: State Space Alternative to Integrated GARCH Processes." *Journal of Econometrics* 60:181–202.

Shephard, Neil. 1995. "Generalized Linear Autoregressions." Unpublished manuscript. Nuffield College.

Skellam, J. G. 1946. "The Frequency Distribution of the Difference Between Two Poisson Variates Belonging to Different Populations." *Journal of the Royal Statistical Society* 109:296.

Squier, John. 1997. "Time-Series Models for Event-Count Data: Properties and Estimation of Integer-Valued Autoregressive Processes with Poisson Marginal Distributions." Unpublished, University of Michigan.

Squier, John. 1996. "Time-Series Models for Event-Count Data: Properties and Perils of Integer-Valued First-Order Autoregressive (INAR(1)) Processes with Poisson Marginal Distributions." Presented at the 1996 Annual meeting of the Southern Political Science Association, Atlanta, Georgia.

Strackee, J., and J. J. Denier van der Gon. 1962. "The frequency distribution of the difference between two Poisson variates." *Statistica Neerlandica* 16:17–23.

Volgy, Thomas J., and Lawrence E. Imwalle. 1995. "Hegemonic and Bipolar Perspectives on the New World Order." *American Journal of Political Science* 39:819–34.

Wallerstein, Immanuel. 1983. "Three Instances of Hegemony in the History of the World Economy." *International Journal of Comparative Sociology* 24:100–108.

Ward, Michael D., and Sheen Rajmaira. 1992. "Reciprocity and Norms in United States-Soviet Foreign Policy." *Journal of Conflict Resolution* 36:369–385.

West, Mike, and P. Jeff Harrison. 1986. "Monitoring and Adaptation in Bayesian Forecasting Models." *Journal of the American Statistical Association* 81(395):741–750.

West, Mike, and Jeff Harrison. 1997. *Bayesian Forecasting and Dynamic Models.* New York: Springer-Verlag.

West, Mike, P. Jeff Harrison, and Helio Migon. 1985. "Dynamic Generalized Linear Models and Bayesian Forecasting (with comments)." *Journal of the American Statistical Association* 80:73–83.

Zeger, Scott L. 1988. "A regression model for time series of counts." *Biometrika* 75:621–629.

Zeger, S. L., and B. Qaqish. 1988. "Markov Regression Models for Time Series: A Quasi-Likelihood Approach." *Biometrics* 44:1019–1031.