Dynamic Modeling for Persistent Event Count Time Series *

Patrick T. Brandt and John T. Williams
Department of Political Science
Workshop in Political Theory and Policy Analysis
Indiana University
Woodburn Hall 210
Bloomington, IN 47405
E-mail: pbrandt@indiana.edu, jotwilli@indiana.edu

December 24, 1998

Abstract

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. We present an application to Supreme Court case agendas as well.

*An earlier version of this work was presented at the 1998 Summer Meetings of the Political Methodology Group, University of California, San Diego, July 22-26, 1998. Williams research was supported by National Science Foundation grant SBR-9422645. All GAUSS code used for this analysis is available with documentation. We wish to thank Rich Pacelle for making an updated version of his data available. We also wish to thank Chris Achen, Neal Beck, Michael Craw, Ben Fordham, John Freeman, Simon Jackman, Gary King, David LeBlang, Michael McGinnis, Walter Mebane, Brian Pollins, Jim Stimson and Pravin Trivedi for their comments and suggestions. Standard disclaimers apply.

1 Introduction

Owing to the impact of King's efforts (1988, 1989a, 1989b, 1989c), Poisson regression and Generalized Event Count estimators are commonly used in political science. It is rare to find analysts in political science assuming a Gaussian distribution when event counts make up the dependent variable. Everyone knows that event counts are not distributed normally, and computer programs and statistical practice acknowledge this fact. The popularization of event count models has led to their widespread application in many areas. Many event count analyses use cross-sectional data. In these cases, heterogeneity and not dynamics are a primary concern.

However, many event counts are time series. Many scholars have begun using event count models to analyze the large sets of data on cooperative and conflictual international events (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; Remmer 1998). The same issues confront the smaller number of comparativists using this data, such as Moore, Lindstrom, and O'Regan (1996).

For time series of event counts, general contagion or dynamics are either ignored in practice or are accounted for by the use of a lagged dependent regressor in the exponential functional form of a Poisson, negative binomial, or Generalized Event Count regression.

Standard methods for modeling the dynamics of a time series assume a normal distribution, so most analyses of event count data pay relatively little attention to the time series properties of the series they have considered. These approaches, lagged dependent variables in event count models and standard (Gaussian) time series models are incorrect for event count time series. Not thinking about the dynamics in event count series can be short-sighted because it leads to incorrect inferences.

The need to consider the time-series properties of event-count data is an important issue confronting a growing body of research in political science. The specification, estimation, and testing of non-linear, non-normal time series models for event counts is not part of the standard tool-kit of political scientists. When faced with time series of counts, many political scientists employ the normality assumption to model dynamics. This approach ignores the basic distributional information about the data in an effort to model the dynamics, resulting in inefficiency. Alternatively, ignoring dynamics also introduces inefficiency. Given these possible inefficiencies, we know that a time-series count data model will produce more efficient estimates than models that use an improper Gaussian assumption or fail to model the dynamics.

1.1 Approaches to Modeling Time Series Event Count Data

Event count data arise from a discrete distribution. Recall, the standard approach to modeling event count data is to assume that the events are generated from a Poisson density (King 1989a). In this model, the probability of observing the count y_t is given by a Poisson distribution with mean arrival rate λ:

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

Using maximum likelihood techniques, one can compute an estimate of the Poisson mean parameter λ. When $\lambda = \exp(X_t\beta)$ the resulting model 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 *over-dispersed.* When the events are over-dispersed, then alternative estimators have been proposed. Among these are the negative binomial and generalized event count (GEC) estimators (King 1989a).

When event counts are a time series, these methods are invalid. For time series data, we typically wish to estimate parameters that characterize the dynamics of the data. Further, when event counts are temporally correlated they almost always are over-dispersed. Since the Poisson regression model assumes independence, it is clearly invalid. The negative binomial model is also inappropriate, since it only accounts for the increase in the variance, but cannot characterize the dynamic properties of the time series of event counts.

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. Suppose that y_t is a time series of event counts. If we model y_t assuming that the errors are normally distributed, then such a series can be modelled easily using existing time series methods, such as ARIMA. To model such data, let Y_{t-1} denote the history of y_t up to period $t-1$. We can then compute the expectations $E[y_t|Y_{t-1}]$ and $V[y_t|Y_{t-1}]$ using standard results (e.g. Hamilton 1994). When these expectations are used to model the series and its covariates, we can used these simple expectations to characterize the distribution and its moments. Inference about the parameters then follows standard techniques.

If y_t is an event count, then ARIMA models assuming normally distributed errors are flawed in three ways. First, unless the values of the observations are very large, the event count distribution cannot be accurately approximated by a normal distribution. Second, the error term for the model must be strictly positive so that conditional on the mean at time t, the distribution of y_t is conditionally Poisson or negative binomial. Third, modeling event counts with a Gaussian distribution leads to bias and inefficiency (King 1988).[1]

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.[2] These models, however, 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 Poisson, negative binomial, and GEC models are appropriate for only a very specific type of data.

For example, suppose that z_t is Poisson distributed with mean $\mu_t = \exp(X_t\delta + \rho z_{t-1})$ and that X_t are independent and identically distributed. The lagged event count model is appropriate only for series with exponential growth rates and no dynamics. The growth rate of this lagged Poisson regression model is given by

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

Taking expectations gives

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

If $\rho \neq 0$ and $E[z_{t-1} - z_{t-2}] \neq 0$, this model implies a non-zero growth rate for the conditional mean. The interpretation of ρ then becomes problematic, since it is a growth rate rather than an autocorrelation or discounting coefficient. 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, this model has limited applicability.

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

[2] This method is used in Poisson, negative binomial or generalized event count regressions by Pollins (1996), Senese (1997), Spriggs and Wahlbeck (1995).

A second problem is that a series generated by a deterministic trend will have infinite variance. This happens because the series grows indefinitely. Thus, the variance is undefined.[3] Unlike the linear case, where including a determinstic trend solves this problem, in the nonlinear case, inclusion of a deterministic trend does not, in general, solve the infinite variance problem (Hamilton 1994: Chapter 15).

1.2 Time series models of for event count data

A variety of time series models for event count data can be found in the statistics and econometrics literature. Davis, Dunsmuir and Wang (1998) classify these models as either "observation driven" or "parameter driven" models. Observation driven models assume that the dynamic process can be described by the observed data. In observation driven models, the current observation can be directly modelled as a function of past values. Alternatively, parameter driven models assume that the dynamic process is a function of the latent parameters of the model. In these models, the latent parameters of the distribution (the mean λ in the Poisson distribution) are a function of an unobserved dynamic process. A third alternative approach defines new distributions and time series processes that generate event count data. These models can be classified as discrete or integer valued autoregressive models (INAR). In these models, a data generation process is described based on a generalization of autoregression for discrete random variables.

Both the observation and parameters driven models are usually estimated using a state-space or Kalman filter approach. Observation driven models are closely related to Bayesian forecasting and use a generalization of the Kalman filter state space model. Different observation driven models are discussed by Davis, Dunsmuir and Wang (1998a, 1998b), Harvey and Fernandes (1989), Kitagawa and Gersch (1996), West, Harrison and Migon (1985), West and Harrison (1997), and Zeger and Qaqish (1988). Latent parameter models for time se-

[3] This problem is a pecularity of frequentist probability and statistics. Bayesian inference is unaffected because of the way the sample is treated in Bayesian statistics.

ries event counts are considered in Zeger (1988), Diggle, Liang and Zeger (1994), Jackman (1998a), Shephard (1995). The integer valued time series approaches are addressed by Al-Osh and Alzaid (1987, 1990), and McKenzie (1988).

Some of these models present difficult numerical and estimation issues, or perform poorly when estimating parameters. Models that suffer from these problems include the integer valued time series models of Al-Osh and Alzaid (1987, 1990, Squier 1997). The latent parameter models use a generalized estimation equation approach (Diggle, Liang and Zeger 1994). These models are quasi-maximum likelihood methods, and may not be fully efficient. The smoothness priors framework of Kitagawa and Gersch (1996) and the Bayesian forecasting models of West et. al (1985, 1986) are similar in spirit to our approach, although they use an explicitly Bayesian framework.

Harvey (1989) details the specification and estimation of state-space models using the Kalman filter for the Gaussian case. Harvey (1989: 348-54, 418-22) and Harvey and Fernandes (1989) build on this approach to develop a state space model for event count data with a Poisson marginal distribution. We extend this model by correcting a misspecification of the transition equation in Harvey and Fernandes, using a result developed by Shephard (1994). The model we derive, the Poisson exponentially weighted moving average (PEWMA), is simple to specify and estimate. Maximum likelihood estimation does not require approximations or Monte Carlo integration (unlike many of the alternative approaches). The estimation of PEWMA models is as easy as standard Poisson or negative binomial models.[4]

In the next section, we outline the PEWMA model for persistent event counts. 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), Harvey (1989) and West and Harrison (1997). We then use

[4]In addition, we have written and are making available software written in GAUSS that can be used to estimate the model.

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 based on recent analyses of U.S. Supreme Court case agendas to illustrate how the model can be used for simple intervention analysis. The consequences of our analysis indicate that if one thinks that event counts have a strong dynamic, one should model the dynamics directly.

2 The Poisson Exponentially Weighted Moving Average Model

2.1 The Model

To address the problem of modeling persistent time series event counts, we use a structural time series model. In general, a structural time series model specifies two equations: 1) the measurement density or system equation and 2) the state density or transition equation to describe the evolution of a time series. The *measurement or system equation* describes the process that generates the observed data. The *state density or transition equation* defines the stochastic transition process for the parameters from period $t-1$ to period t. In the case where the error process is conditionally normal, this can be done using linear equations. In fact, the assumption of normality allows the use of these linear models to specify the distribution to the measurement and state variables in terms of their means and covariances. When the errors in both equations are assumed to be normal, the specification and estimation of the model is done using the Kalman filter (Harvey 1989, Hamilton 1994)

For event count data, specifying a structural time series model is more problematic. The problems with writing down a model where the distribution of event count y_t depends on past observations are identifying and estimating the model. To deal with these problems we adopt the approach of Harvey and Fernandes (1989). They employ natural conjugate densities to simplify the development of the model and numerical calculations. 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 the likelihood function. (Judge et. al 1988: 124; DeGroot 1970: 159). 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 (Judge et. al 1988: 123-4; DeGroot 1970: 164,167). Using natural conjugate densities to describe the unobserved parameters of the model has two benefits. First, it allows us to derive known closed form distributions for the model parameters. Second, in practice our model can be implemented without the use of Markov-Chain-Monte-Carlo methods or other numerical approximations for the density functions (cf. Jackman 1998a, 1998b). The result is a simple model that may be practically estimated to determine the presence of dynamics in time-series count data.

The model for time series count data is based on a Kalman filter setup with a Poisson measurement equation and a gamma distributed state equation. The model captures the changes in the mean arrival rate of the Poisson process at time t. 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 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, $Y_{t-1} = (y_0, y_1, ..., y_{t-1};\ X_0, X_1, ...X_{t-1})$, 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.[5] Throughout the paper, 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:** We assume the observed counts at time t are drawn from Poisson marginal distribution,

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

The unobserved mean arrival rate for the count at time t is μ_t. We assume that 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 $1 \times K$ matrix of explanatory variables (without a constant), and a time varying component μ_{t-1}^*. 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

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

[5] A hyperparmeter is a parameter of a prior distribution not fixed at particular numerical values (Gelman, Carlin, Stern, Rubin 1995: 36).

where $\eta_t \sim Beta\left(\omega a_{t-1}, (1-\omega) a_{t-1}\right)$ (Smith and Miller 1986; Shephard 1994). The parameter $0 < \omega \leq 1$ captures 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, $E[\eta_t] = \omega \; \forall \, t$. The parameter r_t describes the growth in the series and insures that $\mu_t > 0$.

3. **Conjugate Prior:** The prior distribution for the time varying component is a gamma distribution: $\mu_{t-1}^* \sim \Gamma\left(a_{t-1}, b_{t-1}\right)$. To identify the model in equations (1-3), we specify the gamma distribution as the conjugate prior for the distribution of y_t. This is the natural conjugate distribution for the Poisson parameter μ_t, since the domain of the Poisson and the range of the gamma distributions coincide and it is a sufficient statistic for μ_t. The gamma density f is given by

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

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}.

The transition equation determines how the previous history of the time series is reflected in the current observation. The variable μ_t is the conditional mean number of counts in period t. This is computed based on all the previous observations, from $t = 0$ to $t-1$. 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 periuod growth rate is the change in the mean number of counts in the period, $t-1$ to t. In a later section our characterization of the model shows that the expression for the expected growth rate is $E\left[\ln(\omega) - r_t\right] = 0$. 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 $E\left[\eta_t\right] = \omega$.

The covariates X_t, enter the model contemporaneous with the level of the time series at $t-1$, namely μ^*_{t-1}. 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 $100\left(\exp(\delta) - 1\right)$ percent change in y_t, just as in the Poisson model.

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 $t \to \infty$. [6] We avoid this problem by adopting the transition equation with the additional term r_t suggested by Shephard's (1994) work on local scale models.[7]

Estimation of the model requires a time dependent series of recursions for the conditional mean $E\left[\mu_t|Y_{t-1}\right]$ and the posterior mean $E\left[\mu_t|Y_t\right]$. These recursions define a Kalman filter for

[6] 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. A similar result is found in Grunwald, Hamza and Hyndman's (1997: Theorem 1) discussion of power steady models (of which this model is a special case). In their discussion of Harvey and Fernandes' model, Brockwell and Davis (1996) also note this problem, citing Grunwald, Hamza and Hyndman. The addition of the e^{r_t} term to the transition equation of Harvey and Fernandes (1989) implies that the series will have a zero growth rate.

[7] We caution readers who might use the Harvey and Fernandes (1989) result. Their transition equation provides an incorrect model of the stochastic process for the mean of event counts. The transition we use avoids this problem. So far as we know, this is the first application of Shephard's transition equation to a state space model for event count data.

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), and at time t (the posterior). 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.

2.2 Estimation and Forecasting

The PEWMA model is estimated using maximum likelihood techniques. Derivation of the likelihood function requires computing the predictive distribution of the time series of event counts under the assumptions of the model. A proper distribution for μ_t requires that $\mu_t \neq 0$ for $t = 0$. 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.[8] Constructing the likelihood function proceeds in two steps. First, we derive the predictive distribution for the event count series under the PEWMA model. Next, we use this to compute the log-likelihood function.

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

$$\begin{aligned}
\Pr\left(y_{\tau+1}, ..., y_T\right) &= \prod_{t=\tau+1}^{T} \Pr\left(y_t | Y_{t-1}\right) \\
&= \prod_{t=\tau+1}^{T} \underbrace{\Pr\left(y_t | \mu_t\right)}_{Measurement} \cdot \underbrace{\Pr\left(\mu_t | Y_{t-1}\right)}_{Transition} .
\end{aligned} \tag{4}$$

The key to understanding this predictive distribution is the decomposition of the posterior probability $\Pr\left(y_t | Y_{t-1}\right)$. Using the Kalman filter outlined in the appendix, 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

[8] 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 non-zero observation (Harvey and Fernandes 1989: 408). This can be easily implemented by removing the first $\tau - 1$ zero observations from the data.

gamma distribution). Note that these two components correspond to the measurement and transition equations.

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,

$$\begin{aligned}
\Pr(y_t|Y_{t-1}) &= \int_0^\infty \Pr(y_t|\mu_t)\Pr(\mu_t|Y_{t-1})\,d\mu_t \\
&= \frac{\Gamma(y_t+\omega a_{t-1})}{y_t!\Gamma(\omega a_{t-1})}(\omega b_{t-1}\exp(-X_t\delta - r_t))^{\omega a_{t-1}} \cdot \qquad (5)\\
&\qquad (1+\omega b_{t-1}\exp(-X_t\delta - r_t))^{-(y_t+\omega a_{t-1})},
\end{aligned}$$

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

Since the observed predictive distribution of the data is negative binomial, the variance is larger than the mean. Thus, rather than assuming that over-dispersion is caused by unobserved heterogeneity or unknown contagion in the data, a dynamic process produces the over-dispersion. Contrast this with King's (1989b) Generalized Event Count model in which a dispersion parameter is estimated to account for the unknown degree of over- or under-dispersion in the data. In our model, we estimate an additional parameter to account for the discounting of the past observations in the mean level of observations. For other ways to conceptualize over-dispersion in event data, see Winkelmann (1995).

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 in the link function based on (4):

$$\begin{aligned}
\ln L(\omega,\delta|a_{t-1},b_{t-1},y_t,X_t) &= \sum_{t=\tau+1}^{T} \ln\Gamma(y_t+\omega a_{t-1}) - \ln(y_t!) \\
&\quad - \ln\Gamma(\omega a_{t-1}) + \omega a_{t-1}\ln(\omega b_{t-1}\exp(-X_t\delta - r_t)) \qquad (6)\\
&\quad - (\omega a_{t-1}+y_t)\ln(1+\omega b_{t-1}\exp(-X_t\delta - r_t))
\end{aligned}$$

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.

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:

$$E\left(y_{T+1}|Y_T\right) = \frac{a_T}{b_T} = \frac{a_{T+1|T} + y_T}{b_{T+1|T} + 1} = \frac{\omega a_T + y_T}{\omega b_T + \exp\left(X_t \delta + r_t\right)}, \tag{7}$$

$$V\left(y_{T+1}|Y_T\right) = \frac{a_{T+1|T}\left(1 + b_{T+1|T}\right)}{b^2_{T+1|T}} = \omega^{-1} V\left(\mu_T|Y_T\right) + E\left(\mu_T|Y_T\right) \tag{8}$$

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:

$$\bar{y}_{T+1|T} = \exp\left(X_{T+1}\delta + r_{T+1}\right) \frac{\sum_{j=0}^{T-1} \omega^j y_{T-j}}{\sum_{j=0}^{T-1} \omega^j \exp\left(X_{T-j}\delta + r_{T-j}\right)} = \bar{\mu}_{T+1|T}. \tag{9}$$

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 $t = 1, ..., T$. This is the limiting form of the EWMA for $\mu_{t+1|t}$. The coefficients for the X_t vector can be interpreted in the standard way: $100(\exp(\delta) - 1)$ is the percentage change in the mean given the value μ_{t-1} and a unit change in X_t.

Multi-step prediction is also possible. Since this is not essential to the model, we include the discussion of the multi-step prediction function in Appendix B.

2.3 Description of the PEWMA Process

The PEWMA does not have a simple linear structure. Thus, it is less than obvious how the dynamics of the time series evolve over time. In this section we characterize the evolution of the time series by analyzing the properties of the transition equation (3) and the forecast function (9). Our characterization shows how the mean of the process, μ_t evolves over time.

We can show that the transition equation and the forecast function for the PEWMA imply that the mean evolves according to a random walk process. Via some simple calculations, we can derive the theoretical growth rate for the model. Recall that the transition equation is written

$$\begin{aligned} \mu_t &= e^{r_t}\mu_{t-1}\eta_t, \\ so \;\; \ln(\mu_t) - \ln\left(\mu_{t-1}\right) &= r_t + \ln(\eta_t) \end{aligned} \tag{10}$$

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 $E\left[\mu_{t-1}|Y_{t-1}\right] = E\left[\mu_t|Y_{t-1}\right]$. Thus, the expected growth rate is,

$$0 = E\left[r_t\right] + E\left[\ln(\eta_t)\right]. \tag{11}$$

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.[9]

[9] We conjecture that events will not be explosive because with fixed time slices, only so many events can

How then 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).

We would like to be able to diagnose when to use the PEWMA model. Grunwald, Hyndman Tedesco and Tweedie (1997) show that standard autocorrelation function computations can be used to diagnose a linear autoregression process for event counts. Brandt, Williams and Fordham (1998) and Brandt and Williams (1998a) demonstrate that standard autocorrelation function calculations are good diagnostic tools for time series of count data that follow a more general autoregressive process. 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 (ACF) for the series. The final graph shows the ACF for the log difference of the mean, μ_t.

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 about here>

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 mean zero process (Graph 2).[10] 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.

3 Monte Carlo Experiments

We motivated the PEWMA model by arguing that existing ARIMA models could not handle data that is generated from discrete event count processes such as the Poisson or negative binomial. However, it could be argued that existing methods such as Poisson or negative binomial regression could handle the dynamic patterns in some event count data. For time series of counts these estimators are pseudo-maximum likelihood estimators and will generally be less efficient than the true model (Gourieroux, Monfort and Trognon, 1984a, 1984b, and White 1994). We know (from ordinary least squares), however, that failing to account for serial dependence leads to incorrect standard errors. However, the effects of misspecified dynamics in event count models have not been adequately explored. Deriving results about the effects of time dependence for Poisson and negative binomial regression under temporal misspecification is difficult because of the highly non-linear nature of the estimators. Brandt and Williams (1998b) show that the effects of dynamic misspecification in event count models lead to complicated estimators that make comparisons difficult. Since these analytic results present significant complications, we turn to Monte Carlo experiments to demonstrate the conditions under which the PEWMA is more efficient.

In order to assess the properties of this time series model for count data, we present a series of Monte Carlo experiments. These experiments were conducted to investigate several questions:

[10] For Graph 2 of Figure 1, $E\left[\ln(\mu_t) - \ln(\mu_{t-1})\right] = -0.070$ and $Var\left[\ln(\mu_t) - \ln(\mu_{t-1})\right] = 0.57$.

1. What are the convergence properties of the model? That is, how sensitive are the PEWMA parameter estimates to sample size and how does this affect the ability to estimate the time series dynamics relative to alternative specifications?

2. What are the distributional properties of the model specification? Poisson and negative binomial models converge to normality rather rapidly as the mean grows. Given the cost of estimating, specifying and interpreting generalized linear models and structural time series models we want to know under what conditions a Gaussian model might provide approximately correct inferences.

3. What are the implications for the misspecification of dynamics for the efficiency of estimators in Poisson and negative binomial regression? We know from King (1988) that failure to account for the Poisson density assumption for count data leads to bias. What is less obvious is the degree of efficiency loss due to misspecified temporal dynamics.

Our Monte Carlo design includes a series of 27 experiments. We conduct experiments based on data generated from the PEWMA model. 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 3 different values of $\mu_0 = 10, 20, 50$. We also investigated sample sizes of $T = 50, 100, 200$. Finally, we varied the dynamics so that $\omega = 0.2, 0.4, 0.6, 0.8$. The mean number of counts and sample sizes are reflective of data that is typically analyzed in political science applications.

We were not able to produce Monte Carlo results for all experiments with $\omega = 0.2$. For values of ω near zero, the dynamics become so strong that the data generating process is non-stationary. 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. Thus, generating pseudo-random series of PEWMA counts is very difficult for "long" time series or "large" mean counts when $\omega \leq 0.2$. Thus, 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. In fact, few "real" data sets that we have analyzed with the PEWMA model have values of ω less than 0.4. When ω is small, the data tends to be very non-stationary. 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. Since these cases are not consistent with observed time series of counts, we omit them.

For each Monte Carlo experiment, 200 replications were conducted. In each replication, a single covariate X was specified and held constant throughout the experiment. Each element of this $T \times 1$ vector X_t is distributed $N(0,1)$. The vector X_t is fixed and not stochastic.

The data generating process (DGP) we use is based on the PEWMA model:

$$\begin{aligned} \Pr(y_t|\mu_t) &= \frac{e^{-\mu_t}\mu_t^{y_t}}{y_t!} \\ where\ \mu_t &= \mu_{t-1}e^{-r_t}\eta_t \\ and\ \mu_1 &= \frac{a_0}{b_0}\exp(X_1\delta_1 + r_1)\eta_1 \end{aligned}$$

The model does not contain a trend. 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 $\mu_0 = \frac{a_0}{b_0}$. 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}.[11] The value of δ_1 is fixed at 0.5 for all experiments.

Once we generated the data for each replication, we estimated 7 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 models based on Gaussian errors to assess the effects of distributional misspecification. We estimate two Gaussian models, one with the following form:

$$\begin{aligned} \ln y_t &= \delta_0 + \delta_1 X + \epsilon_t \\ with\ \epsilon_t &= u_t - \rho u_{t-1}\ and\ u_t \sim N\left(0, \sigma_u^2\right). \end{aligned}$$

[11] 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 $y_t \sim Po(\mu_t)$, $t = 1, ...T$. Such a process fails to account for the past realizations of $Y_{t-1} = (y_0, y_1, ...y_{t-1})$.

The second Gaussian model has the form:

$$\begin{aligned} \ln y_t &= \rho \ln y_{t-1} + \delta_0 + \delta_1 X + \epsilon_t \\ with\ \epsilon_t &\sim N\left(0, \sigma_\epsilon^2\right) \end{aligned}$$

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 ρ 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).[12]

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

$$\begin{aligned} \lambda_t &= \exp\left(\delta_0 + \delta_1 X_t\right) \\ and\ \lambda_t &= \exp\left(\delta_0 + \delta_1 X_t + \rho y_{t-1}\right) \end{aligned}$$

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 (1989c).

After estimating these models, we saved the coefficients, standard errors and several other diagnostics. Standard errors for MLE methods were estimated using the outer-product of the gradient method, inversion of the numerical Hessian, and the Huber-White heteroscedastic consistent methods. The results we report here are based on the Huber-White standard errors. We choose to use Huber-White standard errors not because they are appropriate because in fact the data does not contain heterogeneity. Rather, we choose these because

[12] 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.

almost all applied work in a time series context use these and they are widely reported in political science applications.

Tables 1 and 2 present Monte Carlo estimates for the single covariate parameter δ_1 and the Monte Carlo standard deviation for δ_1 for the models outlined above. As with any maximum likelihood estimator, the PEWMA performs better in large samples. As the sample size grows, the Monte Carlo standard deviations of the estimated regression parameter, δ_1, fall and the overall fit of the model improves. In small samples ($N = 50$), the PEWMA model may perform worse than the corresponding Poisson and negative binomial models. The PEWMA model performs well in almost all cases: when there are more dynamics (i.e. $\omega \to 0$), it produces more efficient estimates of the regression parameters than the other models.

<Tables 1 and 2 about here>

Based on these results, there appears to be a slight upward bias in the estimates of the regression parameter δ_1 for the two Gaussian models when $\omega = 0.6$ or 0.8. In addition, the two Poisson and negative binomial specifications yield estimates of δ_1 that are slightly downwardly biased. These biases, however, are not very large. In general, these estimates show the consistency of the PEWMA estimates to the true value. Since the other estimates are in the same linear exponential family of estimators, they are also consistent (Gourieroux, Monfort, and Trognon 1984a).

The PEWMA is also more efficient than the other models for estimates of the regression parameter, δ_1. 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 non-linear 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? The estimated Monte Carlo standard deviations for δ_1 in Table 2 and the relative efficiency measures in Table 3 present evidence about the degree of inefficiency of alternative models. The relative efficiency is computed as

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

Relative efficiency values are ratios of the standard deviations computed from the Monte Carlo estimates of δ_1, and thus are positive real numbers. A value of 1 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 relative efficiency value provides a measure of how many times larger (smaller) the alternative estimators' Monte Carlo standard deviations are, relative to the true value (estimated from the Monte Carlo estimates of δ_1 in the PEWMA model).

<Table 3 about here>

The pseudo-maximum likelihood estimators have varying performance. As T gets larger, the PEWMA model performs better than all the alternative models. 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. Reading down the tables of reported standard deviations and relative efficiencies, as the sample size grows the alternative models perform *worse* than the true model in all cases (for a given value of ω) at estimating the true value of δ_1. The degree of this relative inefficiency is striking: as the sample size and the mean increase, the relative efficiency of the estimated regression parameter δ_1 is well over 1.5 times larger than the Monte Carlo standard deviation of the true model (i.e. the PEWMA). When $\omega = 0.8$, the Poisson and negative binomial models have *smaller* Monte Carlo standard deviations than the PEWMA for means less than 20 and sample sizes smaller than 100. This is a small sample bias of the PEWMA. As the mean and sample size grow, the relative efficiency of the Poisson and negative binomial models is worse: for time series of 200 observations, the Poisson model is always at least two time less efficient than the PEWMA; the lagged Poisson, negative binomial and lagged negative binomial models have Monte Carlo standard deviations that are 1.5 to 5 times the Monte Carlo standard deviations of the PEWMA, depending on the value of ω.

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 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 for accounting for time series properties.[13]

The implication of estimating incorrect standard errors with Poisson and negative binomial regressions should not be understated. The results in Tables 2 and 3 demonstrate that the degree of inefficiency is large for a variety of sample sizes, means, and varying degree of dynamics. This means that the presence of *any* dynamics in the count data series being modelled can lead to incorrect inferences about the regression parameters. The results in

[13] 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 ($H_0 : \omega = 1; H_A : \omega < 1$). We note however, that there is a small upward bias in the small sample estimates of ω and this parameter should be interpreted carefully. The alternative models perform poorly when estimating the dynamics. It would appear that using a lagged dependent variable is a poor way to account for dynamics generated from a PEWMA process. The use of these lagged dependent variables performs poorly precisely because there is no growth in these series. Details of these results are available upon request.

Brandt and Williams (1998b) suggest that there could be inefficiency if standard Poisson or negative binomial models are used on time series count data, based on the analytic derivations of the asymptotic covariances for the regression parameter for the Poisson, negative binomial and PEWMA. The results of this section show that the inefficiency of the Poisson and negative binomial model is large, even when a lagged dependent variable is used.

These Monte Carlo results are at best suggestive. If one were to use standard statistical packages and estimation methods, how incorrect would inferences be about the regression parameters of an event count model? Since positive serial correlation induces reported standard errors that are smaller than their true values in time series models, we would like to know what are the implications of failing to account for the dynamics. 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. The relative overconfidence of the estimator is computed as,

$$\text{Relative Overconfidence} = \frac{\text{Monte Carlo Standard Deviation of } \delta_1}{\text{Mean Reported Huber-White Standard Error of } \delta_1}.$$

Relative overconfidence values of 1 indicate that the Huber-White robust standard errors are equal to the Monte Carlo standard deviations for the regression parameters. Relative overconfidence values greater than one indicate that the Huber-White robust standard errors are smaller than the true Monte Carlo standard deviations for the regression parameters. Overconfidence values less than one indicate that the reported standard errors are "too small."

Table 4 presents the average relative overconfidence for each of the estimators used in each of the Monte Carlo experiments. The PEWMA has overconfidence values equal to 1 for almost every experiment. Thus, the reported Huber-White standard errors are very close (on average) to the estimated Monte Carlo standard deviations for the regression parameter. Consistent with time series results on the effects of serial correlation, the standard Poisson regression model has varying relative overconfidence for the regression parameter δ_1. In samples where $T = 50$, and when the hyperparameter $0.4 \leq \omega \leq 0.8$, the Poisson model's robust standard errors are nearly identical to the Poisson Monte Carlo standard deviations. Similar patterns can be seen for the relative overconfidence of the lagged Poisson, negative binomial and lagged negative binomial estimates of the standard error of δ. However, in larger samples (i.e. $T = 100, 200$), the relative overconfidence shows a different pattern. 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.

<Table 4 about here>

While these results depend on the design of the Monte Carlo experiments and the number of replications, the variation in the relative overconfidence and relative efficiency of estimated regression parameters for time series of counts (based on the estimators outlined above) is cause for caution. Using "lagged dependent variables" and robust standard errors are at best imprecise corrections for time series count data models, since the Monte Carlo results include these models with disappointing results. Robust standard errors do not correct the inefficiency problem since their validity generally depends on the absence of dynamic misspecification (White 1994: Chapter 8). The poor performance of robust standard errors is demonstrated by the size of the Huber-White standard errors in our analysis.[14]

[14] We also estimated standard errors based on the outer-product of the gradient (the "BHHH" estimator)

We believe that these results are general and call for additional research and discussion. In addition, the PEWMA model of Section Two (and the accompanying Monte Carlo evidence) is but one model for time series of counts. A more general approach could be derived along the lines of ARIMA-type models for count data. A more general specification raises issues of model selection (MA versus AR processes), as well as the order of the model. These issues clearly call for additional research.[15]

4 Application: U.S. Supreme Court Case Agendas

To demonstrate the usefulness of time-series count models such as the PEWMA, we reanalyzed an updated series of the Supreme Court agenda data discussed in Pacelle (1991,1995) and Pacelle, Farole and Williams (1998). An emerging area of scholarship on the Supreme Court looks at the nature of agenda change over time. Changes in the Court's agenda may result from new membership, landmark decisions, leadership changes, and a new role for the institution. Of great interest is the nature of the dynamic process that accompanies agenda change.

Pacelle (1991) and Pacelle et al. (1998) argue that significant agenda change occurred in two main areas of the Court's agenda since the 1950s. First, there was a significant increase in the number of cases concerning issues of Equality and the 14th Amendment (e.g. race, gender, age, disability). Second, there was a decrease in the number of cases involving Regulation (e.g. the regulation of business by the federal government). These changes in agenda composition and dynamics resulted from three major changes in the Court. The first was the shift to civil rights and civil liberties issues in the wake of the civil rights movement and the Court's decision to abandon efforts to interfere in regulatory policy after Justice

and the inverse of the numerical Hessian (the classic empirical information matrix). For these estimators, the performance of the standard errors is rarely better and generally worse than the standard deviations and relative efficiencies we report in Tables 2 and 3.

[15] Brandt, Williams and Fordham (1998) addresses some of these more advanced specification issues. Autoregressive models and diagnostics are addressed in Brandt and Williams (1998a)

Harlan Stone's famous Footnote Four in *U.S. v. Carolene Products.* More significantly in 1953, the appointment of Earl Warren as Chief Justice altered the leadership of the Court. Warren's leadership is viewed as a significant change from that of his predecessor Fred Vinson. Finally, the Court delivered the landmark decision in *Brown v. Board of Education* during the 1953 term.

Pacelle et al. (1998) test agenda change in the Equality and Regulation cases using time series methods.[16] Their dependent variable is the proportion of the Court's agenda devoted to these two policy areas from 1933 to 1992. They use vector autoregression (VAR) to determine the effect of the 1953 term on the percentage of the agenda allocated to these two areas. It is hypothesized that the Court's increased attention to Equality comes at the expense of Regulation after 1953. Since lag length for the intervention effect cannot be specified *a priori*, they use a deterministic dummy variable to model the effect of the 1953 term and examine the effect of this variable over time.

In our replication, we utilize the number of cases in the Equality and Regulation agenda areas to which the U.S. was a party from 1933 to 1993. Using the counts of the number of cases is preferred because the percentages are influenced by the number of cases that the Court accepts in all areas. The more cases the Court accepts in other areas, the less certain we can be of the changes, since they will affect the denominator in the calculation of the proportion. Thus the actual counts of the cases in each area present a better measure of the agenda.

Based on this data, we are interested in determining the changes in the Court's agenda after the 1953 term. Our expectation is that in the post-1953 period, we should see a rise in the number of cases dealing with issues of Equality (racial, gender, etc.) and a decline in the number of cases dealing with economic Regulation. For our analysis we perform an intervention analysis of the number of cases in the economic Regulation and Equality agendas

[16] See Pacelle (1991) for a description of the data and coding.

before and after 1953.

Each of these series exhibits a strong degree of temporal dependence. Figure 2 plots the Equality and Regulation series and their sample autocorrelation functions. Note that the number of cases involving Equality rises sharply after 1953 (specifically after 1960), and in contrast the number of regulation cases falls, although not as dramatically. The sample autocorrelation functions for the two series demonstrate persistence in the data. The Regulation series has six large spikes, while the Equality series has twelve large spikes.

<Figure 2 about here>

Of interest is whether the decline begins in 1953 for the Regulation cases and whether the rise in Equality cases begins immediately after *Brown* or whether there was a lag. Of additional interest is the year in which the lag intervention occurred. We model each series using the PEWMA model, lagged Poisson regression, lagged negative binomial regression and a Gaussian ARIMA(0,0,1) model each with a temporary intervention. The temporary intervention is used because we believe its effect raises the number of cases in the Equality series (decreases the number of cases in the Regulation series). Such temporary effects, however, should have a permanent effect on the level of the series (see Harvey 1991: 397-99). Since we are not sure of the lag specification for the intervention (i.e. how long after *Brown* the increase/decrease in cases occurred), we empirically determined the intervention specifications from a one period to an 10 period forward lag (1954-1963). The optimal intervention specification will maximize the log-likelihood of the model.

In Table 5, the results for the various models are presented. The coefficient for the dispersion term in the negative binomial model is γ, and the autocorrelation parameter in the models is ρ. Using the PEWMA, the lag specification with the highest likelihood value for the series was 10 lags for the Equality series (1963) and one lag for the Regulation series (1954). These are consistent with the VAR results reported in Pacelle et al. (1998). The coefficient in the PEWMA model for the intervention was 0.9024 for the Equality series and -0.7165 for the economic Regulation series. In the Regulation case the intervention has a negative effect as predicted and a large absolute t-ratio greater than three. For the equality series, a similar conclusion holds.

<Table 5 about here>

The results of the lagged Poisson, lagged negative binomial and Gaussian ARIMA models tell a similar story. If we were to use the Poisson regression, negative binomial or ARIMA models, we would determine that the intervention effect occurred in the same periods that we found from using the PEWMA model for each series. However, we would conclude that the intervention has no effect in the Equality series based on the ARIMA model. Based on the Poisson and negative binomial models, the intervention does have an effect on the Equality series. For the Regulation series, the negative binomial model shows the intervention to be smaller and has a larger reported standard error than either the PEWMA or Poisson models. More troubling is that we would reject the negative binomial model for the Regulation series, since the value of the over–dispersion parameter (γ) is not significantly different from one. These differing conclusions highlight both the need to get the distributional and dynamic aspects of the specification correct.

Note also that the effect of the intervention captures a large part of the change in the series. The intervention effect for the PEWMA model for the Equality case agenda implies that there was a 147 percent increase in the number of cases after 1963. The intervention coefficient from the PEWMA model for the Regulation series implies that there was a 51 percent decrease in the number of cases in the year after *Brown*. This is consistent with the data. However, the other models fail to account for the continuing increase or decrease in cases after the intervention.

For the Regulation agenda cases, the Poisson and, negative binomial and Gaussian

ARIMA models all predict around a 45 percent decrease in the number of cases after 1953. However, these models systematically underestimate the number Regulation cases prior to 1953 and overestimate the number of cases after 1953. The reverse happens with the Equality series before and after 1963. Prior to 1963, the Poisson, negative binomial and Gaussian ARIMA models overstate the mean of the series. After the intervention, these models underestimate the number of Equality cases.[17] The Poisson model for the Equality series implies a 218 percent effect for the (1963) intervention; the negative binomial, a 229 percent increase and the Gaussian ARIMA model, a 178 percent increase. These effects are clearly at odds with the observed data.

This reanalysis of the Supreme Court case agenda makes several points about modeling dynamics in time series of event counts. First, the PEWMA model outperforms the alternative models based on the value of the log-likelihood. Second, the PEWMA offers a good prediction for real, persistent time series such as the Supreme Court case agenda on Equality and Regulation. Finally, the PEWMA offers a more sensible estimate of the intervention effects of the 1953 term.

5 Conclusion

The PEWMA model offers a straightforward and easily estimable method for modeling event count data with persistent dynamics. 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

[17] The relative over- and underestimates of the series can be seen by computing the prediction errors. Prior to the intervention, the regulation series has a positive mean prediction error and a negative mean prediction error after the intervention. The reverse occurs with the Equality series.

will be much too small. This means that scientific judgments can be seriously flawed.

Our experience is that many time series event count 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. A third alternative is that the data have dynamics but are mean reverting. The proper specification for such a case is an autoregressive model (see Brandt and Williams 1998a). 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.

A PEWMA Filtering Algorithm

The PEWMA utilizes a version of the Kalman filter to estimate the time varying mean for event count data. This filter can be derived using the three basic assumptions that define the PEWMA model.

The filter must compute three quantities of interest: the unconditional mean of the process $\left(E\left[\mu_{t-1}|Y_{t-1}\right]\right)$, the conditional mean of the process $\left(E\left[\mu_t|Y_{t-1}\right]\right)$, and the posterior mean of the process $\left(E\left[\mu_t|Y_t\right]\right)$. Similar to Gaussian ARIMA models the Kalman filter for count data should require (1) $E\left[\mu_t|Y_{t-1}\right] = E\left[\mu_{t-1}|Y_{t-1}\right]$, and (2) $Var\left[\mu_t|Y_{t-1}\right] > Var\left[\mu_{t-1}|Y_{t-1}\right]$.[18] 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 $\Pr\left(\mu_t|Y_{t-1}\right)$. 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).

From the transition equation, and the properties of the gamma distribution, conditional on Y_{t-1} we can compute the distribution of $\mu_t|Y_{t-1}$. First, substituting the gamma prior into the transition equation yields

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

Conditional on Y_{t-1}, μ_t is gamma distributed and we write $\mu_t|Y_{t-1} \sim \Gamma\left(a_{t|t-1}, b_{t|t-1}\right)$.

From the properties of the gamma distribution, $\mu_t|Y_{t-1}$ has a gamma distribution with

[18] Note that this justifies using the local level from the last period in the multiplicative function for the mean μ_t.

parameters $a_{t|t-1}$and $b_{t|t-1}$ such that

$$a_{t|t-1} = \omega a_{t-1} \tag{12a}$$

$$b_{t|t-1} = \omega b_{t-1} \exp\left(-X_t\delta - r_t\right), \quad \omega \in (0,1]. \tag{12b}$$

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

$$E\left[\mu_t|Y_{t-1}\right] = \frac{a_{t|t-1}}{b_{t|t-1}} = \frac{a_{t-1}}{b_{t-1}\exp\left(-X_t\delta - r_t\right)} = E\left[\mu_{t-1}|Y_{t-1}\right] \tag{13}$$

$$Var\left[\mu_t|Y_{t-1}\right] = \frac{a_{t|t-1}}{b^2_{t|t-1}} = \frac{\omega a_{t-1}}{\omega^2\left(b_{t-1}\right)^2\left(\exp\left(-X_t\delta - r_t\right)\right)^2} = \omega^{-1}Var\left[\mu_{t-1}|Y_{t-1}\right]. \tag{14}$$

Thus, 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.[19]

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

$$\begin{aligned} \mu_t &= e^{r_t}\mu_{t-1}\eta_t \\ \ln\left(\mu_t\right) - \ln\left(\mu_{t-1}\right) &= r_t + \ln\left(\eta_t\right) \end{aligned} \tag{15}$$

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 $E\left[\mu_{t-1}|Y_{t-1}\right] = E\left[\mu_t|Y_{t-1}\right]$ and the model has a zero growth rate. Therefore, since η_t follows a beta distribution, this expectation can be evaluated using a standard integral formula:

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

[19] 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 intercept term of the Poisson regression model.

$$
\begin{aligned}
&= -\frac{1}{B\left(\omega a_{t-1},(1-\omega)\,a_{t-1}\right)}\int_0^1 \eta_t^{\omega a_{t-1}-1}\left(1-\eta_t\right)^{a_{t-1}(1-\omega)-1}\ln\left(\eta_t\right)d\eta_t \\
&= \Psi\left(a_{t-1}\right)-\Psi\left(\omega a_{t-1}\right),
\end{aligned}
$$

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, $\Pr(y_1...y_T|\omega,\delta,X_t)$. We compute the posterior using Bayes rule. Given the observed count y_t 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:

$$
\begin{aligned}
\Pr\left(\mu_t|Y_t\right) &= \frac{\Pr\left(Y_t|\mu_t\right)\cdot\Pr\left(\mu_t\right)}{\int_0^\infty \Pr\left(Y_t|\theta\right)\cdot\Pr\left(\theta\right)d\theta} \\
&= \frac{e^{-\mu_t\left(1+b_{t|t-1}\right)}\mu^{y_t+a_{t|t-1}-1}\left(1+b_{t|t-1}\right)^{-\left(y+a_{t|t-1}\right)}}{\Gamma\left(y_t+a_{t|t-1}\right)}.
\end{aligned}
$$

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

$$
a_t = a_{t|t-1}+y_t=\omega a_{t-1}+y_t \;\; and \tag{16a}
$$

$$
b_t = b_{t|t-1}+1=\omega b_{t-1}+\exp\left(X_t\delta+r_t\right), \tag{16b}
$$

where (16a) 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 (12a-16b). These filter parameters can then be used to compute the predictive distribution and maximum likelihood estimates in equations (5 and 6).

B PEWMA Multi-Step Forecasts and Predictions

This appendix details several technical results concerning the forecast function for the PEWMA model. While these results are not necessary to estimate or interpret the model, we present them for completeness.

Multi-step prediction requires integration of the conditional density function, equation (5). 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

$$\Pr\left(y_{T+\ell}|Y_T\right) = \int_0^\infty \Pr\left(y_{T+\ell}|\mu_{T+\ell}\right) \cdot \Pr\left(\mu_{T+\ell}|Y_T\right) \partial\mu_{T+\ell}. \tag{17}$$

Simply using the recursions (12a-16b) ignores the evolution of μ_t for observations at $T+1, T+2, \ldots, T+\ell-1$. Thus, accounting for the conditional distribution of μ_t in these interim periods is accomplished by rewriting (17) as

$$\Pr\left(y_{T+\ell}|Y_T\right) = \sum_{y_{T+\ell-1}} \cdots \sum_{y_{T+1}} \prod_{j=1}^{\ell} \Pr\left(y_{Y+j}|Y_{T+j-1}\right). \tag{18}$$

In practice, equation (18) can be evaluated numerically. As Harvey and Fernandes note, closed form solutions for $\Pr\left(y_{T+\ell|T}\right)$ for $\ell > 1$ can also be evaluated numerically.

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

$$E\left[y_{T+\ell}|Y_T\right] = \frac{a_T}{b_T} \exp\left(X_{T+\ell}\delta + r_{T+\ell}\right) \quad \forall \ell \geq 1.$$

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).

References

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

Al-Osh, M.A. and A.A. Alzaid. 1987. "First-Order Integer-Valued Autoregressive (INAR(1)) Process." *Journal of Time Series Analysis* 8(3):261–275.

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(4):636–657.

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

Brandt, Patrick T. and John T. Williams. 1998b. "Analytical Covariances for Pseudo-Maximum Likelihood Estimators of Event Count Time Series." Unpublished manuscript, Indiana University.

Brandt, Patrick T., John T. Williams and Benjamin Fordham. 1998. "Modeling Time Series Count Data: A State-Space Approach to Event Counts." presented at the Annual Meeting of the Society for Political Methodology, San Diego, California, July 22-26, 1998.

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(1):196–210.

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. 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.

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.

Eyerman, Joe and Robert A. Hart. 1996. "An Empirical Test of the Audience Cost Proposition." *Journal of Conflict Resolution* 40(4):597–616.

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

Goldstein, Joshua S. 1991. "Reciprocity in Superpower Relations: An Empirical Analysis." *International Studies Quarterly* 35(2):195–209.

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

Gourieroux, C., A. Monfort and A. Trognon. 1984*a*. "Pseudo Maximum Likelihood Methods: Theory." *Econometrica* 52(3):681–700.

Gourieroux, C., A. Monfort and A. Trognon. 1984*b*. "Pseudo Maximum Likelihood Methods: Applications to Poisson Models." *Econometrica* 52(3):701–720.

Gowa, Joanne. 1998. "Politics at the Water's Edge: Parties, Voters, and the Use of Force Abroad." *International Organization* 52(2):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(3):615–626.

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

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

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(4):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(6):35–58.

Jackman, Simon. 1998a. "Analyzing time series counts of international disputes." presented at the Annual Meeting of the Midwest Political Science Association, April 1998.

Jackman, Simon. 1998b. "Time Series Models for Discrete Data: solutions to a problem with quantitative studies of international conflict." presented at the Annual Meeting of the Society for Political Methodology, San Diego, California, July 22-26, 1998.

Judge, George G., R. Carter Hill William E. Griffths Helmut Lutkepohl and Tsoung-Chao Lee. 1988. *Introduction to the Theory and Practice of Econometrics, 2nd Ed.* New York: John Wiley.

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

King, Gary. 1989*a*. "Event Count Models for International Relations: Generalizations and Applications." *International Studies Quarterly* 33(2):123–147.

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

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

Kinsella, David. 1995. "Nested Rivalries: Superpower Competition, Arms Transfers, and Regional Conflict, 1950-1990." *International Interactions* 21(2):109–125.

Kinsella, David and Herbert K. Tillema. 1995. "Arms and Aggression in the Middle East: Overt Military Interventions." *Journal of Conflict Resolution* 39(2):306–329.

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

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

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

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(4):335–63.

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(2):320–35.

Pacelle, Richard L. 1991. *The Transformation of the Supreme Court's Agenda: From the New Deal to the Reagan Administration*. Boulder: Westview Press.

Pacelle, Richard L. 1995. The Dynamics of Agenda Change in the Rehnquist Court. In *Contemplating Courts*, ed. Lee Epstein. Washington D.C.: Congressional Quarterly chapter 11.

Pacelle, Richard L. Donald Farole and John T. Williams. 1998. "The Dynamics of Supreme Court Agenda Change." Unpublished manuscript, Indiana University.

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

Remmer, Karen L. 1998. "Does Democracy Promote Interstate Cooperation? Lessons from the Mercosur Region." *International Studies Quarterly* 42(1):25–52.

Sayrs, Lois W. 1992. "The Effect of Provocation on Foreign Policy Response: A Test of the Matching Hypothesis." *International Interactions* 18(2):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(4):407–27.

Senese, Paul D. 1997. "Costs and Demands: International Sources of Dispute Challanges and reciprocation." *Journal of Conflict Resolution* 41(3):407–27.

Shephard, Neil. 1994. "Local scale models: State space alternative to integrated GARCH processes." *Journal of Econometrics* 60(2):181–202.

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

Smith, J.Q. and J.E. Miller. 1986. "Predictive records." *Journal of the Royal Statistical Society, B* 48(1):79–88.

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

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.

Volgy, Thomas J. and Lawrence E. Imwalle. 1995. “Hegemonic and Bipolar Perspectives on the New World Order.” *American Journal of Political Science* 39(4):819–34.

Ward, Michael D. and Sheen Rajmaira. 1992. “Reciprocity and Norms in United States-Soviet Foreign Policy.” *Journal of Conflict Resolution* 36(2):369–85.

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

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, P. Jeff Harrison and Helio Migon. 1985. “Dynamic Generalized Linear Models and Bayesian Forecasting (with comments).” *Journal of the American Statistical Association* 80(389):73–83.

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

Winkelmann, Rainer. 1995. “Duration Dependence and Dispersion in Count-Data Models.” *Journal of Business and Economic Statistics* 13(4):467–474.

Zeger, Scott L. 1988. “A regression model for time series of counts.” *Biometrika* 75(4):621–9.

Zeger, S.L. and B. Qaqish. 1988. “Markov Regression Models for Time Series: A Quasi-Likelihood Approach.” *Biometrics* 44:1019–31.

Table 1: Monte Carlo Estimates for regression parameter δ_1

Results based on 200 replications

N	μ_0	ω	PEWMA	LLOLS	ARIMA	Poisson	Lagged Poisson	Negative Binomial	Lagged Neg. Bin.
50	10	0.8	0.4982	0.5954	0.5993	0.4934	0.4927	0.4922	0.4922
50	10	0.6	0.5022	0.5411	0.5566	0.4861	0.4828	0.4758	0.4782
50	10	0.4	0.5016	0.5221	0.5491	0.4756	0.4766	0.4568	0.4661
50	10	0.2	0.5030	0.4863	0.5188	0.4763	0.4766	0.4529	0.4623
50	20	0.8	0.5013	0.5124	0.5154	0.4972	0.4963	0.4966	0.4958
50	20	0.6	0.5000	0.5067	0.5135	0.4979	0.4941	0.4933	0.4910
50	20	0.4	0.5014	0.4937	0.5076	0.4920	0.4877	0.4806	0.4808
50	20	0.2	0.4990	0.4901	0.5042	0.4896	0.4877	0.4785	0.4802
50	50	0.8	0.5012	0.5048	0.5056	0.4999	0.4995	0.4997	0.4993
50	50	0.6	0.5009	0.4994	0.5024	0.4979	0.4961	0.4963	0.4948
50	50	0.4	0.5008	0.5003	0.5038	0.5003	0.4974	0.4965	0.4943
100	10	0.8	0.4998	0.5834	0.5943	0.4965	0.4931	0.4884	0.4884
100	10	0.6	0.5059	0.5525	0.5877	0.4949	0.4912	0.4527	0.4666
100	10	0.4	0.5016	0.4908	0.5231	0.4948	0.4883	0.4563	0.4654
100	20	0.8	0.4849	0.5124	0.5183	0.4995	0.4975	0.4968	0.4957
100	20	0.6	0.4986	0.5059	0.5201	0.4956	0.4912	0.4794	0.4820
100	20	0.4	0.5014	0.4927	0.5134	0.4962	0.4913	0.4700	0.4766
100	50	0.8	0.4995	0.5012	0.5032	0.4993	0.4983	0.4984	0.4975
100	20	0.6	0.5002	0.4981	0.5028	0.4997	0.4971	0.4942	0.4929
100	20	0.4	0.5001	0.4923	0.5019	0.4989	0.4946	0.4870	0.4867
200	20	0.8	0.4995	0.5255	0.5308	0.5072	0.4960	0.4973	0.4897
200	20	0.6	0.4999	0.5249	0.5406	0.5132	0.4920	0.4706	0.4652
200	20	0.4	0.5026	0.5350	0.5552	0.5104	0.4880	0.4381	0.4387
200	50	0.8	0.5004	0.5027	0.5057	0.5031	0.4879	0.4997	0.4951
200	50	0.6	0.5010	0.4995	0.5073	0.5082	0.4947	0.4906	0.4826
200	50	0.4	0.5009	0.4969	0.5083	0.5144	0.4944	0.4819	0.4734

See text for discussion. True value of $\delta_1 = 0.5$.

Table 2: Monte Carlo Standard Deviations of δ_1

Results based on 200 replications

N	μ_0	ω	PEWMA	LLOLS	ARIMA	Poisson	Lagged Poisson	Negative Binomial	Lagged Neg. Bin.
50	10	0.8	0.0542	0.1778	0.1769	0.0428	0.0410	0.0445	0.0417
50	10	0.6	0.0364	0.1387	0.1480	0.0488	0.0460	0.0617	0.0497
50	10	0.4	0.0284	0.1315	0.1460	0.0556	0.0419	0.0721	0.0485
50	10	0.2	0.0217	0.0582	0.0616	0.0651	0.0551	0.0802	0.0613
50	20	0.8	0.0251	0.0457	0.0444	0.0267	0.0260	0.0272	0.0263
50	20	0.6	0.0239	0.0737	0.0715	0.0323	0.0290	0.0356	0.0309
50	20	0.4	0.0209	0.0287	0.0264	0.0361	0.0296	0.0423	0.0322
50	20	0.2	0.0161	0.0254	0.0168	0.0356	0.0308	0.0462	0.0353
50	50	0.8	0.0169	0.0222	0.0217	0.0175	0.0173	0.0176	0.0173
50	50	0.6	0.0121	0.0192	0.0184	0.0178	0.0166	0.0184	0.0173
50	50	0.4	0.0116	0.0184	0.0145	0.0221	0.0195	0.0237	0.0206
100	10	0.8	0.0294	0.1522	0.1604	0.0328	0.0338	0.0381	0.0358
100	10	0.6	0.0293	0.1405	0.1627	0.0393	0.0438	0.0768	0.0584
100	10	0.4	0.0152	0.0670	0.0947	0.0304	0.0267	0.0766	0.0569
100	20	0.8	0.2238	0.0333	0.0343	0.0187	0.0185	0.0209	0.0197
100	20	0.6	0.0173	0.0732	0.0959	0.0257	0.0328	0.0423	0.0293
100	20	0.4	0.0159	0.0418	0.0440	0.0261	0.0236	0.0488	0.0325
100	50	0.8	0.0120	0.0161	0.0160	0.0125	0.0125	0.0126	0.0127
100	50	0.6	0.0105	0.0169	0.0153	0.0151	0.0140	0.0177	0.0156
100	50	0.4	0.0092	0.0165	0.0113	0.0158	0.0142	0.0238	0.0180
200	20	0.8	0.0147	0.0598	0.0567	0.0307	0.0225	0.0257	0.0230
200	20	0.6	0.0138	0.1018	0.0994	0.0593	0.0447	0.0469	0.0381
200	20	0.4	0.0128	0.1338	0.1279	0.0712	0.0624	0.0636	0.0566
200	50	0.8	0.0085	0.0189	0.0136	0.0197	0.0165	0.0187	0.0166
200	50	0.6	0.0078	0.0330	0.0251	0.0360	0.0247	0.0317	0.0268
200	50	0.4	0.0066	0.0332	0.0271	0.0429	0.0271	0.0349	0.0292

See text for discussion of experiments and results.

Table 3: Relative Efficiency for regression parameter δ_1

with respect to PEWMA Monte Carlo Standard Deviation

Results based on 200 replications.

N	μ_0	ω	LLOLS	ARIMA	Poisson	Lagged Poisson	Negative Binomial	Lagged Neg. Bin.
50	10	0.8	3.280	3.263	0.7907	0.7556	0.8201	0.7692
50	10	0.6	3.807	4.063	1.340	1.261	1.694	1.364
50	10	0.4	4.638	5.151	1.963	1.478	2.544	1.710
50	10	0.2	2.689	2.846	3.007	2.543	3.704	2.831
50	20	0.8	1.824	1.770	1.066	1.036	1.084	1.050
50	20	0.6	3.090	2.996	1.353	1.218	1.494	1.297
50	20	0.4	1.371	1.263	1.726	1.416	2.024	1.539
50	20	0.2	1.572	1.039	2.209	1.911	2.860	1.211
50	50	0.8	1.315	1.284	1.037	1.023	1.039	1.022
50	50	0.6	1.586	1.514	1.469	1.372	1.515	1.426
50	50	0.4	1.580	1.249	1.897	1.679	2.032	1.772
100	10	0.8	5.176	5.454	1.114	1.148	1.294	1.218
100	10	0.6	4.800	5.559	1.341	1.497	2.624	1.996
100	10	0.4	4.408	6.227	1.997	1.759	5.040	3.746
100	20	0.8	0.1488	0.1532	0.0834	0.0826	0.0935	0.0880
100	20	0.6	4.222	5.532	1.481	1.389	2.439	1.690
100	20	0.4	2.640	2.775	1.647	1.491	3.082	2.048
100	50	0.8	1.344	1.334	1.038	1.045	1.052	1.058
100	50	0.6	1.615	1.460	1.446	1.333	1.692	1.487
100	50	0.4	1.784	1.225	1.708	1.534	2.576	1.951
200	20	0.8	4.061	3.849	2.082	1.526	1.744	1.563
200	20	0.6	7.368	7.194	4.288	3.232	3.397	2.756
200	20	0.4	10.49	10.03	5.578	4.895	4.988	4.434
200	50	0.8	2.229	1.605	2.334	1.956	2.207	1.964
200	50	0.6	4.216	3.207	4.606	3.153	4.054	3.432
200	50	0.4	5.029	4.104	6.052	4.112	5.289	4.428

See text for discussion.

$$\text{Relative Efficiency} = \frac{\text{Indicated Model's Monte Carlo Standard Deviation for } \delta_1}{\text{PEWMA Monte Carlo Standard Deviation for } \delta_1}$$

Table 4: Mean Relative Overconfidence

Relative overconfidence for regression parameter δ_1

Monte Carlo Results based on 200 replications.

N	μ_0	ω	PEWMA	LLOLS	ARIMA	Poisson	Lagged Poisson	Negative Binomial	Lagged Neg. Bin.
50	10	0.8	1.256	2.378	2.431	1.031	1.008	1.003	0.9405
50	10	0.6	1.068	2.507	2.462	1.051	1.076	1.292	1.113
50	10	0.4	0.9395	1.475	2.809	1.125	0.9545	1.462	1.100
50	10	0.2	0.9228	1.159	1.991	1.403	1.341	1.781	1.516
50	20	0.8	0.9956	1.419	1.451	0.9769	0.9645	0.9614	0.9239
50	20	0.6	1.077	2.160	2.561	1.017	0.9644	1.098	0.9950
50	20	0.4	1.115	0.8676	1.371	1.042	0.9545	1.224	1.015
50	20	0.2	1.045	1.317	0.9187	1.143	1.105	1.471	1.239
50	50	0.8	1.082	1.187	1.217	1.045	1.040	1.016	1.011
50	50	0.6	0.8763	0.9873	1.189	0.926	0.8988	0.9435	0.9159
50	50	0.4	0.9534	0.9130	1.273	1.039	0.9774	1.056	0.9739
100	10	0.8	1.005	2.259	2.490	0.8622	0.9207	0.9613	0.9349
100	10	0.6	1.075	1.603	2.175	0.7821	0.9832	1.461	1.283
100	10	0.4	0.9939	1.424	3.134	0.7648	0.7618	1.848	1.619
100	20	0.8	0.8872†	1.401	1.234	0.7502	0.7671	0.8222	0.7951
100	20	0.6	0.9994	2.121	3.849	0.8037	0.8300	1.288	1.004
100	20	0.4	1.065	1.247	2.328	0.7227	0.7508	1.332	1.031
100	50	0.8	0.9951	1.005	1.127	0.8068	0.8261	0.8143	0.8349
100	50	0.6	0.9832	0.9139	1.324	0.7641	0.7526	0.8727	0.8111
100	50	0.4	0.9969	0.8036	1.289	0.6609	0.6560	0.9612	0.8223
200	20	0.8	1.059	2.143	2.288	1.381	1.058	1.118	1.074
200	20	0.6	1.020	2.334	3.079	1.706	1.420	1.312	1.237
200	20	0.4	1.082	1.845	2.300	1.416	1.360	1.359	1.385
200	50	0.8	0.9961	1.377	1.249	1.415	1.199	1.317	1.198
200	50	0.6	1.024	1.739	2.361	1.709	1.253	1.472	1.358
200	50	0.4	0.9767	1.484	2.776	1.635	1.145	1.298	1.263

See text for discussion.

$$\text{Relative Overconfidence} = \frac{\text{Monte Carlo Standard Deviation for } \delta_1}{\text{Huber-White Standard Error for } \delta_1}$$

† indicates that one replication removed because it was an outlier that skewed the result.

Table 5: Intervention Analysis of U.S. Supreme Court Agenda Series Cases

Regulation Policy Cases

Estimated Regression Parameters for intervention models

Parameter	PEWMA	Poisson	Negative Binomial	Gaussian ARIMA(0,0,1)
Intervention (1954)	-0.7165	-0.6107	-0.5855	-0.6448
	(0.2613)	(0.2440)	(0.3264)	(0.1825)
ω	0.6078			
	(0.0639)			
Constant		2.8757	2.8768	3.467
		(0.0778)	(0.1047)	(0.1457)
γ			1.1390	
			(0.3265)	
ρ		0.0189	0.0189	0.7890
		(0.0020)	(0.0026)	(0.09724)
Final Log-likelihood	-201.55	-216.92	-209.81	
AIC	405.09	439.85	425.62	

Note: Standard errors in parentheses. No likelihood or AIC values are computed for the ARIMA(0,0,1) model, since the scales are not comparable. See text for discussion.

Equality Policy Cases

Estimated Regression Parameters for intervention models

Parameter	PEWMA	Poisson	Negative Binomial	Gaussian ARIMA(0,0,1)
Intervention (1963)	0.9024	1.158	1.1913	1.021
	(0.2396)	(0.1950)	(0.3291)	(1.674)
ω	0.6284			
	(0.0668)			
Constant		1.5574	1.5666	1.847
		(0.0793)	(0.1276)	(0.3623)
γ			2.070	
			(0.2880)	
ρ		0.0617	0.0610	0.3650
		(0.0040)	(0.0063)	(0.1233)
Final Log-likelihood	-162.14	-205.57	-179.42	
AIC	326.28	417.14	364.84	

Note: Standard errors in parentheses. No likelihood or AIC values are computed for the ARIMA(0,0,1) model, since the scales are not comparable. See text for discussion.