Time Series Models for Event Count Data

Patrick T. Brandt
and
John T. Williams
Department of Political Science
Indiana University, Bloomington
August 1999

1 Overview of Event Count Models

Standard approach to modeling event count data is to use a Poisson distribution:

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

Estimation of the mean parameter is accomplished via maximum likelihood methods.

Note that for the Poisson model,
$E(y) = V(y) = \mu.$

A Poisson regression model can be created by including covariates in the parameterization

$$\mu = \exp\left(X_t\delta\right).$$

- Poisson regression models are limited because they assume events are independent.

- Alternative models assume dependence: negative binomial and generalized event count (GEC)

(a) GEC and negative binomial are not appropriate for time series data

(b) Including a lagged endogenous count implies a growth rate model

2 Current Practice with Event Count Models

- Many political scientists now use event count models.
- However, much of this data exhibits time series properties.
- Time dependence is an important issue in many research areas that use event count models:

(a) Presidential vetoes and executive actions

(b) International relations studies of conflict

(c) Supreme Court case agendas

- Little is known about the efficiency properties of event count models in the presence of **dynamic misspecification.**

- If count data demonstrate serial dependence, how can we model this dependence?

- Further, if we fail to model this dependence, how inefficient are the estimates we get?

3 Limits of a lagged count model

- Exponentiated coefficient on the lagged variable (ρ) is no longer an autocorrelation coefficient.
  It is a growth rate.
- Model is only stationary if $\rho < 0$
- Model is only appropriate for non-stationary event counts, since the mean is an exponential function of time.

Suppose that $z_t \sim Po(\mu_t)$,
$\mu_t = \exp(X_t\delta + \rho z_{t-1})$ and X_t are i.i.d.

The growth rate of this lagged Poisson regression model is

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

Taking expectations gives

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

- Unless $\rho = 0$ or $E\left[z_{t-1} - z_{t-2}\right] = 0$, this model implies a non-zero growth rate for the conditional mean.

- The coefficient ρ is a growth rate rather than an autocorrelation or discounting coefficient.

4 Two Approaches to Event Count Time Series

- **PEWMA:** Poisson Exponentially Weighted Moving Average

  Models a moving mean for persistent event count data

- **PAR(p)**: Poisson Autoregressive Model of Order p

  Models a linear autoregressive, mean reverting event count series.

5 A Model for Persistent Series of Counts

- Model for persistent count data: the **Poisson exponentially weighted moving average (PEWMA)**

(1) Has a general specification of the dynamics in persistent event count data.

(2) Generalizes earlier work on event count models.

(3) Based on the well known Kalman filter.

(4) Easily implemented

6 The PEWMA model

The PEWMA is a structural time series model. The model and method of estimation were originally developed by Harvey and Fernandes (1989), Harvey (1991), and Shephard (1994).

It is based on a measurement equation, a state equation, and a prior:

Measurement Equation:

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

Transition Equation:

$$\mu_t = \mu_{t-1}^{*} \exp(X_t\delta + r_t)\,\eta_t,$$

$$where \quad \eta_t \sim Beta(\omega a_{t-1}, (1-\omega)\, a_{t-1})$$

Conjugate Prior:

$$\mu_{t-1} \sim \Gamma(a_{t-1}, b_{t-1}).$$

- The hyperparameter $\omega \in (0, 1]$, measures the dependence. μ^*_{t-1} is the period $t-1$ conditional mean.

- Small values indicate more dynamics and dependence in the data, while values near one indicate independence.

- Transition equation differs from Harvey and Fernandes. It is based on the gamma distributed transition results of Shephard (1994). It allows for a separate growth rate (r_t) in each period. The mean growth rate is zero.

6.1 Estimation

To estimate the model, construct the joint density conditional on Y_τ for observations $y_{\tau+1}, ..., y_T$:

$$\Pr\left(y_{\tau+1}, ..., y_T\right) = \prod_{t=\tau+1}^{T} \Pr\left(y_t | Y_{t-1}\right).$$

where:

$$\begin{aligned}\Pr\left(y_t|Y_{t-1}\right) &= \int_0^\infty \underbrace{\Pr\left(y_t|\mu_t\right)}_{Measurement}\ \underbrace{\Pr\left(\mu_t|Y_{t-1}\right)}_{Transition} d\mu_t \\ &= \frac{\Gamma\left(y_t + a_{t|t-1}\right)}{y_t!\Gamma\left(a_{t|t-1}\right)} b_{t|t-1}^{a_{t|t-1}} \\ &\quad \cdot \left(1 + b_{t|t+1}\right)^{-\left(y_t + a_{t|t-1}\right)},\end{aligned}$$

where $a_{t|t-1}$ and $b_{t|t-1}$ are the conditional values of a_{t-1} and b_{t-1}.

Given the conditional probability density of $y_t|Y_{t-1}$, construct the log-likelihood function $(\ln L)$ based on the joint density:

$$\ln L = \sum_{t=\tau+1}^{T} \ln\Gamma\left(y_t + \omega a_{t-1}\right) - \ln\left(y_t!\right)$$
$$- \ln\Gamma\left(\omega a_{t-1}\right)$$
$$+\omega a_{t-1} \ln\left(\omega b_{t-1} \exp\left(-X_t\delta - r_t\right)\right)$$
$$- \left(\omega a_{t-1} + y_t\right) \ln\left(1 + \omega b_{t-1} \exp\left(-X_t\delta - r_t\right.\right.$$

6.2 PEWMA Forecast Function

- The forecast function for the PEWMA model is derived given estimates of ω and δ. The mean and variance of the predictive distribution are :

$$
\begin{aligned}
E\left(y_{T+1}|Y_T\right) &= \frac{a_{T+1|T}}{b_{T+1|T}} = \frac{a_T}{b_T} = \mu_T, \\
V\left(y_{T+1}|Y_T\right) &= \frac{a_{T+1|T}\left(1 + b_{T+1|T}\right)}{b^2_{T+1|T}} \\
&= \omega^{-1} V\left(\mu_T|Y_T\right) + E\left(\mu_T|Y_T\right)
\end{aligned}
$$

where a_T and b_T are given as above.

- Based on repeated substitutions of a and b the forecast function for the *one-step ahead prediction* is:

$$\bar{y}_{T+1|T} = \frac{\exp\left(X_{T+1}\delta + r_{T+1}\right)\sum_{j=0}^{T-1}\omega^j y_{T-j}}{\sum_{j=0}^{T-1}\omega^j \exp\left(X'_{T-j}\delta + r_{T-j}\right)}.$$

- When T is large, $\bar{y}_{t+1|t}$ approaches $\mu_t = \omega\bar{y}_{t|t-1} + (1-\omega)\,y_t \;\; for \;\; t = 1, ..., T.$

7 Application: Supreme Court Case Agenda Series

- **Data:** Number of cases in civil rights and economic regulation agendas for each Court term, 1933-1993.

- **Independent variable:** Intervention specification for effects of 1953 term.

(a) Empirically determine the intervention using log-likelihood / AIC values

(b) Look at interventions in years after 1953.

- **Models:** Estimate intervention effects using,

(a) PEWMA

(b) Poisson

(c) Negative Binomial

(d) Gaussian ARIMA(0,0,1)

7.1 Results: Supreme Court Case Agenda Analysis

- Poisson, negative binomial and ARIMA models predict a **smaller** decline in economic regulation cases than the PEWMA.
- After intervention, the alternative models over estimate the number of economic regulation cases.
- Poisson, negative binomial and ARIMA models predict a **larger** increase in equality cases than the PEWMA.
- After intervention, the alternative models under estimate the number of equality regulation cases.

- PEWMA predictions of the effect of the intervention are closer to the observed data. Getting the distribution and dynamics right makes a difference in estimating the effect of the covariates.

8 Benefits of Using PEWMA Model

- **Efficiency Gains:** PEWMA standard errors are 1.5 and 5 times smaller than Poisson or negative binomial regressions, based on Monte Carlo evidence.
- **Lagged dependent variables are a poor correction:** Monte Carlo results show that Poisson and negative binomial estimates do not improve with lag count.

- **Diagnosing dynamics:** PEWMA can uncover the presence of dynamics in event count data. After accounting for dynamics, static models perform poorly.

- **Identifying Dynamics using ACF:** Can be used to identify dynamics in a manner similar to continuous, Gaussian processes.

- **Using Gaussian ARIMA models?:** ARIMA models appear to be a poor approximation for time series event count data. For nonstationary series, Gaussian models may provide some insights.

9 The PAR(p) Process

- An event count series can also be modelled using a ”linear autoregressive process.”
- This AR(p) process can be used to define a new transition equation for the state space model.

9.1 Advantages of PAR(p)

- **Alternative specification:** another model for the conditional mean of a stationary count data series.
- **Ease of interpreting:** predictive distribution is based on a linear function.
- **Generalization:** The AR model can be account for higher order lag structures.
- **Diagnostics:** Standard ACF and PACF routines can be used for diagnostics.

9.2 Linear AR(p) Processes

- This work generalizes earlier work by Grunwald et al. (1997, 1998) on linear AR(1) processes.

- **Assumptions:**

(a) Let Y_{t-1} be all the prior information about the event count at time t.

(b) y_t is a time-homogeneous Markov process with the conditional transition $\Pr(y_t|Y_{t-1})$, and that $E[Y_0] = \mu < \infty$.

(c) The expectation of the conditional count at time t has a finite mean, $E[y_t|Y_{t-1}] = m_t$.

Then y_t is a p-th order linear autoregressive process if

$$E\left[y_t|Y_{t-1}\right] = \sum_{i=1}^{p} \rho_i Y_{t-i} + \lambda \qquad (1)$$

where ρ and λ are any real numbers.

- No restriction on the probability density $\Pr\left(y_t|Y_{t-1}\right)$.
- Choice of this probability density for y_t places constraints on the admissible values of λ and ρ. For event counts this implies, $\lambda, \rho > 0$.

9.3 Stationarity of Linear AR(p) processes

Equation (1) generates a stationary mean.

Using iterated expectations for an AR(p) process:

$$E\left[y_t|Y_{t-1}\right] = \sum_{i=1}^{p} \rho_i Y_{t-i} + \lambda \quad (2)$$

$$E\left[E\left[y_t|Y_{t-1}\right]\right] = E\left[\sum_{i=1}^{p} \rho_i Y_{t-i} + \lambda\right] \quad (3)$$

$$E\left[Y_t\right] = \sum_{i=1}^{p} \rho_i E\left[Y_{t-i}\right] + \lambda. \quad (4)$$

Noting that (4) is a geometric series for ρ_i outside the unit circle, then

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

Since $E[Y_0] = \mu$ (by definition), equation (2) can be written as

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

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

9.4 PAR(p) model

Based on a measurement equation, state equation, and a prior

Measurement Equation:

$$\Pr(y_t|m_t) = \frac{m_t^{y_t} e^{-m_t}}{y_t!},$$

Transition Equation:

$$m_t = \sum_{i=1}^{p} \rho_i Y_{t-i} + \left(1 - \sum_{i=1}^{p} \rho_i\right) \mu$$

Conjugate Prior:

$$\Pr(m_t|Y_{t-1}) = \Gamma(\sigma_{t-1} m_{t-1}, \sigma_{t-1})$$
$$m_{t-1} > 0, \sigma_{t-1} > 0$$

where $m_{t-1} = E[y_t|Y_{t-1}]$
and $\sigma_{t-1} = Var[y_t|Y_{t-1}]$.

9.5 PAR(p) likelihood

The forecast density for the one-step ahead distribution is

$$\Pr\left(y_t|Y_{t-1}\right) = \int_\theta \underbrace{\Pr\left(y_t|\theta_t\right)}_{Measurement} \underbrace{\Pr\left(\theta_t|Y_{t-1}\right)}_{Transition} d\theta$$

$$= \int_\theta \frac{\theta_t^{y_t} e^{-\theta_t}}{y_t!} \cdot \frac{e^{-\sigma_{t|t-1}\theta}\theta^{\sigma_{t|t-1}m_{t|t-1}-1}\sigma_t^{\sigma_{t|t-1}m_{t|t-1}}}{\Gamma\left(\sigma_{t|t-1}m_{t|t-1}\right)}$$

$$= \frac{\Gamma\left(\sigma_{t|t-1}m_{t|t-1}+y_t\right)}{\Gamma\left(y_t+1\right)\Gamma\left(\sigma_{t|t-1}m_{t|t-1}\right)}\left(\sigma_{t|t-1}\right)^{\sigma_{t|t-1}m_{t|t-1}}$$

$$\cdot\left(1+\sigma_{t|t-1}\right)^{-\left(\sigma_{t|t-1}m_{t|t-1}+y_t\right)}.$$

This is a negative binomial distribution.

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

$$\begin{aligned}
\mathcal{L}\left(y_t \ldots y_T | Y_{t-1}\right) &= \ln \prod_{t=1}^{T} \Pr\left(y_t | Y_{t-1}\right) \\
&= \sum_{t=1}^{T} \ln \Gamma\left(\sigma_{t-1} m_{t-1} + y_t\right) \\
&\quad - \ln \Gamma\left(y_t + 1\right) \\
&\quad - \ln \Gamma\left(\sigma_{t-1} m_{t-1}\right) \\
&\quad + \sigma_{t-1} m_{t-1} \ln\left(\sigma_{t-1}\right) \\
&\quad - \left(\sigma_{t-1} m_{t-1} + y_t\right) \ln\left(1 + \sigma_{t-1}\right)
\end{aligned}$$

- Substituting the linear AR(1) process for m_t yields a PAR(1) model with a negative binomial predictive distribution.

- Covariates can be introduced by replacing μ with $\exp(X_t\delta)$ in (5).

10 Application: Hospital Deaths Series

- **Data:** Number of monthly deaths in Vermillion County (Indiana) Hospital, January 1991-December 1995.

- **Independent variables:**
  Two temporary interventions,

(a) Accused nurse working in hospital: captures the epidemic effects.

(b) Post-nurse period: captures the post-epidemic effects.

- **Models:** Estimate intervention effects using,

(a) PEWMA

(b) PAR(1)

(c) PAR(2)

(d) Poisson

(e) Lagged Poisson

(f) Negative Binomial

(g) Lagged Negative Binomial

10.1 Results: Hospital Deaths Series

- PAR(p) models are more consistent with the data
- Static event count models predict a constant mean, rather than capturing the change in the number of deaths over time
- Only event count time series models capture the effects of the nurse's arrival and departure.

WHERE TO FIND PESTS AND PAPERS:

http://php.indiana.edu/~pbrandt

This is the website where we have made available the current version of our GAUSS software, PESTS. PESTS can be used to estimate the PEWMA and PAR(p) models that we have developed. The site also contains a detailed users manual for PESTS, as well as links to our papers on event count time series models.

Contact information:

Patrick T. Brandt
Department of Political Science
210 Woodburn Hall
Indiana University
Bloomington, IN 47405
E-mail: pbrandt@indiana.edu

John T. Williams
Department of Political Science
210 Woodburn Hall
Indiana University
Bloomington, IN 47405
E-mail: jotwilli@indiana.edu