A Bayesian Analysis of Time-Series Event Count Data: An Application to Legislative Production *

Kentaro Fukumoto[†]

September 29, 2006

Abstract

Political science is abundant of time-series event count (TSEC) data. This paper argues that state space form makes TSEC models easy to interpret, compare and extend. Its merit is illustrated by extension of the Poisson exponentially weighted moving average model. With help of Markov Chain Monte Carlo (MCMC), I propose to use negative binomial instead of Poisson in measurement equation, which has not been used in TSEC models. Monte Carlo simulation demonstrates that my model is more robust against violation of Poisson assumption such as omitted variables. This paper also reanalyzes Tanabe (1995)'s study on the annual number of laws to derive the opposite finding. Moreover, I rewrite other existent TSEC models in state space form.

*This is a paper prepared for Workshop on Bayesian Statistics and Applied Econometrics, Tohoku University, Sendai, Japan, October 31 - November 1, 2006. Its earlier versions were presented at the Annual Meeting of the Midwest Political Science Association, Chicago, IL, USA, April 20-23, 2006 and the Annual Meeting of the Japan Election Studies Association, Tokyo, Japan, May 20-21, 2006. I appreciate Patrick Brandt, Andrew Martin and Kevin Quinn for their comments. This is work in progress. Comments are really welcome.

[†]Associate Professor, Department of Political Science, Gakushuin University, Tokyo, E-mail: `FirstName dot LastName at gakushuin dot ac dot jp`, URL: `http://www-cc.gakushuin.ac.jp/~e982440/index_e.htm`.

1 Introduction

Political scientists have studied various kinds of time-series event count (TSEC) data such as disputes between countries (Brandt et al., 2000), presidential use of forces (Mitchell and Moore, 2002), effect of the seat belt law on the number of killed and seriously injured van drivers (Harvey, 1989; Durbin and Koopman, 2000, 421-2) and the annual number of enacted laws (Fukumoto, 2004; Howell et al., 2000), to name only a few. Analyzing TSEC data requires them to pay attention to *both* dynamic structure and the feature of non-negative integer. Researchers in political science have shirked this difficulty. A few exceptions in political science are the Poisson Exponentially Weighted Moving Average (P-EWMA) model (Brandt et al., 2000) and the Poisson Auto-Regressive model (Brandt and Williams, 2001). Statisticians have developed many models and estimators for TSEC data (Cameron and Trivedi, 1998, ch. 7), while most existent models use Poisson for count distributiuon.

Assumption of Poisson, however, is often violated in the real data generation process in politics as well as in other fields. Some independent variables may be omitted and count variables may not be identically and independently distributed. But previous models cling to Poisson because other count distributions such as negative binomial do not have conjugate prior which is used for serial errors and this fact makes estimation difficult.

This paper proposes a new TSEC model which uses negative binomial for count distributiuon. Its estimation is possible via Markov Chain Monte Carlo (MCMC) method. Monte Carlo simulation will show that it is more robust against omitted variables than P-EWMA. This model can also contribute to statistical research as a TSEC model which utilizes non exponential family distribution.

Models are expressed in state space form, which facilitates our understanding of models as well as their extension. In the first place, since TSEC models have been presented in various ways, it is not easy to compare them when analysts choose a model appropriate for the data at hand. State space form offers a common format to express TSEC models.

The organization of the paper is as follows. The next section explains state space form of TSEC models and maintains its merits. In the following section, using state space form and MCMC, I extend P-WEMA model to introduce negative binomial random walk error model and demonstrate that the latter is more robust against omitted variables than the former by Monte Carlo simulation. In section four, this model is applied to Tanabe (1995)'s study on the number of laws in social policy the Japanese legislature passes every year and sheds new light on the data. The second last section discusses state space representation of other existent models. Finally, I conclude.

2 State Space Form of TSEC Model

This section explains how to express TSEC models in state space form. Preceding works have already expressed *some* TSEC models in state space form, though I emphasize that state space form can represent almost *all* TSEC models, enables us to interpret TSEC models in the same way as ordinary Gaussian time series models, and makes it easy to extend existent TSEC models.

State space form is composed of measurement equation and transition equation. Measurement equation expresses the distribution of an event count dependent variable (Y_t)

conditioned on a current measurement covariates vector (x_t) and a current state variable (or vector, θ_t).

$$Y_t \sim f(y_t|\theta_t, x_t, \beta, \varsigma^2),$$

where β is a coefficients vector and ς^2 is an ancillary parameter (of variance). θ_t may or may not be count.

Transition equation represents how the past observations ($y_{s<t}$) or states ($\theta_{s<t}$) affects the present state (θ_t) conditioned on a current transition vector (z_t). Observation driven model is written as

$$\theta_t \sim g(\theta_t|y_{s<t}, z_t, \gamma, \sigma^2),$$

and parameter driven model is written as

$$\theta_t \sim g(\theta_t|\theta_{s<t}, z_t, \gamma, \sigma^2),$$

where γ is a coefficients vector and σ^2 is an ancillary parameter. Only a state variable θ_t delivers information from the past to the present. In parameter driven model, covariates and errors in transition equation continue to affect future observations, while those in measurement equation do not. Theoretically, this difference is important. For example, autocorrelation structure of Y_t's should not depend on transition equation parameters and, therefore, should reduce to that of θ_t's. In observation driven model, covariates and errors in both equations have persistent effects on future event counts. In my opinion, the former is better than the latter because only parameter driven model distinguishes long term effects from short term ones. If analysts can not estimate state variable (or its moments), however, they can not fail to use observation driven model.

The most straight-forward model will be a Poisson ARIMA(p,d,q) model. Measurement equations is

$$Y_t \sim Poisson(y_t| \log(\text{mean}) = \theta_t + x_t\beta).$$

Transition equation is

$$\Delta^d\theta_t \sim Normal(\Delta^d\theta_t|\text{mean} = z_t\gamma_z + \sum_{s=1}^{p} \Delta^d\gamma_{\theta,s}\theta_{t-s} + \sum_{u=1}^{q} \gamma_{\epsilon,s}\epsilon_{t-u}, \quad \text{var} = \sigma^2),$$

where

$$\begin{aligned}
\Delta^{d>2}\theta_t &= \Delta^{d-1}\theta_t - \Delta^{d-1}\theta_{t-1} \\
\Delta^{d=1}\theta_t &= \theta_t - \theta_{t-1} \\
\epsilon_t &= \Delta^d\theta_t - E(\Delta^d\theta_t)
\end{aligned}$$

For instance, Chan and Ledolter (1995)'s model can be regarded as the ARIMA(1,0,0) version of this model. Similarly, transfer function model is also easily available.

State space form has some advantages. First, it makes it easy to understand TSEC models. It divides a time series event count model into time series part and event count part. Transition equation models the former, while measurement equation models the latter. We can interpret transition equation of TSEC models like Gaussian time series models such as ARIMA. Measurement equations can take not only Poisson but also other distributions for count data such as (negative or beta) binomial, general event count, hurdle Poisson, zero-inflated Poisson, etc (King, 1989a,b).

Second, state space form offers a common format to express TSEC models and makes it easy to compare them. Though previous TSEC models are expressed in various ways, to my knowledge, every model can be rewritten in state space form as long as its likelihood function is parametrically specified.[1] I will show some below. Comparison in the same format makes difference among models clear and enables us to choose an appropriate model depending on data generation process.

Third, state space form is flexible enough to model various kinds of data generation process and extend existent models. Moreover, parameters of every model in state space form, including a state variable, can be estimated by MCMC, even if closed form of likelihood function is not available and maximum likelihood estimation is difficult (though identification problems may still remain). This relieves scholars from making a model so that they can estimate it. Rather, scholars can pay more attention to data generation process.

3 Extension of the Poisson EWMA Model

In political science, few works have addressed *both* serial dependence and the nature of non-negative integer when they analyze TSEC data. Brandt et al. (2000) and Brandt and Williams (2001) are rare exceptions. Brandt et al. (2000) introduce Poisson exponentially weighted moving average (P-EWMA) model from Harvey and Fernandes (1989) to political methodology. Thus, I explain it as an example of TSEC model. P-EWMA is a observation driven model. The present paper proposes a parameter driven model which resembles P-EWMA data generation process. Then, I extend it so that measurement equation follows negative binomial distribution instead of Poisson and estimates become robust. I will show robustness of this new model by Monte Carlo simulation. Previous non Gaussian time series model, including Bayesians', have mostly focused on exponential family distributions. In this sense, too, a negative binomial model of TSEC data is new and important.

3.1 Poisson EWMA as Observation Driven Model

Motivation of P-EWMA is to mimic the following Gaussian model;

$$\begin{aligned} Y_t &\sim Normal(y_t|\log(\text{mean}) = \log(\theta_t) + x_t\beta, \quad \text{var} = \varsigma^2) \\ \log(\theta_t) &\sim Normal(\log(\theta_t)|\text{mean} = \log(\theta_{t-1}), \quad \text{var} = \sigma^2). \end{aligned}$$

Difficulty in transforming this model into TSEC version lies in how to estimate $\theta_{t-1} > 0$. $y_{t-1}/\exp(x_{t-1}\beta)$ is a good candidate. But y_{t-1} can be zero, which is a very important

[1] To put it another way, if a model specifies parameters' moments only, it can not be written in state space form. Such an example is serially correlated error model (Zeger, 1988). I do not consider them here.

feature of TSEC data. In that case, $\theta_{t-1} = 0$, which is not allowed. For an arbitrary small positive number, δ, $(y_{t-1} + \delta)/\exp(x_{t-1}\beta)$ may be used. Cameron and Trivedi (1998) call this "autoregressive model" and criticize δ is "ad hoc". King (1988) also argues that δ is arbitrary and makes estimation biased and less efficient. Moreover, this estimator is inefficient (Brandt et al., 2000).

To address these problems, Harvey and Fernandes (1989) propose the P-EWMA model. Measurement equation has Poisson distribution instead of normal in order to be a distribution of non-negative integers. Transition equation follows gamma instead of normal in order to utilize its conjugacy with Poisson.

$$\begin{aligned} Y_t &\sim Poisson(y_t|\text{mean} = \theta_t \exp(x_t\beta)) &(1)\\ \theta_t &\sim Gamma(\theta_t|\text{shape} = a_{t|t-1}, \text{scale} = b_{t|t-1}), &(2)\end{aligned}$$

where

$$\begin{aligned} a_{t|t-1} &= \omega a_{t-1|t-1} = \omega(a_{t-1|t-2} + y_{t-1}) &(3)\\ b_{t|t-1} &= \omega b_{t-1|t-1} = \omega(b_{t-1|t-2} + \exp(x_{t-1}\beta)) &(4)\\ 0 < \omega < 1 \end{aligned}$$

and x does not contain constant term for identification.[2] Moments of the state variable are

$$\begin{aligned} E(\theta_t|y_{t-1}) &= \frac{a_{t|t-1}}{b_{t|t-1}} = E(\theta_{t-1}|y_{t-1}) = \frac{\Omega_{s=1}^{t-1}(y_{t-s}, \omega, a_{1|0})}{\Omega_{s=1}^{t-1}(\exp(x_{t-s}\beta), \omega, b_{1|0})} &(5)\\ V(\theta_t|y_{t-1}) &= \frac{a_{t|t-1}}{b^2_{t|t-1}} = \frac{V(\theta_{t-1}|y_{t-1})}{\omega}, &(6)\end{aligned}$$

where Ω is an EWMA function, that is,

$$\Omega_{s=1}^{t-1}(z_{t-s}, \omega, z_1) = \left(\sum_{s=1}^{t-1} (\omega z_{t-s})^s\right) + \omega^{t-1} z_0.$$

Using Kalman filter, their model estimates not θ_t's themselves but derives parameters of their distribution (a's and b's) and, therefore, their moments from observed y's and x's as well as model parameters β and ω. Then, thanks to conjugacy between distributions of both equations, they integrate out θ_t's and have negative binomial as marginal distribution of y_t conditioned on $x_t, \beta, a_{t|t-1}$ and $b_{t|t-1}$. Then, maximum likelihood estimates of β and ω are obtained.

P-EWMA model has some assumption which is required (partly) in order to make its estimation possible. MCMC makes them unnecessary and enables us to construct more flexible model. To begin with, conjugacy is required for integrating out state variable and make closed form of likelihood function available. To put it another way, if we use MCMC and do not need analytic expression of likelihood function, we do not have to cling to conjugacy any longer. Next, Eqs.(3) and (4) enable Kalman filter. Once we

[2]If x contains constant term, it would have multicollinearity problem θ_1 (and, therefore, other θ's, too).

use MCMC, however, we do not have to rely on Kalman filter and can employ other kinds of transition equation. Third, as observation driven model, PEMWA uses y_{t-1} in transition equation in order to estimate θ_t's moments. But if we can estimate θ_{t-1} itself by MCMC in parameter driven model, we had better include it in transition equation. Finally, P-EWMA offers moments of filtered $\theta_t|y_t$ but not smoothed $\theta_t|y_T$ $(t = 1, \ldots, T)$, which MCMC provides.

3.2 Parameter Driven Model like Poisson EWMA

MCMC enables us to estimate not just moments of θ_t's conditioned on y_{t-1} but also $\theta_{t|T}$ themselves conditioned on y_T. Thus it will be more straight-forward to model transition process by θ_t's, not by their moments. Instead of Eqs. (2) and (5), I propose to use the following transition equation;

$$\theta_t \sim Gamma(\theta_t|\text{mean} = \theta_{t-1}, \quad \text{var} = \text{mean} \times \sigma^2), \tag{7}$$

where σ^2 is a dispersion parameter. This is a parameter driven model like P-EWMA. For example, moments of θ_t are similar to that of P-EWMA (Eqs. (5) and (6)). In fact, Monte Carlo simulation (not reported) indicates that one of the two models can estimate parameters using data generated by the other.

σ^2 is inversely related with ω in P-EWMA model. In P-EWMA, ω is discount parameter and implies how long past effects persist. Thus, large ω does not deviate current state from the past so much. This means, in my parameter driven model, that variance of error terms is small and, therefore, σ^2 should be small. I can show this analytically to some degree. To begin,

$$\begin{aligned} V(\theta_t|y_{t-1}) &= \frac{E(\theta_t|y_{t-1})}{\omega b_{t-1|t-1}} \\ &= \frac{E(\theta_t|y_{t-1})}{\Omega_{s=1}^{t-1}(\exp(x_{t-s}\beta), \omega, b_{1|0})} \end{aligned}$$

If $x_t\beta = 0 \forall t$ and $b_{1|0} = 0$,

$$\lim_{t\to\infty} V(\theta_t|y_{t-1}) = \frac{1-\omega}{\omega} E(\theta_t|y_{t-1})$$

Comparing this with Eq.(7), I use as estimator of ω

$$\hat{\omega} = (1 + \sigma^2)^{-1}. \tag{8}$$

I rewrite measurement and transition equations so that it resembles ordinary time series models;

$$\begin{aligned} \log(E(y_t)) &= x_t\beta + \log(\theta_t) \\ \log(\theta_t) &= \log(\theta_{t-1}) + \epsilon(\sigma^2), \end{aligned}$$

where ϵ is random error variable whose variance is an increasing function of σ^2. If we regard $\log(\theta_t)$ as error term, $\log(E(y_t))$ in this model (and P-EWMA) looks like (or mimics) random walk error process and $\epsilon(\sigma^2)$ works as if it were white noise.[3] Thus, we can understand that, in P-EWMA model, event count component is Poisson conditioned on state variable, whose temporal structure is random walk error process. Errors are never discounted, while covariates have nothing to do with future observations.

Since this model does (or can) not use conjugacy between measurement and transition equations, we do not have to use gamma distribution for transition equation. It will be more convenient to assume that state variable follows normal distribution because it enables us to use well established technique for MCMC of Gaussian time series model. To begin, inverse gamma is conjugate prior for σ^2 and makes MCMC efficient. Moreover, when transition equation includes covariates whose coefficient prior has multivariate normal distribution, conjugacy between normal and multivariate normal works as well.

3.3 Negative Binomial I(1) Model

Previous non Gaussian time series model, including Bayesians' (e.g. West, Harrison and Migon, 1985), have focused on exponential family distributions. Grunwald, Hamza and Hyndman (1997, p. 619) argue that, in power steady models including P-EWMA, "[b]oth the temporal characteristics of the model $\cdots$ and the dispersion of the forecast distribution are controlled by" a single model parameter (Actually, in my opinion, this is the trick to make estimation possible). "As a result, the range of possible models is quite limited." Even if one adds a scale parameter, no discrete distributions of exponential family with support on the non-negative integers exist. Thus, in order to take into consideration conditional overdispersion, we should depart from Poisson.

In actual data of political science, violation of Poisson assumption is probably common. Events may not be identically and independently distributed. Some variables may be omitted. Usually, negative binomial is a robust alternative to Poisson in this situation. But since negative binomial has no conjugate prior, it has not been used for TSEC model.[4]

Fortunately, as I argued, MCMC makes it possible to estimate negative binomial measurement equation. My model just replaces measurement equation of P-EWMA, Eq.(1), with

$$Y_t \sim NegativeBinomial(y_t|\text{mean} = \theta_t \exp(x_t\beta), \quad \text{var} = \text{mean} \times \varsigma^2), \tag{9}$$

where $\varsigma^2 > 0$ is a dispersion parameter (transition equation is still Eq.(7)). Below, I

[3]This is why data generated by Harvey and Fernandes (1989)'s P-EWMA model may reach and stick in zero counts. In order to avoid this, Brandt et al. (2000) replace Eq.(4) with

$$b_{t|t-1} = \omega[b_{t-1|t-2} + \exp\{x_{t-1}\beta + \Psi(a_{t-1|t-1}) - \Psi(a_{t|t-1})\}]$$

where $\Psi(x)$ is digamma function,

$$\Psi(x) = \frac{\partial \log \Gamma(x)}{\partial x}.$$

But note that $E(\epsilon(\sigma^2)) \neq 0$ and, strictly speaking, $\epsilon(\sigma^2)$ can not be white noise.

[4]Bradlow, Hardie and Fader (2002) propose *approximate* conjugate prior for negative binomial. If one parameterize size and probability, beta is conjugate prior (Harvey and Fernandes, 1989). But this parameterization does not help us to model overdispersion without changing model of mean.

call this model NB-I(1) to emphasize that transition equation is first order integrated (i.e., I(1)) error series. Note that, in P-EWMA model, marginal distribution of y_t conditioned y_{t-1} is negative binomial. Eq.(9) is conditional, not marginal, distribution given θ_t. Difference lies in where errors come. NB-I(1) model considers that there may be omitted variable in x, measurement errors in x or some contagion among subjects may exist, while P-EWMA does not. Thus, when P-EWMA estimator faces overdispersion in measurement equation, it takes it as errors in transition equation. This results into underevaluation of ω because of overevaluation of σ^2 as shown in the following Monte Carlo simulation. By contrast, NB-I(1) model correctly classifies systematic parts and stochastic parts into persistent ones (σ^2 (and γ in Eq.(1))) and temporary ones (β and ς^2).

3.4 MCMC Estimator of NB-I(1)

Since NB-I(1) model does not have a closed form, we can not rely on maximum likelihood estimation. Instead, I use MCMC. As prior distribution of parameters, I assume multivariate normal for β, inverse gamma for σ^2, gamma for ς^2 and θ_1, respectively.[5] *A priori*, these are statistically independent of each other.

$$\begin{aligned}
\beta &\sim MultiVariateNormal(\beta|\ \text{mean} = \beta_0, \quad \text{variance covariance matrix} = B_0) \\
\sigma^2 &\sim InverseGamma(\sigma^2|\ \text{shape} = a_0/2, \quad \text{scale} = b_0/2) \\
\varsigma^2 &\sim Gamma(\varsigma^2|\ \text{shape} = c_0, \quad \text{scale} = d_0) \\
\theta_1 &\sim Gamma(\theta_1|\ \text{shape} = t_0, \quad \text{scale} = s_0)
\end{aligned}$$

The algorithm of my estimator is random walk Metropolis-Hastings sampling nested in Gibbs sampling.[6] In order to take advantage of Gibbs sampling algorithm, I derive the following full posterior from the above model and Bayes Theorem, namely, likelihood (measurement times transition) times priors:

$$\begin{aligned}
p(\boldsymbol{\theta}, \beta, \sigma^2, \varsigma^2|\ \mathbf{y}) \propto & \left\{ \prod_{t=1}^{t=T} NegativeBinomial(y_t|\ \theta_t \exp(x_t\beta), \varsigma^2) \right\} \times \\
& \left\{ \prod_{t=2}^{t=T} Gamma(\theta_t|\ \theta_{t-1}, \sigma^2) \right\} \times \\
& MultiVariateNormal(\beta|\ \beta_0, B_0) \times \\
& InverseGamma(\sigma^2|\ a_0/2, b_0/2) \times \\
& Gamma(\varsigma^2|\ c_0, d_0) \times Gamma(\theta_1|\ t_0, s_0).
\end{aligned}$$

Thus, I can sample each parameter from its full conditional. The full conditional for

[5] If you use normal for transition equation, inverse gamma will be conjugate with it and make it easy to sample σ^2.

[6] Alternative estimators are posterior mode (Fahrmeir, 1992, with extended Kalman filter and smoother) and piecewise linear function (Kitagawa, 1987, also with filtering and smoothing, but this is not so attractive).

$\theta_t (t \neq 1, T)$ is

$$p(\theta_t | \ \theta_{\sim t}, \beta, \sigma^2, \varsigma^2, \mathbf{y}) \propto NegativeBinomial(y_t | \ \theta_t \exp(x_t \beta), \varsigma^2) \times Gamma(\theta_t | \ \theta_{t-1}, \sigma^2) \times$$
$$Gamma(\theta_{t+1} | \ \theta_t, \sigma^2).$$

The full conditional for θ_T is

$$p(\theta_T | \ \theta_{\sim T}, \beta, \sigma^2, \varsigma^2, \mathbf{y}) \ \propto NegativeBinomial(y_T | \ \theta_T \exp(x_T \beta), \varsigma^2) \times Gamma(\theta_T | \ \theta_{T-1}, \sigma^2).$$

The full conditional for θ_1 is

$$p(\theta_1 | \ \theta_{\sim 1}, \beta, \sigma^2, \varsigma^2, \mathbf{y}) \propto NegativeBinomial(y_1 | \ \theta_1 \exp(x_1 \beta), \varsigma^2) \times Gamma(\theta_1 | \ t_0, s_0).$$

The full conditional for ς^2 is

$$p(\varsigma^2 | \ \boldsymbol{\theta}, \beta, \sigma^2, \mathbf{y}) \ \propto \left\{ \prod_{t=1}^{t=T} NegativeBinomial(y_t | \ \theta_t \exp(x_t \beta), \varsigma^2) \right\} \times Gamma(\varsigma^2 | \ c_0, d_0).$$

The full conditional for σ^2 is

$$p(\sigma^2 | \ \boldsymbol{\theta}, \beta, \varsigma^2, \mathbf{y}) \ \propto \left\{ \prod_{t=2}^{t=T} Gamma(\theta_t | \ \theta_{t-1}, \sigma^2) \right\} \times InverseGamma(\sigma^2 | \ a_0/2, b_0/2).$$

And the full conditional for β is

$$p(\beta | \ \boldsymbol{\theta}, \sigma^2, \varsigma^2, \mathbf{y}) \ \propto \left\{ \prod_{t=1}^{t=T} NegativeBinomial(y_t | \ \theta_t \exp(x_t \beta), \varsigma^2) \right\} \times$$
$$MultiVariateNormal(\beta | \ \beta_0, B_0).$$

We do not have closed forms of posterior densities and, therefore, can not use canned random generators. Target densities are not known, either. Hence, I employ random walk Metropolis-Hastings samplings. Candidate values of β, $\log \theta_t$'s, $\log \sigma^2$ and $\log \varsigma^2$ are generated by using normal. During implementation, one must tune standard deviations of these normal densities.

Obviously, since θ_t's are highly serially dependent on θ_{t-1} and θ_{t+1}, estimation of them is inefficient. If measurement equation and transition one were normal, forward filtering and backward smoothing would provide more efficient estimation (West and Harrison, 1997, 569-71). As the following Monte Carlo simulation shows, however, this estimator still works better than the maximum likelihood estimator of P-EWMA model.

3.5 Monte Carlo Simulation

In order to show that NB-I(1) model is more robust against omitted variable (or random effect) than P-EWMA, I demonstrate Monte Carlo simulation.

1. I draw 500 observations of $x = (x_1, x_2, x_3)$ from multivariate normal distribution only once. Its mean is $(0, 0, 0)$, $Var(X_1) = Var(X_2) = Var(X_3) = 0.5$ and $Cor(X_1, X_2) = Cor(X_1, X_3) = 0$ but $Cor(X_2, X_3) = 0.5$.

2. I set $\beta = (\beta_1 = 0.75, \beta_2 = -0.5, \beta_3 = 2)$ and $\omega = 0.9$.

3. For one simulation, I sample y_t's ($t = 1, \ldots, 500$) using P-EWMA data generation process. I estimate P-EWMA parameters and NB-I(1) ones except β_3 using these same 500 y's, x_1's and x_2's. [7] As for hyperparameters of NB-I(1) model priors, I choose $\beta_0 = (0, 0), B_0 = diag(0.01, 0.01), a_0 = 0.2, b_0 = 0.2, c_0 = 0.01, d_0 = 0.01, t_0 = 16/9$ and $s_0 = 1/9$.[8] I discard first 100 draws as burn-in and use next 1,000 draws without thinning.

4. I repeat this simulation (the third step only) 40 times and obtain 40 sets of parameter estimates.[9]

Table 1 summarizes median and 80% confidence interval of acceptance rates for every parameter or block of parameters. Acceptance rates are calculated using draws after burn-in period. According to this table, estimates do not well converge. Since 80% confidence intervals are not so wide, I illustrate traceplots and densities of arbitrary chosen simulation in Figure 1. β_1 and β_2 are serially dependent but behave well. Clearly, σ^2 has not reach to convergence. ς^2's are not well sampled.

	10%Q	Median	90%Q
(β_1, β_2)	0.649	0.671	0.687
σ^2	0.234	0.247	0.264
ς	0.125	0.138	0.149
θ_t's	0.150	0.472	0.709

Table 1: Summary of Acceptance Rates

Though MCMC estimates are not satisfactory, they are still better than PEWMA results. Table 2 reports summary of estimates. For P-EWMA, they are maximum likelihood estimates. For NB-I(1), they are median of posterior parameter samples. Clearly, NB-I(1) is more robust against an omitted variable than P-EWMA. $\hat{\beta}_1$ is less biased in NB-I(1). P-EWMA has smaller bias of $\hat{\beta}_2$ which, however, should be not zero but one due to omitted variable bias:

$$E(\hat{\beta}_2 - \beta_2) = Cor(X_2, X_3)\beta_3 = 1.$$

[7] For data generation and P-EWMA estimation, I use Brandt et al. (2000)'s `R` code, "pests.r", on a statistical software `R`. The code is version 1.1.2 (revised on September 20, 2005) and is downloaded from `http://www.utdallas.edu/~pbrandt/codepage.html` on March 28, 2006. I appreciate Patrick Brandt for making the code public. Their data generation code uses not random draw but expectation of θ_t as mean of Y_t, which is wrong. I write and run the NB-I(1) code by myself.

[8] These are non-informative except the last two, which are not non-informative by mistake.

[9] All simulation process takes about 20 hours. For one of 40 simulations, estimates of P-EWMA are ill behaved (ω is larger than one, which should not happen, the hessian is not positive definite and, therefore, standard errors are not available for some variables). Thus, I exclude estimates of this data by both models from this summary.

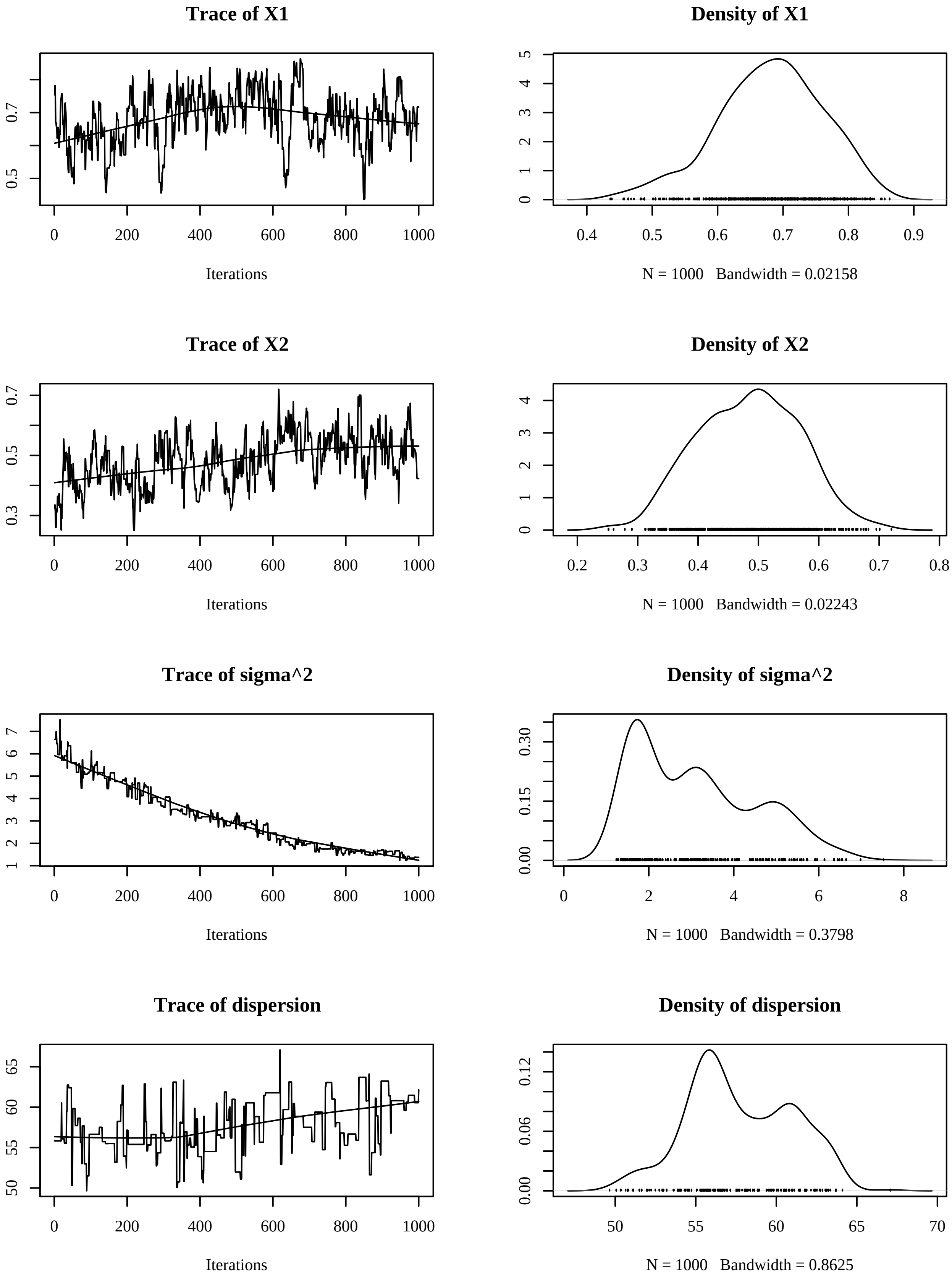

Figure 1: Traceplots and Densities of Main Parameters

In this sense, NB-I(1) has a correctly biased estimate closer to what it should have compared with P-EWMA.[10] NB-I(1) estimates are more efficient. I compare -2 log likelihood of P-EWMA and sample mean of -2 log likelihood of NB-I(1) as discrepancy statistic because both models are non nested in each other. For every set of y's, NB-I(1) has smaller -2 log likelihood, which means that NB-I(1) fits data more than P-EWMA from Bayesian perspective. Hence, mean discrepancy statistic is also smaller as indicated in Table 2. $\hat{\omega}$ of NB-I(1) is estimated according to Eq.(8). This estimate is very rough approximation but is still less biased than P-EWMA estimate, which is very small in order to absorb overdispersion in measurement equation. $\hat{\varsigma^2}$ (only for NB-I(1) by construction) explicitly shows that there is overdispersion and something is wrong with Poisson assumption.

Point Estimate		P-EWMA MLE	NB-I(1) Median
Bias	β_1	-0.218	-0.067
	β_2	0.737	0.980
	ω	-0.777	-0.374
Median	ς^2		37.663
Variance	β_1	0.006	0.001
	β_2	0.002	0.001
	ω	0.034	0.009
	ς^2		134.802
E(-2 Log Likelihood)		11291.996	4966.032

Table 2: Estimates Summary of Monte Carlo Simulation

4 Application: Reanalysis of Legislative Production in the Japanese Social Policies

By applying my NB-I(1) model to Tanabe (1995)'s data and specification, I show that my model changes substantial implication of political analysis. Tanabe (1995) explains what decides how many laws in social policy area the Diet passes in every fiscal year from 1955 to 1992. The dependent variable I reanalyze is the number of enacted laws which include not only new laws but also revised ones. Values are shown in Figure 2. Independent variables are unemployment rate, gini index, change in gini index, GNP growth rate, seat share of the Liberal Democratic Party in the House of Representatives, and a dummy variable to indicate that an election of either house of the Diet is held in that calendar year.

The first two columns of Tables 3 show my own maximum likelihood estimates (MLE) and their standard errors (SE) of Tanabe's model, which assumes Poisson without serial dependence. I can not replicate the original author's result, probably because I construct

[10]Negative binomial assumes that random effect is independent of mean of dependent variable. Thus, it is expectedly robust against inefficiency but not against bias due to omitted variables.

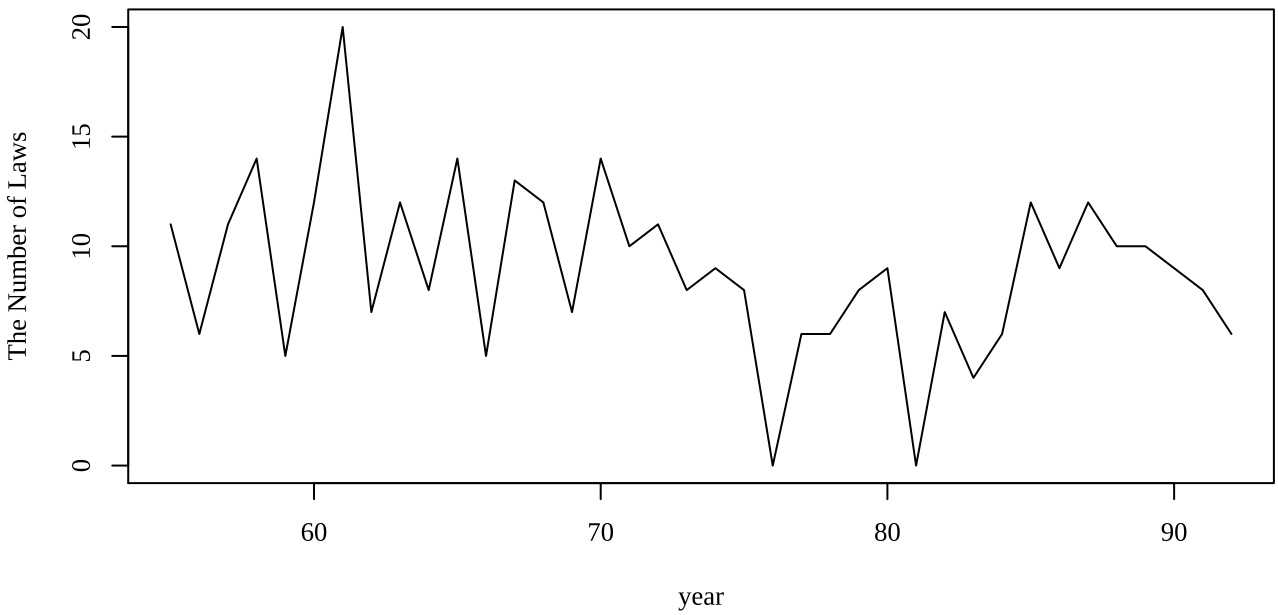

Figure 2: The Number of Laws in Social Policy the Japanese Diet Passes Annually

the data from scratch by myself and it is not the same as his.[11] No independent variables are significantly different from zero at 5% level. For your information, I add "**+" (or "**−") next to variable's name if Tanabe (1995, Table 2, the first column) tells that the coefficient of the variable is positive (or negative) and significantly different from zero at 1% level.

Tanabe (1995)'s model, however, are problematic in two ways. First, his model does not take into consideration dynamic mechanism. A glance at time series of dependent variables strongly suggests that there is serial dependence (Figure 2). In terms of legislative study, too, it is reasonable that bureaucrats and politicians refer to the previous year's number of laws when they decide how many they propose to or pass in the Diet. This is incrementalism. Thus, the next two columns of Table 3 present the results of P-EWMA model. Unfortunately, ω is larger than one (which should not happen) and standard errors of some independent variables are not obtained ("n.a" is shown, which

[11] His definitions of variables are not clear enough for me to replicate his data. Dependent variables are based on Tanabe (1995, Figure 3). Gini index is calculated from disposable income in Table 20-12 at `http://www.stat.go.jp/data/chouki/20.htm`, Statistics Bureau, Ministry of Internal Affairs and Communications, Government of Japan. Unemployment rate is calculated from Table 1 (13) at `http://www.stat.go.jp/data/roudou/longtime/03roudou.htm`, idem. As for GNP, after 1956, I refer to the former SNA68 series (benchmark year 1990, at constant price, calendar year) at `http://www.esri.cao.go.jp/jp/sna/toukei.html`, Economic and Social Research Institute, Cabinet Office, Government of Japan. For 1955, the source is Okawa, Takamatsu and Yamamoto (1974, 214). LDP seat share is average of the values for the first day of each session in a given legislative year; these are weighted with the length of each session. As for importance of replication, see King (1995).

Model	Poisson		P-EWMA			NB-I(1)			
	MLE	SE	MLE	SE		Median	95% CI		
Constant **−	1.69	0.92							
Unemployment	−0.14	0.15	−0.14	0.13		−0.000	−0.004	0.004	
Gini **+	2.20	2.09	0.19	n.a.	n.a.	2.076	2.066	2.097	**
Δ Gini **+	4.03	5.84	4.59	n.a.	n.a.	3.816	3.810	3.821	**
GNP **+	0.03	0.02	0.02	0.02		0.000	−0.003	0.003	
LDP **+	0.00	0.01	0.02	0.01		−0.094	−0.235	−0.040	**
Election **−	−0.02	0.13	−0.03	0.12		−0.000	−0.004	0.004	
ω			1.61	0.40		0.676	0.592	0.747	n.a.
ς^2						0.000	0.000	0.000	n.a.

Table 3: Results of Reanalysis: Legislative Production in the Japanese Social Policies

stands for "not applicable"). Moreover, no covariates are significantly different from zero at 5% level, again.

The second shortcoming of Tanabe (1995)'s model is that assumption of Poisson may be violated. Poisson model implies that the legislature decides whether every bill or to-be bill is passed or not, regardless of other laws' fate. But this is against what we know. Some bills compose a package deal. For example, National Pension, Assistance to the Wounded and Survivors, and Special Measure to Victims of the Atomic Bombs (and sometimes Veteran's Pension) are treated as one set repeatedly. Health Insurance, Seamen's Insurance, and Health Insurance for Day Laborers are another group. Mesures to Disaters are deliberated as if they were one. It is inappropriate to think that one bill of a package is passed or rejected irrespective of its colleague bills'. Besides, we may not be able to specify the correct set of covariates.

Therefore, I employ NB-I(1) model. The last three columns of Table 3 report median and 95% confidence interval of estimates. When the 95% confidence interval includes 0, I mark one star except for ω and ς^2 which are always positive by construction. Similarly, two stars are attached based on the 99% confidence intervals.[12] Gini index and its change significantly increase the number of laws, whereas share of LDP seats decrease it. That is, social inequality and its increase have positive impact on the number of social policy laws and strength of the conservative governing party has negative one. The latter finding is important contradiction to the original author's result which indicates that the LDP promoted social policy. ω shows that there is serial dependence. Unexpectedly, ς^2 is zero, which does not, however, imply that we do not have to construct negative binomial model. Unless we check it, we would never know whether it fits data or not. Though I have not compared validity of each model yet, say, by way of Bayes factors, it is very hard to imagine that any model without dynamic structure is superior to those with it. My analysis suspects that Tanabe (1995)'s argument may be flawed by methodological errors.

[12] All hyperparameters of priors are the same as those used in Monte Carlo simulation except t_0 and s_0, which are set at $y_1 = 11$ and 1, respectively. I scan 110,000 times, burn in first 60,000 scans and thin chains at every 100 scans to get 500 draws in a chain. Acceptance rates are 0.72 for β, 0.42 for σ^2 and 0.04 for ς^2. Acceptance rates of θ's range from 0.04 to 0.10.

5 State Space Representation of Other Models

Cameron and Trivedi (1998) classify TSEC models into six types. I have already referred to three of them. I rewrite two of the others in state space form using common variable notations so that it is easy to compare models.

5.1 Integer Valued ARMA (INARMA)

Alzaid and Al-Osh (1990) and McKenzie (1988) propose similar models, which are together called Integer Valued ARMA (INARMA). p th order autoregressive and q th order moving average model (INARMA(p,q)) is as follows.

$$\begin{aligned} Y_t &= \sum_{s=1}^{p} \omega \circ y_{t-s} + \sum_{s=0}^{q} \epsilon_{t-s} \\ \epsilon_t &\sim Poisson(\epsilon_t|\text{mean} = \varsigma^2) \end{aligned}$$

where $\circ$ is a binomial thinning operator,

$$\begin{aligned} \omega \circ y_{t-s} &= \theta_{t,s} \\ &\sim Binomial(\theta_{t,s}|\text{size} = y_{t-s}, \text{probability} = \omega) \end{aligned}$$

Their presentation intends to remind us of similarity with ARMA model, especially similarity of moments. MA component represents arrival counts, while AR component expresses survival counts. Thus, this model is appropriate for stock count but not for flow one.

It is easy to rewrite this model in state space form. Measurement equation is

$$\begin{aligned} Y_t &= \theta_t + \epsilon_t \\ \epsilon_t &\sim Poisson(\epsilon_t|\text{mean} = \varsigma^2). \end{aligned}$$

Transition equation is

$$\begin{aligned} \theta_t &= \sum_{s=1}^{p} \theta_{t,s} + \sum_{s=1}^{q} \epsilon_{t-s} \\ \theta_{t,s} &\sim Binomial(\theta_{t,s}|\text{size} = y_{t-s}, \text{probability} = \omega). \end{aligned}$$

State variable θ_t is a non-negative integer.

5.2 Discrete ARMA

Cameron and Trivedi (1998, 245-6) give this model as

$$\begin{aligned} Y_t &= \vartheta_t y_{t-1} + (1 - \vartheta_t)\epsilon_t \\ \epsilon_t &\sim Poisson(\epsilon_t|\varsigma^2) \\ \vartheta_t &\sim Bernoulli(\vartheta_t|\omega). \end{aligned}$$

In this model, two variables are passed from the previous period to the current one: lagged dependent variable and whether it affects the present one or not. Thus, I assume state vector $\theta_t = \{\theta_{t,1}, \theta_{t,2}\}$ instead of state variable. Measurement equation is

$$\begin{aligned} Y_t &= \theta_{t,1}\theta_{t,2} + (1 - \theta_{t,1})\epsilon_t \\ \epsilon_t &\sim Poisson(\epsilon_t|\varsigma^2). \end{aligned}$$

Transition equation is

$$\begin{aligned} \theta_{t,1} &\sim Bernoulli(\theta_{t,1}|\omega) \\ \theta_{t,2} &= y_{t-1}. \end{aligned}$$

6 Conclusion

Political science is abundant of TSEC data. So far, most scholars have paid attention to either time series side or event count one. Recently, new models come to address both features at the same time. Even they, however, have limitation. Most assume Poisson. It is not easy to compare models.

The present paper proposes NB-I(1) model in state space form where measurement equation has negative binomial and transition equation follows random walk error process. Monte Carlo simulation illustrates that my model is more robust against omitted variable than currently used P-EWMA model. As readers may see from this example, state space form makes models easy to interpret, compare and extend.

A lot of future agendas are still ahead. Features of the estimator for a finite number of observations (such as efficiency and speed of asymptotic convergence) have not been explored well yet. Even though Monte Carlo simulation in the present paper employs data composed of as large as 500 observations, estimate of dispersion is still biased. When, in another paper, I examine another TSEC model where measurement equation has negative binomial and perform simulation using 1,000 observations, estimate of dispersion is unbiased. Moreover, as is always the case with Bayesian analysis, computational time can be very long. From my experience, 1,000 scans are enough for covariates but not for parameters of variance and dispersion. More severe is observation size. For 100 observation and 1,100 scans, it takes just 10 to 20 seconds. But computational time increase geometrically as observation size. I will address them in a revised version.

References

Alzaid, A. A and M. Al-Osh. 1990. “An Integer-Valued pth-Order Autoregressive Structure (INAR(p)) Process.” *Journal of Applied Probability* 27:314–24.

Bradlow, Eric T., Bruce G. S. Hardie and Peter S. Fader. 2002. “Bayesian Inference for the Negative Binomial Distribution via Polynomial Expansions.” *Journal of Computational and Graphical Statistics* 11:189–201.

Brandt, Patrick T. and John T. Williams. 2001. “A Linear Poisson Autoregressive Model: The Poisson AR(p) Model.” *Political Analysis* 9:164–84.

Brandt, Patrick T., John T. Williams, Benjamin O. Fordham and Brain Pollins. 2000. “Dynamic Modeling for Persistent Event-Count Time Series.” *American Journal of Political Science* 44:823–43.

Cameron, Adrian Colin and P. K. Trivedi. 1998. *Regression Analysis of Count Data.* Cambridge, UK ; New York, NY, USA: Cambridge University Press.

Chan, K. S. and Johannes Ledolter. 1995. “Monte Carlo EM Estimation for Times Series Models Involving Counts.” *Journal of American Statistical Association* 90:242–52.

Durbin, J. and S.J. Koopman. 2000. “Time Series Analysis of Non-Gaussian Observations Based on State Space Models from Both Classical and Bayesian Perspectives.” *Journal of the Royal Statistical Society. Series B* 62:3–29.

Fahrmeir, Ludwig. 1992. “State Space Modeling and Conditional Mode Estimation for Categorical Time Series.” In *New Directions in Time Series Analysis, Part1*, ed. David Brillinger. New York: Springer-Verlag pp. 87–109.

Fukumoto, Kentaro. 2004. ”How Many Laws does the Legislature Make? Cross Country Comparison and Cointegrated Time Series of Japan.” Paper Prepared for the Annual Meeting of the Midwest Political Science Association, April 15-8, 2004. Chicago, IL, USA: .

Grunwald, Garay K., Kais Hamza and Rob J. Hyndman. 1997. “Some Properties and Generalizations of Non-Negative Bayesian Time Series Models.” *Journal of the Royal Statistical Society. Series B* 59:615–26.

Harvey, A. C. and C. Fernandes. 1989. “Time Series Models for Count or Qualitative Observations.” *Journal of Business and Economic Statistics* 7:407–17.

Harvey, Andrew. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge ; New York: Cambridge University Press.

Howell, William, Scott Adler, Charles Cameron and Charles Riemann. 2000. “Divided Government and the Legislative Productivity of Congress, 1945-94.” *Legislative Studies Quarterly* 25:285–312.

King, Gary. 1988. "Statistical Models for Political Science Event Counts: Bias in Conventional Procedures and Evidence for the Exponential Poisson Regression Model." *American Journal of Political Science* 32:838–62.

King, Gary. 1989a. *Unifying Political Methodology.* Cambridge: Cambridge Univeristy Press.

King, Gary. 1989b. "Variance Specification in Event Count Models: From Restrictive Assumptions to a General Estimator." *American Journal of Political Science* 33:762–84.

King, Gary. 1995. "Replication, Replication." *PS: Political Science and Politics* 28:444–52.

Kitagawa, Genshiro. 1987. "Non-Gaussian State-Space Modeling of Nonstationary Time Series." *Journal of American Statistical Association* 82:1032–41.

McKenzie, Ed. 1988. "Some ARMA Models for Dependent Sequences of Poisson Counts." *Advances in Applied Probability* 20:822–35.

Mitchell, Sara M. and Will H. Moore. 2002. "Presidential Uses of Force during the Cold War: Aggregation, Truncation, and Temporal Dynamics." *American Journal of Political Science* 46:438–52.

Okawa, Kazushi, Nobukiyo Takamatsu and Yuzo Yamamoto. 1974. *National Income [in Japanese].* Tokyo: Toyo Keizai Shimpo Sha.

Tanabe, Kuniaki. 1995. "Social Policy in Japan: Building a Welfare State in a Conservative One Party Dominant System [in Japanese]." *Rebaiasan* 16:89–120.

West, Mike and Jeff Harrison. 1997. *Bayesian Forecasting and Dynamic Models.* 2nd ed. New York: Springer.

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

Zeger, Scott l. 1988. "A Regression Model for Time Series of Counts." *Biometrika* 75:621–29.