ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

An Assessment of Alternative State Space Models for Count Time Series

Ralph D Snyder, Gael M. Martin, Phillip Gould and Paul D. Feigin

May 2007

Working Paper 4/07

An Assessment of Alternative State Space Models for Count Time Series*

Ralph D Snyder, Gael M. Martin†, Phillip Gould
Department of Econometrics and Business Statistics, Monash University
and Paul D. Feigin
Faculty of Industrial Engineering and Management, Technion — Israel Institute of Technology

May 3, 2007

Abstract

This paper compares two alternative models for autocorrelated count time series. The first model can be viewed as a 'single source of error' discrete state space model, in which a time-varying parameter is specified as a function of lagged counts, with no additional source of error introduced. The second model is the more conventional 'dual source of error' discrete state space model, in which the time-varying parameter is driven by a random autocorrelated process. Using the nomenclature of the literature, the two representations can be viewed as observation-driven and parameter-driven respectively, with the distinction between the two models mimicking that between analogous models for other non-Gaussian data such as financial returns and trade durations. The paper demonstrates that when adopting a conditional Poisson specification, the two models have vastly different dispersion/correlation properties, with the dual source model having properties that are a much closer match to the empirical properties of observed count series than are those of the single source model. Simulation experiments are used to measure the finite sample performance of maximum likelihood (ML) estimators of the parameters of each model, and ML-based predictors, with ML estimation implemented for the dual source model via a deterministic hidden Markov chain approach. Most notably, the numerical results indicate that despite the very different properties of the two models, predictive accuracy is reasonably robust to misspecification of the state space form.

Key Words: Discrete state-space model; single source of error model; hidden Markov chain; overdispersion; forecasting low count data.

JEL Codes: C13, C22, C46, C53.

*This research has been supported by Australian Research Council Discovery Grant No. DP0450257.

†Corresponding author: Department of Econometrics and Business Statistics, P.O. Box, 11E, Monash University, Victoria, 3800, Australia. *(Email: Gael.Martin@Buseco.monash.edu.au).*

1 INTRODUCTION

Models for time series of counts are typically divided into two broad categories: 'observation-driven' and 'parameter-driven' models; see Cox (1981). In the former case, serial correlation in the counts is modelled directly via lagged values of the dependent variables, with various strategies adopted to ensure that the positive integer nature of the data is preserved (e.g. the binomial thinning operation used in the integer-valued autoregressive (INAR) class of models of Al-Osh and Alzaid, 1987, and McKenzie, 1988). In the case of parameter-driven models, correlation in the counts is introduced indirectly by specifying the parameter(s) of the conditional distribution for the counts as a function of a correlated latent stochastic process. The random parameter approach is, in turn, equivalent to the specification of a non-Gaussian state-space model for the discrete counts, in which both measurement and state equation contain a source of randomness. That is, this approach amounts to the specification of a dual source of error (DSOE) state-space model; see, for example, West, Harrison and Mignon (1985), Zeger (1988), Harvey and Fernandes (1989), West and Harrison (1997), Davis (2000), Durbin and Koopman (2000, 2001) and McCabe, Martin and Freeland (2006).[1]

An intermediate class of models comprises specifications such as the generalized linear autoregressive moving average (GLARMA) models of Shephard (1995) and Davis et al. (1999, 2003), the autoregressive conditional Poisson (ACP) model of Heinen (2003) and the autoregressive conditional ordered probit (ACOP) model of Jung et al. (2005). In these models the correlation in the counts is modelled indirectly by allowing (functions of) the parameter(s) of the conditional distribution for the observations to be both serially correlated and dependent on lagged counts. Such models are thus, in style, parameter-driven. However, in contrast with a conventional DSOE model, in which the latent parameter(s) are stochastic, the latent parameters in these models, conditional on lagged values of the counts, are deterministic, with no additional source of randomness introduced. As a consequence, such models can be referred to as single source of error (SSOE) models and would typically be classified as observation-driven.

The contrast between the SSOE and DSOE models for count data is analogous to the contrast between an (observation-driven) generalized autoregressive conditionally heteroscedastic (GARCH) model and a stochastic volatility (SV) model for financial returns; see Kim, Shephard and Chib (1998). It also mimics the contrast between the autoregressive conditional duration (ACD) model for trade durations (Engle and Russell, 1998) and the alternative stochastic conditional duration (SCD) model for the same data type (e.g. Bauwens

[1] The term 'dual', rather than 'multiple' is used here in order to emphasize the fact that randomness characterizes *both* the measurement and state equations. These equations could, of course, be defined for vectors, in which case the dual sources of error encompass multiple scalar error terms.

and Veradas, 2004; Strickland, Forbes and Martin, 2006). As in these other non-Gaussian settings, the relative merits of the SSOE and DSOE approaches to modelling counts remains an open empirical question. Whilst the latter approach may potentially yield more flexibility than the former approach, via the introduction of the additional source of random error, it does so at the cost of computational ease, with estimation of the dual source models often entailing the use of some form of computationally intensive simulation methodology (e.g. Chan and Ledolter, 1995, Durbin and Koopman, 2001, Fruhwirth-Schnatter and Wagner, 2004, and Jung et al., 2005).[2]

The aim of this paper is to provide a comparison of the alternative forms of discrete state space model. To keep the comparison manageable, we adopt a conditional Poisson distribution for the observations in each case, and allow the single parameter of that distribution to be (a function of) a latent process with autocorrelation of order one only. The SSOE model thus corresponds to the ACP(1,1) model considered by Heinen (2003) and Jung (2005), and the DSOE model to a special case of the model analysed by Chan and Ledolter (1995), Fruhwirth-Schnatter and Wagner (2004) and Jung et al. (2005), amongst others.

The nature of the comparison is two-fold. First, we use the theoretical properties of each model to produce a characterisation of the data types for which each model is suitable. Specifically, we derive a map of the feasible combinations of dispersion ($D = \frac{variance}{mean}$) and first-order autocorrelation (C) for each model. We demonstrate that, as well as there being no overlap in the feasible regions for the two alternative models, the DSOE model is suitable for a much wider range of data types than is the SSOE model, for which the feasible set of (D, C) pairs is very narrowly defined. Analysis of the (estimated) (D, C) characteristics of 13 empirical count time series indicates that for only one series is the SSOE model suitable (according to this criterion), with the DSOE model being appropriate in ten cases. For two series neither model is justified.

The second aspect of the comparison focusses on the estimation and predictive performance of the two models, with simulation experiments used to measure the finite sample performance of maximum likelihood (ML) estimators of the parameters of each model, and ML-based predictors. In order to place the two types of model on a similar computational footing, we present a deterministic ML estimation method for the DSOE model. The methodology is an extension of an algorithm suggested in McDonald and Zucchini (1997), and is based on a discretization of the space for the continuous random state variable. The discretized state process is treated as a hidden Markov chain, with evaluation of the likelihood

[2] In the work of West, Harrison and Mignon (1985) and Harvey and Fernandez (1989), natural conjugate distributions are chosen in order to enable non-simulation based inferential treatment of the discrete DSOE models adopted therein.

function occurring via a straightforward application of Markov chain methodology. In order to mimic an empirical scenario in which the true underlying process is not known, we also measure the accuracy with which a misspecified model predicts future values of the time series. The results suggest that there is some robustness to misspecification at the level of prediction, despite the very different characteristics of the two types of model.

The structure of the remainder of the paper is as follows. In Section 2 we introduce the proposed models for the integer data, outlining some of their important properties, including the feasible (D, C) regions. Estimation details are provided in Section 3, including an outline of the Markov chain approach used to estimate the DSOE model. Section 4 reports the results of simulation experiments, in which the estimation and forecasting performance of the two models, in a range of relevant scenarios, is documented, including forecasting performance under model misspecification. Section 5 concludes with some discussion of possible extensions of the current analysis. Various technical details are included in Appendices A to C.

2 TWO MODELS FOR COUNT TIME SERIES

We begin by defining a Poisson distribution, P(.), for the count time series variable, y_t, with (possibly) time-varying mean (and variance) parameter λ_t. Clearly, other discrete distributions could be chosen, with the choice being based on the empirical features of the data. For simplicity we use the Poisson distribution throughout the paper, as well as focussing on models without covariates.

The two models to be discussed, SSOE and DSOE, are distinguished one from the other by the specification of the error source for the state equation which, for the purpose of illustration, is assumed to contain first-order lags only, although models with higher-order lags could also be entertained. Specifically, the SSOE model is defined as:

$$y_t \sim \mathrm{P}\left(\lambda_t\right) \tag{1}$$

$$\lambda_t = \lambda + \phi\lambda_{t-1} + \alpha(y_{t-1} - \lambda_{t-1}), \tag{2}$$

for $t = 2, \ldots, T$, where the restrictions $\lambda > 0$, $\phi \geq \alpha \geq 0$ and $\phi \leq 1$ are imposed. When $\phi = \alpha = 0$, the model collapses to a Poisson process with constant mean. Denoting by $\mathcal{I}_{t-1}$ the information contained in $\{\lambda_1, y_1, y_2, \ldots, y_{t-1}\}$, from (2) it follows that, conditional on $\mathcal{I}_{t-1}$, the mean (and variance) of y_t is λ_t. The SSOE model in (1) and (2) is equivalent to the ACP(1,1) model specified in Heinen (2003) and also investigated in Jung *et al.* (2005). Replacing the conditional Poisson specification in (1) with a conditionally Gaussian specification, with mean λ_t and fixed variance, and imposing $\lambda = 0$ and $\phi \leq 1$ in (2), produces a local level model (Ord, Koehler and Snyder, 1997) that underlies the method of simple

exponential smoothing of Brown (1959) for y_t defined on the real line.

The DSOE model is defined as (1), but with

$$\lambda_t = h(x_t) \tag{3}$$

$$x_t = a + \kappa x_{t-1} + \eta_t;\ \eta_t \sim iidN(0, \sigma_\eta^2), \tag{4}$$

for $t = 1, 2, \ldots, T$, where $h(x_t)$ is any function that maps x_t into the positive space of λ_t and the stationarity restriction $|\kappa| < 1$ is imposed. Throughout the paper, for simplicity, we assume that $h(.)$ defines the exponential function. Conditionally on $\{\lambda_t; t = 1, 2, \ldots, T\}$, $\{y_t\}$ is assumed to be an independent sequence of Poisson counts, with corresponding mean and variance sequence $\{\lambda_t\}$. The model in (1), (3) and (4), extended to cater for covariates, is analysed in Chan and Ledolter (1995) and Jung et al. (2005), amongst other studies. Replacing the conditional Poisson specification in (1) with a conditionally Gaussian specification, with mean λ_t and fixed variance, and specifying $h(.)$ in (3) as the identity function, reproduces a standard linear Gaussian state space representation for y_t, appropriate when the random variable is defined on the real line (e.g. Harvey, 1991).

When comparing the alternative models we see two differences. One is the aforementioned use of a single source of error, $\{y_{t-1} - \lambda_{t-1}\}$, in the SSOE model, versus the use of $\{\eta_t\}$ in the DSOE model. The second difference is the scale for the recursion on λ_t : in the SSOE case (2) defines a recursion on λ_t itself, whereas in the DSOE case, assuming $h(\cdot) = \exp(\cdot)$ as we do, (4) defines a recursion on $\log(\lambda_t)$. Although it may have been more natural to also apply the SSOE paradigm to the natural Poisson parameters $\{\log(\lambda_t); t = 1, \ldots, T\}$, we follow the approach of Heinen (2003) and Jung et al. (2005) here and use the formulation in (2).

In Sections 2.1 and 2.2 we summarize the main properties of the two models. Some more details are provided in Appendices A and B. Note that although our results refer to the existence of a (unique) limiting stationary distribution of a process, we will often simply state that a process *is* stationary, even if it starts from y_1, which is not chosen from the stationary distribution. Moreover, moment and related calculations are performed under the stationary regime, even if not stated specifically.

2.1 Properties of the SSOE Model

We denote the (stationary) mean of $\{\lambda_t\}$ by $\mu = \frac{\lambda}{(1-\phi)}$.

Theorem 1 *Given (2), with* $0 \leq \alpha \leq \phi < 1$, $\{\lambda_t\}$ *is a stationary time series with moments,*

$$E(\lambda_t) = \mu \tag{5}$$

$$var(\lambda_t) = \mu \frac{\alpha^2}{(1-\phi^2)}. \tag{6}$$

Theorem 2 *Given (1) and (2), with* $0 \leq \alpha \leq \phi < 1$, *the count variable* $\{y_t\}$ *is a stationary time series with unconditional stationary moments*

$$\begin{aligned} M \equiv E(y_t) &= \mu \\ V \equiv var(y_t) & \\ &= \mu + var\left(\lambda_t\right) \\ &= \frac{\mu[1-\phi^2+\alpha^2]}{1-\phi^2} > \mu \\ cov(y_t, y_{t-1}) &= \frac{\mu\alpha[1-\phi^2+\phi\alpha]}{1-\phi^2}. \end{aligned}$$

First-order correlation and dispersion are given respectively by

$$C \equiv cor(y_t, y_{t-1}) = \frac{\alpha[1-\phi^2+\phi\alpha]}{[1-\phi^2+\alpha^2]} \tag{7}$$

$$D \equiv \frac{var(y_t)}{E(y_t)} = 1 + \frac{\alpha^2}{1-\phi^2}. \tag{8}$$

Heinen (2003) provides proofs of the above moment results and refers to the durations analysis of Engle and Russell (1998) for a demonstration of how the proof of stationarity would proceed. For completeness, in Appendix A we provide detailed information about the nature of the stationary distribution of the count model via a characterization of its Laplace transform. This information could be used to generate observations from the stationary distribution if one wanted to perform an exact likelihood analysis for this model.

Given Theorems 1 and 2, the following properties can also be derived.

SSOE: Property 1 Given $\phi < 1$ and $\lambda > 0$ and the resulting non-degenerate stationary distribution, the model does not suffer the fixed point problem ($y_t \to 0$ a.s. when $\lambda = 0$) highlighted in Grunwald, Hamza and Hyndman (1997).

SSOE: Property 2 Given $\phi < 1$, and conditional on λ_1, if $\alpha = 0$, the λ_t process is deterministic with a limiting value of μ, such that the stationary distribution of the system is $\lambda_t = \mu$ and $y_t \sim \mathrm{P}(\mu)$.

SSOE: Property 3 For fixed $0 \leq \alpha < 1$, C increases from $\alpha > 0$ to 1 as ϕ increases from α to 1. Negative correlation is thus not possible with the SSOE model.

SSOE: Property 4 Given $\{(\alpha,\phi): 0 \leq \alpha \leq \phi < 1\}$, $D \geq 1$. If $\alpha = 0$ then $D = 1$ (and y_t is equidispersed as a consequence), for any $\phi < 1$.

The following Corollary also follows from Theorem 2:

Corollary 1 *Given* $\{(\alpha,\phi): 0 \leq \alpha \leq \phi < 1\}$, *the feasible region for the dispersion* D, *given* $0 \leq C < 1$ *is*

$$\frac{1}{1-C^2} \quad \leq \quad D < \frac{1}{1-C}. \tag{9}$$

Proof: provided in Appendix B.

As is clear from Figure 1, the feasible region of (D, C) values implied by (9) is very limited, with there being only a narrow range of quite modest values of D possible for a given value of C, unless C is quite large ($C > 0.9$, say).

2.2 Properties of the DSOE Model

We denote the (stationary) mean and variance of x_t by $\mu_X = \frac{a}{1-\kappa}$ and $\sigma_X^2 = \frac{\sigma_\eta^2}{(1-\kappa^2)}$ respectively.

Theorem 3 *Given (3) and (4), with* $|\kappa| < 1$, $\{\lambda_t\}$ *is a stationary time series with moments,*

$$E(\lambda_t) = e^{\left\{\mu_X + 0.5\sigma_X^2\right\}} \tag{10}$$

$$var(\lambda_t) = e^{2\left\{\mu_X + \sigma_X^2\right\}} - e^{2\mu_X + \sigma_X^2}. \tag{11}$$

Proof: straightforward using the properties of the lognormal distribution.

Theorem 4 *Given (1), (3) and (4), with* $|\kappa| < 1$, *the count variable* $\{y_t\}$ *is a stationary time series with unconditional stationary moments*

$$\begin{aligned} M &\equiv E(y_t) \\ &= E\left(E\left[Y_t|\lambda_t\right]\right) \\ &= E(\lambda_t) \\ &= e^{\left\{\mu_X + 0.5\sigma_X^2\right\}} \\ V &\equiv var(y_t) \\ &= E(var\left[Y_t|\lambda_t\right]) + var(E\left[Y_t|\lambda_t\right]) \\ &= M + M^2(e^{\left\{\sigma_X^2\right\}} - 1) \\ cov(y_t, y_{t-1}) &= M^2\left[e^{\left\{\kappa\sigma_X^2\right\}} - 1\right], \end{aligned}$$

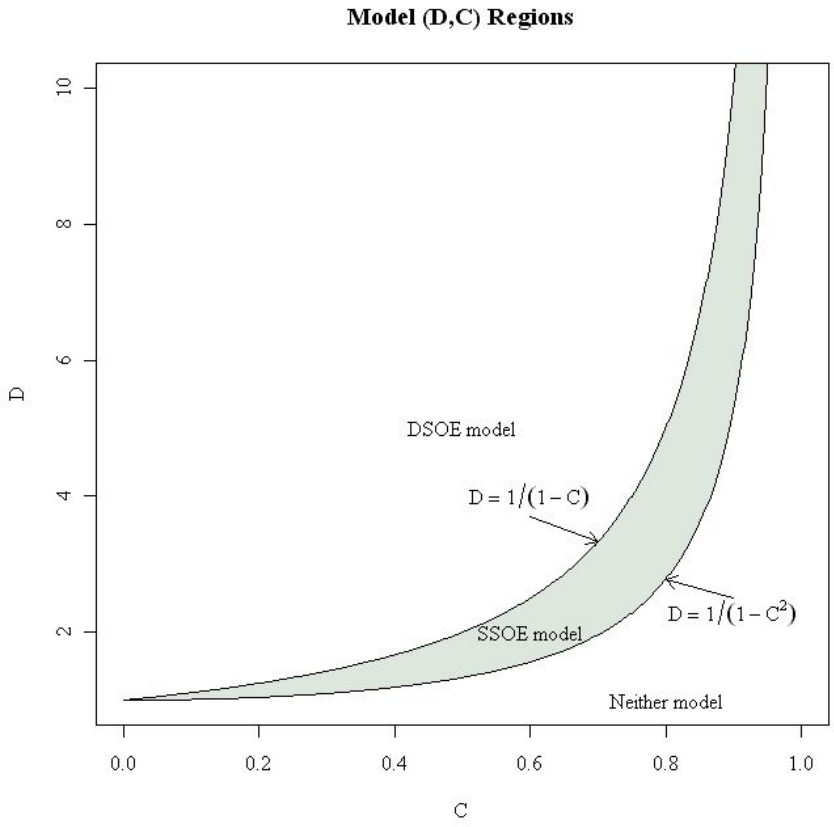

Figure 1: Feasible values of dispersion (D) and first-order autocorrelation (C) for the SSOE and DSOE models.

and correlation and dispersion given respectively given by

$$C \equiv cor(y_t, y_{t-1}) = \frac{e^{\left\{\kappa\sigma_X^2\right\}} - 1}{e^{\left\{\sigma_X^2\right\}} - 1 + 1/M} \tag{12}$$

$$D \equiv \frac{var(y_t)}{E(y_t)} = 1 + M(e^{\left\{\sigma_X^2\right\}} - 1). \tag{13}$$

Proof: straightforward using iterated expectations.

Given Theorems 3 and 4, the following properties can also be derived:

DSOE: Property 1 Given $\kappa < 1$, and conditional on x_0, if $\sigma_\eta = 0$, the x_t process is deterministic with a limiting value of μ_X, such that the stationary distribution of the system is $\lambda_t = e^{\mu_X}$ and $y_t \sim \mathrm{P}\left(e^{\mu_X}\right)$.

DSOE: Property 2 $C = 0 \iff$ either $\kappa = 0$ or $\sigma_\eta = 0$. C is an increasing function of κ and σ_η and is negative for $\kappa < 0$.

DSOE: Property 3 Given $\{(a, \kappa, \sigma_\eta) : a \in \mathcal{R}, |\kappa| < 1, \sigma_\eta > 0\}$, $D \geq 1$. *If* $\sigma_\eta = 0$, $D = 1$, for any $|\kappa| < 1$. As $|\kappa| \to 1$, $D \to \infty$, for any $\sigma_\eta > 0$.

The following Corollary also follows from Theorem 4:

Corollary 2 *Given* $\{(a, \kappa, \sigma_\eta) : a \in \mathcal{R}, |\kappa| < 1, \sigma_\eta > 0\}$, *the feasible region for the dispersion* D, *for* $0 \leq C < 1$, *is*

$$D \;>\; \frac{1}{1-C}. \tag{14}$$

For $-1 < C < 0$, $D > 1$.

Proof: provided in Appendix B.

Most notably, in comparison with the feasible region in (9) for the SSOE model, for any given value of C the possible values of D, beyond the lower bound of $\frac{1}{1-C}$, are unrestricted. Furthermore, in the positive correlation region, in which both models apply, there is *no overlap* in the feasible regions for (D, C) for the two alternatives. This result implies that the two models are suitable for analysing *different* types of correlated count data, rather than being alternative models for any given data set. Only the DSOE model is applicable to the negative correlation case.

2.3 Empirical Count Time Series

In Table 1 the descriptive properties of 13 empirical count time series are summarized. The data sets fall into six different categories which span a wide range of empirical settings. In particular, all data series are *low* count series, thereby requiring explicit modelling via a discrete conditional distribution.[3]

1. Australian monthly road accident deaths and injuries: (i) fatalities in 60km/h zones in Victoria, Australia from January 1996 to June 2002 (FAT_60); (ii) fatalities in 40-60km/h zones in Victoria from January 1996 to February 2005 (FAT_4060); (iii) non-fatal police-car-related injuries in New South Wales from January 1987 to December 2004 (POLICE).

2. Australian monthly deaths by other (non-road-related) causes: (i) medical-related deaths (MEDICAL); (ii) choking-related deaths (CHOKE); (iii) death inflicted by another but with undetermined intent (UNDETERM); (iv) death inflicted by another but unintentionally (UNINTENT); (v) deaths by late effects (LATE).

3. Monthly wage loss benefit claims for injuries sustained in the British Columbia (Canada) logging industry: (i) claims for burn injuries from January 1984 to December 1994 (BURNS); (ii) claims for cut and/or laceration injuries for the same period (CUTS).

4. Annual corporate failures (defaults) of investment grade firms from the U.S., over the period 1920-2000.

5. One-minute trade counts for the Australian firm Broken Hill Proprietary (BHP) Limited for 1 August 2001, 10.00am to 4.00pm.

6. Daily admissions for asthma treatment to a Sydney hospital from 1 January 1990 to 31 December 1993.

The claims data sets in 3. have been extensively analysed in the literature using INAR-type specifications, most notably in Freeland and McCabe (2004a,b), McCabe and Martin (2005) and Zhu and Joe (2006). The admissions data in 6. have been analysed previously by Davis et al. (1999) using a GLARMA model and Davis et al. (2000) using a generalized linear model (with allowance made for an autocorrelated latent process). Jung et al. (2005) also use the admissions data in an empirical comparison of a range of count time series models, including SSOE and DSOE models, with the dual source model estimated via a

[3]This is in contrast to one of the earliest analyses of count times series, Harvey and Durbin (1986), in which a conditional Gaussian assumption was adopted due to the large count nature of the data.

simulation-based method. All of the road accident data in 1. and death data in 2. have been provided to the authors by the Monash University Accident Research Centre.

In Table 1, the term in the superscript indicates which model fits the sample (D, C) characteristics of the observed data (S = SSOE model; D = DSOE model; N = neither model). As is clear, only one series (CUTS) falls into the narrow feasible set for the SSOE model ($\frac{1}{1-C^2} \leq D < \frac{1}{1-C}$), with nine of the series with positive correlation satisfying the DSOE criterion ($D > \frac{1}{1-C}$). Two of the series are underdispersed; hence neither model is suitable. The dispersion value of the single series with negative autocorrelation (LATE) exceeds the lower bound of the DSOE model ($D > 1$), the only model of the two valid for the negative correlation case. Overall then, the DSOE model appears to be vastly more suitable for empirical analysis, according to this particular criterion. In Section 4, however, we using artificially simulated data to document the fact that despite this mismatch of the SSOE model with the typical dispersion/correlation characteristics of observed data, its forecasting performance, even when misspecified for the data, is still competitive with that of a correctly specified DSOE model.

3 ML ESTIMATION

3.1 SSOE Model

ML estimation of the full set of unknown parameters in (1) and (2), $\Phi = (\lambda, \phi, \alpha)$, is based on maximizing the likelihood function (conditioned on λ_1 and y_1),

$$L(\Phi) = \sum_{t=2}^{T} P(y_t|\mathcal{I}_{t-1}; \Phi),$$

where

$$P(y_t|\mathcal{I}_{t-1}; \Phi) = \frac{\lambda_t^{y_t} \exp(-\lambda_t)}{y_t!}$$

and $\mathcal{I}_{t-1}$ is as defined previously, with $\mathcal{I}_1 = \{\lambda_1, y_1\}$. The initial value of λ_t, λ_1, is equated with $\mu = \frac{\lambda}{(1-\phi)}$ and y_1 to the observed value at $t = 1$. Note that one could approximate the exact likelihood by using the stationary distributions specified in Appendix A for y_1 and λ_1 to simulate initial values and average with respect to each conditional likelihood so produced. However, the less computationally intensive conditional ML method is used in the simulation work in Section 4.

Table 1:

Summary Statistics for Empirical Count Time Series

Time Series $y_t^{(a)}$	Freq.	Sample Size (T)	Sample Mean (M)	Sample Variance (V)	Sample Dispersion $(D = V/M)$	Sample ACF(1) (C)	SSOE lb $\frac{1}{1-C^2}$	DSOE lb $\frac{1}{1-C}$
CUTS$^{(S)}$	MT$^{(b)}$	120	6.13	11.80	1.92	0.56	1.46	2.27
MEDICAL$^{(D)}$	MT	84	2.08	2.66	1.27	0.08	1.01	1.09
FAT_40-60$^{(D)}$	MT	110	6.97	10.32	1.48	0.14	1.02	1.17
POLICE$^{(D)}$	MT	72	7.22	14.25	1.97	0.15	1.02	1.25
CHOKE$^{(D)}$	MT	84	3.24	7.12	2.20	0.18	1.03	1.22
UNDETERM$^{(D)}$	MT	84	1.73	2.73	1.58	0.22	1.05	1.28
ASTHMA$^{(D)}$	D$^{(c)}$	1461	1.94	2.70	1.39	0.25	1.07	1.33
TRADES$^{(D)}$	ID$^{(d)}$	360	4.52	15.07	3.33	0.33	1.13	1.50
UNINTENT$^{(D)}$	MT	84	2.96	10.25	3.46	0.39	1.18	1.65
DEFAULTS$^{(D)}$	A$^{(e)}$	85	1.74	9.62	5.53	0.52	1.38	2.09
LATE$^{(D)}$	MT	84	1.62	1.97	1.22	-0.14	1.02	0.88
FAT_60$^{(N)}$	MT	78	2.06	2.01	0.97	0.15	1.02	1.18
BURNS$^{(N)}$	MT	120	0.18	0.16	0.93	0.22	1.05	1.29

(a) Model that fits the sample (D,C) values for y_t is denoted by the superscript: S = SSOE model; D = DSOE model; N = neither model.

(b) MT = monthly frequency

(c) D = daily frequency

(d) ID = intraday frequency

(e) A = annual frequency

3.2 DSOE Model

In the case of the DSOE model in (1) and (4), the likelihood function for the unknown parameter vector, $\boldsymbol{\Theta} = (a, \kappa, \sigma_\eta)$, is defined by the T-dimensional integral

$$L(\boldsymbol{\Theta}) = p(\mathbf{y}|\boldsymbol{\Theta}) = \int_{\boldsymbol{\lambda}} p(\mathbf{y}|\boldsymbol{\lambda}) p(\boldsymbol{\lambda}|\boldsymbol{\Theta}) d\boldsymbol{\lambda}, \tag{15}$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_T)'$ and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_T)'$. Typically, the integral in (15) that defines the (marginal) distribution of $\mathbf{y}$, $p(\mathbf{y}|\boldsymbol{\Theta})$, would be approximated via some form of simulation method; see, for example, Durbin and Koopman (2001), Fruhwirth-Schnatter and Wagner (2004) and Jung et al. (2005). The alternative approach of adopting conjugate forms for $p(\boldsymbol{\lambda}|\boldsymbol{\Theta})$ to enable analytical treatment of the integral has been adopted by Harvey and Fernandez (1989), amongst others.

We propose a deterministic numerical approach, based on an extension of an algorithm outlined in McDonald and Zucchini (1997). This approach produces a direct numerical evaluation of $p(\mathbf{y}|\boldsymbol{\Theta})$ via the treatment of the continuous state variable x_t in (4) as a discrete hidden Markov chain process.[4] Specifically, we approximate the continuous normal distribution of x_t with a discrete mass function, defined over a grid of N_x discrete states $x_{(1)}, x_{(2)}, ..., x_{(N_x)}$. Let $\mathcal{I}_{t-1}$ again denote the information set up to and including period $t-1$. Given (4), $\mathcal{I}_{t-1} = \{x_0, y_1, y_2, \ldots, y_{t-1}\}$ with $\mathcal{I}_0 = \{x_0\}$. Transitions between states over time are governed by an ergodic first-order Markov chain with transition probability matrix $\mathbf{P} = [p_{ij}]$ where the elements of $\mathbf{P}$,

$$p_{ij} = \Pr\left\{x_t = x_{(j)} | x_{t-1} = x_{(i)}, \mathcal{I}_{t-1}\right\}, \tag{16}$$

are calculated as normal probability increments. Defining the $(N_x \times 1)$ state probability vector in period t by $\boldsymbol{\pi}(t|\mathcal{I}_{t-1}) = [\pi_j(t|\mathcal{I}_{t-1})]$, where

$$\pi_j(t|\mathcal{I}_{t-1}) = \Pr\left\{x_t = x_{(j)} | \mathcal{I}_{t-1}\right\},$$

successive state probability vectors may be generated via

$$\boldsymbol{\pi}(t|\mathcal{I}_{t-1}) = \boldsymbol{\pi}(t-1|\mathcal{I}_{t-2})\mathbf{P}. \tag{17}$$

Given $|\kappa| < 1$, the chain is ergodic.

The unobserved state variable x_t is related to the observed count variable y_t by an $(N_x \times N_y)$ probability matrix $\mathbf{Q} = [q_{ij}]$, where the elements of $\mathbf{Q}$,

$$q_{ij} = \Pr\left\{y_t = y_{(j)} | x_t = x_{(i)}, \mathcal{I}_{t-1}\right\}, \tag{18}$$

[4] See also White et al. (2004) for a similar algorithm developed independently of the algorithm in this paper, and applied in the context of a stochastic volatility model for continuous returns.

are determined by the conditional Poisson distribution in (1). The marginal distribution of y_t is in turn defined by

$$\Pr(y_t = y_{(j)}|\mathcal{I}_{t-1}) = \sum_{k=1}^{N_x} q_{kj}\pi_k(t|\mathcal{I}_{t-1}). \tag{19}$$

Using this notation, the likelihood function in (15) can be re-written as

$$L(\boldsymbol{\Theta}) = p(\mathbf{y}|\boldsymbol{\Theta}) = \prod_{t=1}^{T} P(y_t|\mathcal{I}_{t-1}, \boldsymbol{\Theta}). \tag{20}$$

At any given point in the parameter space, each component of (20) can be numerically evaluated via application of a Markov chain filter, using the expressions in (16) to (19), with a standard optimization algorithm then used to produce the ML estimator of $\boldsymbol{\Theta}$. Details of the filter are given in Appendix C.

4 SIMULATION EXPERIMENTS

In Section 2.3 we documented the fact that the DSOE model matches the (estimated) dispersion/correlation characteristics of typical empirical count data much more closely than does the SSOE model. Nevertheless, in any specific instance, either model may be appropriate and it is of interest to assess the accuracy with which the parameters of each model are estimated, in finite samples. More importantly, it is of interest to gauge the relative forecasting accuracy of the correctly specified and misspecified models and, in particular, ascertain the extent to which any mismatch of the SSOE model with the dispersion/correlation features of a given data set impinges on its forecasting performance.

Results associated with particular DSOE and SSOE data generating processes (dgps) are presented. The parameter settings are chosen to resemble, to some extent, the range of empirical features of the multiple data sets summarized in Section 2.3. Three dgps are entertained for each model specification, corresponding to the low first-order autocorrelation ($C = 0.1$) and medium autocorrelation ($C = 0.5$) that characterize various of the observed data sets, and with a high autocorrelation case ($C = 0.8$) included for completeness. The three intrinsic parameters are then set so as to produce a corresponding value for D that is in the feasible region for the model and a value for M which is approximately equal to 5, a mean value that is appropriate for the low count data that is the focus here. For each dgp, 1000 samples of length $T = 100$ and $T = 500$ are generated, with ML estimation of the DSOE model implemented via the grid-based method described in Section 3.2, with $N_x = 100$ grid-points.

Measures of estimation performance are designed to allow for cases in which extreme estimates are produced, in particular of the variance parameter of the DSOE model. Specifically, we calculate the median, rather than the mean, of the estimator $\hat{\theta}$ of any unknown parameter θ, over simulated samples, measuring bias by the difference between this figure and the true parameter value. In order to allow for comparison across the two different models, we report accuracy using relative bias — denoted by RBIAS in the tables — i.e. bias as a proportion of the true value of the parameter. We measure precision via the median of the relative absolute deviations ($|\hat{\theta} - \theta|/\theta$) — denoted by MRAD.

Invoking a mean squared error loss criterion, the optimal prediction of y_{T+1} is the conditional mean, $E(y_{T+1}|\mathcal{I}_T) = \lambda_{T+1}$.[5] When the DSOE model is used to produce the prediction, λ_{T+1} itself is a random quantity and is, in turn, predicted by $E(\lambda_{T+1}|\mathcal{I}_T) = E(\exp(x_{T+1})|\mathcal{I}_T)$. This expectation is estimated using the discretized distribution $\boldsymbol{\pi}(T+1|\mathcal{I}_T)$, based on the ML estimates of the DSOE parameters (see Appendix C). When the SSOE model is used to produce the prediction, conditional on y_T and λ_T, λ_{T+1} is a deterministic function of the parameters. As such, λ_{T+1} is estimated from the recursion (2) with ML estimates, $\widehat{\lambda}$, $\widehat{\phi}$ and $\widehat{\alpha}$, of the SSOE parameters replacing λ, ϕ and α respectively, and λ_T specified via the recursion, with $\lambda_1 = \widehat{\lambda}/(1-\widehat{\phi})$. In order to be consistent with the way in which the within-sample results are reported we use median quantities to summarize the prediction results. Specifically, as a measure of relative prediction bias we report median value of $(\widehat{\lambda}_{T+1} - y_{T+1})/y_{T+1}$ in the 1000 replications. MRAD is calculated as the median of the relative absolute deviations, $(|\widehat{\lambda}_{T+1} - y_{T+1}|/y_{T+1})$.

Tables 2, 3 and 4 respectively report the results for the low, medium and high correlation cases, with results relating to the DSOE dgp appearing in the upper panel of each table, and results for the SSOE dgp appearing in the lower panel. Results for both sample sizes, $T = 100$ and $T = 500$, are recorded in each table, with RBIAS and MRAD reported both for individual model parameters, and as averages across the three parameters that characterize each of the two models. In the averaging of the RBIAS results across parameters, the absolute value of the bias figures is taken before averaging, in order to avoid the cancellation of negative and positive biases.

The results in Table 2 indicate that for the low correlation setting ($C = 0.1$) and for $T = 100$, the parameters of the SSOE model are estimated with much larger relative bias than those of the DSOE model, but with similar precision, as measured by MRAD. For $T = 500$, however, the SSOE parameters are estimated with less bias and (slightly) more

[5] We refrain in this paper from discussing the issues associated with using a non-integer forecast of the integer random variable. See Freeland and McCabe (2004a) and McCabe and Martin (2005) for further discussion of this point.

precision than the DSOE parameters. More detailed consideration of the results for the SSOE dgp highlight the fact that there are particular problems with estimating α near its lower bound of zero (necessary to ensure the low value for C) when the sample size is only 100.

In the medium correlation case ($C = 0.5$; Table 3), the DSOE model is estimated with less bias and more precision than the SSOE model, for $T = 100$, and more precisely, according to the MRAD measure, for $T = 500$. For the high correlation setting however ($C = 0.8$; Table 4), all results unambiguously favour the SSOE model. Considering the average |RBIAS| and MRAD values for each parameter set, the anticipated reduction in bias and increase in precision associated with the larger sample size is uniform for the SSOE model, for all three values of C. For the DSOE model, the MRAD is smaller, as expected, for the larger sample size (and for all C); however |RBIAS| increases with T in the low and medium correlation cases, although only very slightly in the latter case.

In summary then, there is variation, over C, in the relative bias and precision with which the parameters of the two alternative models are estimated. The SSOE model is the easier of the two to estimate accurately when the data accords with that model and has high autocorrelation. In the low correlation case, as long as the sample size is not too low, the SSOE model again appears to be preferable in terms of estimation accuracy. However, in the medium correlation case, the DSOE model is able to be estimated more accurately, overall, again when that model matches the true dgp.

In contrast to the inconclusive ranking of estimation performance, the ranking of predictive performance is uniform: across all three values of C, for both sample sizes, and according to the respective measures of bias and precision (RBIAS and MRAD), the SSOE model produces more accurate and precise predictions of y_{T+1}, as generated from the SSOE model, than does the DSOE model of y_{T+1} generated from the DSOE model. Hence, notwithstanding the fact that the SSOE model appears to be the more restrictive of the two according to the analysis in Section 2, if it *were* the correct model for a particular count time series, it would enable more accurate predictions to be produced than if the DSOE model had been the appropriate model.

Crucially, the results in Tables 2 to 3 also illustrate that predictions from a *misspecified* model - whether a misspecified SSOE model under a DSOE dgp, or vice-versa - produce MRAD and RBIAS values that are usually quite similar to (sometimes even smaller than) the corresponding values for the correctly specified model. This result has practical import, as it suggests that in any empirical situation, in which the investigator does not know the true model, and may not wish to select a model on the basis of estimated dispersion/correlation properties alone, predictions are quite robust to misspecification of that model, at least within

the confines of the dual possibility set considered here. In particular, the fact that an SSOE model is unlikely to be suitable for many count series, in terms of dispersion/correlation properties, does not appear to preclude it from producing acceptable forecasts.

5 CONCLUSIONS

In this paper some results have been documented regarding the performance of two alternative discrete state space models, one of which drives the parameter of the conditional Poisson distribution via a first-order autoregressive stochastic process, and the other of which expresses that parameter as a function of last period's observed count. Most notably, it has been demonstrated that the dispersion/correlation regions over which the two models are valid do not overlap, with the single source of error model having a feasible region that is much more narrowly defined than that of the dual source of error model. The empirical properties of multiple count time series match the theoretical dual source model, with only one of the 13 series considered matching the dispersion/correlation properties of the single source model. That said, simulation experiments are used to demonstrate the fact that when the single source model *is* appropriate, and for certain correlation settings, its parameters are estimated more accurately, via maximum likelihood, than the dual source model, at least when the latter is estimated using a Markov chain-based approximate maximum likelihood approach. Moreover, the single source model appears able to predict future values from that model more accurately than the dual source model can predict its own values. Perhaps most importantly however, the simulation results also indicate that one-step-ahead forecasts are reasonably robust to misspecification of the state space form.

Obvious extensions to this analysis include the generalization of the conditional distribution to more flexible distributions than the Poisson, such as the negative binomial and the double Poisson, that allow for separate dynamic specifications for the mean and variance (see, for e.g. Heinen, 2003, and McCabe et al., 2006). In particular, a more flexible parameterization of the conditional distribution may lead to a larger feasible dispersion/correlation region for the single source model than is associated with the single parameter Poisson case. The extension to higher order lags also need to be considered, including the impact of that extension on the computational efficiency of the grid-based estimation method introduced here for the dual source model. Certainly, more fine tuning of the latter estimation method also needs to be conducted, including an investigation of the impact of the number and distribution of the grid points on estimation accuracy. A comparison of the accuracy of the deterministic method with the simulation-based methods commonly adopted in the dual source framework would also be of interest. Finally, although the focus here is on count data,

Table 2:

Simulation Results: low correlation case ($C = 0.1$) for both the DSOE and SSOE dgp;
$T = 100$, 500; No. of replications $= 1000$

		$T = 100$		$T = 500$	
Estimand	True Value	RBIAS[(a)]	MRAD[(b)]	RBIAS	MRAD
DSOE DGP					
a	1.077	-0.102	0.388	-0.207	0.242
κ	0.310	0.226	0.883	0.481	0.540
σ_η	0.294	-0.099	0.200	-0.058	0.102
		0.142[(c)]	0.490[(d)]	0.249[(c)]	0.295[(d)]
		Correct forecast model (DSOE)		Correct forecast model (DSOE)	
y_{T+1}	5.000[(e)]	0.070	0.321	0.084	0.325
		Misspecified forecast model (SSOE)		Misspecified forecast model (SSOE)	
y_{T+1}	5.000[(e)]	0.121	0.368	0.036	0.330
SSOE DGP					
λ	0.163	0.288	0.332	0.331	0.541
ϕ	0.967	-0.010	0.013	-0.011	0.018
α	0.057	-1.000	1.000	-0.035	0.275
		0.433[(c)]	0.448[(d)]	0.126[(c)]	0.278[(d)]
		Correct forecast model (SSOE)		Correct forecast model (SSOE)	
y_{T+1}	5.000[(e)]	-0.023	0.304	-0.028	0.291
		Misspecified forecast model (DSOE)		Misspecified forecast model (DSOE)	
y_{T+1}	5.000[(e)]	0.007	0.299	0.029	0.289

(a) Difference between the median of the 1000 parameter estimates (or predictions) and the true value, as a ratio of the true value.

(b) Median of the relative absolute deviations between the estimate (or prediction) and the true value. (Note this is NOT necessarily what is calculated and reported as yet for the prediction case)

(c) The average of the absolute value of RBIAS over the three individual parameters.

(d) The average of MRAD over the three individual parameters.

(e) Median of the y_{T+1} values simulated in the 1000 replications of the relevant dgp.

Table 3:

Simulation Results: medium correlation case ($C = 0.5$) for both the DSOE and SSOE dgp; $T = 100$, 500; No. of replications $= 1000$

		$T = 100$		$T = 500$	
Estimand	True Value	RBIAS[(a)]	MRAD[(b)]	RBIAS	MRAD
DSOE DGP					
a	0.221	0.023	0.230	-0.063	0.123
κ	0.851	-0.006	0.048	0.013	0.024
σ_η	0.269	-0.108	0.184	-0.093	0.103
		0.046[(c)]	0.154[(d)]	0.056[(c)]	0.083[(d)]
		Correct forecast model (DSOE)		Correct forecast model (DSOE)	
y_{T+1}	4.000[(e)]	0.118	0.445	0.102	0.422
		Misspecified forecast model (SSOE)		Misspecified forecast model (SSOE)	
y_{T+1}	4.000[(e)]	0.154	0.394	0.165	0.383
SSOE DGP					
λ	0.459	0.468	0.585	0.057	0.205
ϕ	0.908	-0.046	0.059	-0.007	0.022
α	0.296	-0.045	0.199	-0.013	0.087
		0.186[(c)]	0.281[(d)]	0.026 [(c)]	0.105[(d)]
		Correct forecast model (SSOE)		Correct forecast model (SSOE)	
y_{T+1}	5.000[(e)]	0.012	0.319	0.052	0.332
		Misspecified forecast model (DSOE)		Misspecified forecast model (DSOE)	
y_{T+1}	5.000[(e)]	-0.021	0.309	0.046	0.310

(a) Difference between the median of the 1000 parameter estimates (or predictions) and the true value, as a ratio of the true value.

(b) Median of the relative absolute deviations between the estimate (or prediction) and the true value.

(c) The average of the absolute value of RBIAS over the three individual parameters.

(d) The average of MRAD over the three individual parameters.

(e) Median of the y_{T+1} values simulated in the 1000 replications of the relevant dgp.

Table 4:

Simulation Results: high correlation case ($C = 0.8$) for both the DSOE and SSOE dgp;
$T = 100$, 500; No. of replications $= 1000$

		$T = 100$		$T = 500$	
Estimand	True Value	RBIAS[(a)]	MRAD[(b)]	RBIAS	MRAD
DSOE DGP					
a	0.037	1.027	1.081	0.027	0.236
κ	0.971	-0.036	0.036	0.003	0.012
σ_η	0.200	-0.045	0.185	-0.065	0.097
		0.369[(c)]	0.434[(d)]	0.032[(c)]	0.115[(d)]
		Correct forecast model (DSOE)		Correct forecast model (DSOE)	
y_{T+1}	3.000[(e)]	0.227	0.680	0.145	0.606
		Misspecified forecast model (SSOE)		Misspecified forecast model (SSOE)	
y_{T+1}	3.000[(e)]	0.241	0.479	0.351	0.477
SSOE DGP					
λ	0.801	0.128	0.265	0.025	0.118
ϕ	0.840	-0.036	0.056	-0.007	0.025
α	0.728	-0.030	0.084	-0.005	0.041
		0.065[(c)]	0.135[(d)]	0.012 [(c)]	0.061[(d)]
		Correct forecast model (SSOE)		Correct forecast model (SSOE)	
y_{T+1}	4.000[(e)]	0.053	0.392	0.066	0.397
		Misspecified forecast model (DSOE)		Misspecified forecast model (DSOE)	
y_{T+1}	4.000[(e)]	0.003	0.355	0.047	0.332

(a) Difference between the median of the 1000 parameter estimates (or predictions) and the true value, as a ratio of the true value.

(b) Median of the relative absolute deviations between the estimate (or prediction) and the true value.

(c) The average of the absolute value of RBIAS over the three individual parameters.

(d) The average of MRAD over the three individual parameters.

(e) Median of the y_{T+1} values simulated in the 1000 replications of the relevant dgp.

these results suggest that similar results may obtain for any data defined on a restricted domain. In particular, an obvious topic for future research is the production of comparable results for the alternative state space representations of positive durations data.

Appendix A: The Stationary Distribution of $\{\lambda_t\}$ and $\{y_t\}$

A.1 The SSOE model

We seek to characterize the stationary distribution of y_t and λ_t when $\phi < 1$, and derive an expression for their Laplace transforms. Note that the pair $\{y_t, \lambda_t\}$ form a bivariate Markov chain. Denote the Laplace transform of y_t by $L_t(\cdot)$ and that of λ_t by $M_t(\cdot)$. Given the conditional Poisson distribution for y_t in (1), and defining $\mathcal{I}_{t-1}$ as in the text, it follows that

$$\begin{aligned} L_t(u) &= E\left[E(\exp(-uy_t)|\mathcal{I}_{t-1}\right] \\ &= E\exp\left[-\lambda_t(1-e^{-u})\right] \\ &= M_t(1-e^{-u}). \end{aligned} \tag{21}$$

If the limit $M(\cdot)$ of $M_t(\cdot)$ exists, then the Laplace transform, $L(\cdot)$, of the stationary distribution of y_t will also exist and satisfy

$$L(u) = M(1-e^{-u}). \tag{22}$$

We begin by focussing on the convergence of $M_t(\cdot)$ to $M(\cdot)$. We provide an outline here — a complete proof can be obtained from the authors.

From (2) we deduce that

$$\begin{aligned} M_t(v) &= E\left[E(\exp\left(-v\lambda_t\right)|\mathcal{I}_{t-2})\right] \\ &= \exp\left\{-v\lambda\right\} M_{t-1}\left(g(v;\delta,\alpha)\right), \end{aligned} \tag{23}$$

where

$$\begin{aligned} g(v) &= g(v;\delta,\alpha) \\ &= v\delta + 1 - e^{-v\alpha}. \end{aligned}$$

and $1 > \delta \equiv \phi - \alpha \geq 0$. Define the k-th iterate $g^{(k)}$ of g by

$$g^{(k)}(v) = g\left(g^{(k-1)}(v)\right);\ g^{(0)}(v) \equiv v \tag{24}$$

Applying (23) iteratively we deduce

Theorem 5 *Given* $0 \leq \phi < 1$, $M_t(v)$ *converges to*

$$M(v) = \exp\left\{-\lambda \sum_{k=0}^{\infty} g^{(k)}(v)\right\} \tag{25}$$

as $t \to \infty$, *where* $\sum_{k=0}^{\infty} g^{(k)}(v) < \infty$.

By (22) and (25), the Laplace transform of the stationary distribution of y_t satisfies

$$L(u) = \exp\left\{-\lambda \sum_{k=0}^{\infty} g^{(k)}(1 - e^{-u})\right\}$$

for $u \geq 0$.

A.2 The DSOE model

The existence of a unique limiting stationary distribution for y_t follows from the fact that $\{y_t, \lambda_t\}$ is a bivariate Markov chain and that $\{x_t = \log(\lambda_t)\}$ is a Gaussian autoregressive process with known stationary distribution under the conditions given (namely, $|\kappa| < 1$). Conditionally on $\{\lambda_t\}$, the y_t are independent Poisson variables with $E(y_s|\{\lambda_t\}) = \lambda_s$.

Appendix B: Feasible (D,C) Regions

B.1 The SSOE model

From

$$\begin{aligned} M &= E(y_t) = \frac{\lambda}{1-\phi} \\ V &= var(y_t) = \frac{M[1-\phi^2+\alpha^2]}{1-\phi^2} \\ C &= \mathrm{cor}(y_t, y_{t-1}) = \frac{\alpha[1-\phi^2+\phi\alpha]}{[1-\phi^2+\alpha^2]} \qquad (26) \\ D &= \frac{\mathrm{var}(y_t)}{\mathrm{E}(y_t)} = 1 + \frac{\alpha^2}{1-\phi^2}, \qquad (27) \end{aligned}$$

we can deduce the following inversions from (M, V, C) to (λ, ϕ, α):

$$
\begin{aligned}
D &= \frac{V}{M} \\
\phi &= C + \sqrt{\frac{D(1-C^2)-1}{D(D-1)}} \qquad (28) \\
\alpha &= CD - \phi(D-1) \qquad (29) \\
\lambda &= M(1-\phi).
\end{aligned}
$$

The solution (28) for ϕ in terms of C and D follows by noting that (29) follows from (26) and (27), with substitution of (29) into (27) yielding

$$D = 1 + \frac{\left(CD - \phi\left(D-1\right)\right)^2}{1-\phi^2} \ . \qquad (30)$$

Solving the following quadratic equation in ϕ, given C and D,

$$\phi^2 - 2C\phi + \frac{C^2D^2 - D + 1}{D(D-1)} = 0$$

we obtain (after showing from (26) that $C \leq \phi$) the solution in (28), as well as the condition that $D(1-C^2) \geq 1$; that is,

$$D \geq \frac{1}{(1-C^2)}. \qquad (31)$$

Solving (26) for α we obtain:

$$(\phi - C)\alpha^2 + (1-\phi^2)\alpha - (1-\phi^2)C = 0,$$

with (positive) solution

$$\alpha = \frac{-1 + \sqrt{1 + 4C(\phi - C)/(1-\phi^2)}}{2(\phi - C)/(1-\phi^2)} \ .$$

Thus D from (27) can be written as:

$$D = 1 + \frac{1 + 2C(\phi - C)/(1-\phi^2) - \sqrt{1 + 4C(\phi - C)/(1-\phi^2)}}{2(\phi - C)^2/(1-\phi^2)}.$$

Now $\alpha \leq C \leq \phi < 1$ follows from (26) — since $\phi \geq \alpha$ so that $C/\alpha \geq 1$ — and from (28). Thus, the range of values of D as a function of given C can be determined by allowing ϕ to vary from C through to 1. As $\phi \uparrow 1$, $D \to 1 + C\frac{1-C}{(1-C)^2} = \frac{1}{1-C}$. Defining $U = (\phi - C)$ and $k = 4C/(1-\phi^2)$, as $\phi \downarrow C$ it follows that $U = (\phi - C) \downarrow 0$ and $k \downarrow 4C/(1-C^2)$. Given that

$$D = 1 + \frac{2 + kU - 2\sqrt{1+kU}}{kU^2/C} \sim 1 + \frac{2 + kU - 2(1 + \frac{1}{2}kU - \frac{1}{8}k^2U^2)}{kU^2/C} \sim 1 + \frac{C^2}{(1-C^2)},$$

it follows that as $\phi \downarrow C$, $D \to \frac{1}{1-C^2}$. From (31) it follows that D approaches $\frac{1}{1-C^2}$ from above. Moreover, from (26) one can also show that $\frac{C-\alpha}{\phi-C} = \frac{\alpha^2}{1-\phi^2}$, so that for each ϕ there exists α such that $D = 1 + \frac{C-\alpha}{\phi-C} \leq 1 + \frac{C}{\phi-C}$. As the latter holds for any $1 > \phi \geq C$, we can conclude that D approaches $\frac{1}{1-C}$ from below, i.e. that

$$D < \frac{1}{1-C}. \tag{32}$$

From (31) and (32) it follows that (9) defines the feasible range of D values for any $0 \leq C < 1$, with both bounds being tight. Note that when $C = 0$ then $D = 1$ (and $\phi = \alpha = 0$).

B.2 The DSOE model

From

$$\begin{aligned} M &= E(y_t) = e^{\{\mu_X + 0.5\sigma_X^2\}} \\ V &= var(y_t) = M + M^2(e^{\{\sigma_X^2\}} - 1) \\ C &= \text{cor}(y_t, y_{t-1}) = \frac{e^{\{\kappa\sigma_X^2\}} - 1}{e^{\{\sigma_X^2\}} - 1 + 1/M} \qquad (33) \\ D &= \frac{\text{var}(y_t)}{\text{E}(y_t)} = 1 + M(e^{\{\sigma_X^2\}} - 1), \qquad (34) \end{aligned}$$

we can deduce the following inversions from (M, V, C) to (a, κ, σ_η):

$$\begin{aligned} D &= \frac{V}{M} \\ \kappa &= \frac{\log\left[\frac{CD}{M} + 1\right]}{\log\left[\frac{D-1}{M} + 1\right]} \qquad (35) \\ a &= (1-\kappa)\left(\log(M) - \frac{1}{2}\log\left[\frac{D-1}{M} + 1\right]\right) \\ \sigma_\eta^2 &= (1-\kappa^2)\log\left[\frac{D-1}{M} + 1\right]. \end{aligned}$$

Note that if $C = 0$ we obtain $\kappa = 0$ from (35), but we need not have $D = 1$ as in the case of the SSOE model.

From equations (33) and (34), and using the fact that $\exp\{\kappa\sigma_X^2\} < \exp\{\sigma_X^2\}$ for $0 \leq \kappa < 1$, we can conclude that for $\kappa \geq 0$ ($\Rightarrow C \geq 0$), $C < \frac{D-1}{D}$ which, in turn, implies that $D > \frac{1}{1-C}$. It also follows from (35) that for fixed C, D can be arbitrarily close to $\frac{1}{1-C}$ by choosing κ sufficiently close to 1. Thus this lower bound for D is tight.

Appendix C: Markov Chain Filter for the DSOE Model

In what follows, we explicitly distinguish between the state probability vector at time t based on information set $\mathcal{I}_{t-1}$, $\boldsymbol{\pi}(t|\mathcal{I}_{t-1})$, and the revision of that state vector based on the observed value of y_t at time t, $\boldsymbol{\pi}(t|\mathcal{I}_t)$. At any given point in the parameter space, each component of (20) can be calculated via application of the following filter:

Step 1. Define a vector of N_x initial state probabilities, $\boldsymbol{\pi}(0) = (1/N_x, 1/N_x, \ldots, 1/N_x)'$

Step 2. Given $\boldsymbol{\pi}(0)$, use the inverse cumulative distribution function technique to determine the initial grid of N_x states, $x_{(1)}, x_{(2)}, ..., x_{(N_x)}$, from the steady state distribution of x_t, $N(\mu_X, \sigma_x^2)$, where $\mu_X = \frac{a}{1-\kappa}$ and $\sigma_X^2 = \frac{\sigma_\eta^2}{1-\kappa^2}$.

Step 3. Determine the elements of the transition probability matrix $\mathbf{P}$, as

$$p_{ij} = \frac{\varphi\left(x_{(j)} - a - \kappa x_{(i)}\right)}{\sum_{k=1}^{N_x} \varphi\left(x_{(k)} - a - \kappa x_{(i)}\right)} \quad \text{for } i = 1, 2, ..., N_x,\ j = 1, 2, ..., N_x.$$

Step 4. Update the state probability vector as $\boldsymbol{\pi}(1|\mathcal{I}_0) = \boldsymbol{\pi}(0)\,\mathbf{P}$

Step 5. Assume that the observed value of y_t at $t = 1$, is equal to the jth value in the grid of N_y possible values for y_t. Revise the ith component of $\boldsymbol{\pi}(1|\mathcal{I}_0)$ to accommodate the observed value $y_{(j)}$, as

$$\boldsymbol{\pi}(1|\mathcal{I}_1) = \frac{q_{ij}\boldsymbol{\pi}_i(1|\mathcal{I}_0)}{\sum_{k=1}^{N_x} q_{kj}\boldsymbol{\pi}_k(1|\mathcal{I}_0)}, \quad i = 1, 2, \ldots, N_x$$

where

$$q_{ij} = \frac{\exp\left[-\lambda_{(i)}\right]\lambda_{(i)}^{y_{(j)}}}{y_{(j)}!}$$

and $\lambda_{(i)} = \exp(x_{(i)})$, assuming an exponential link function in (3).

Step 6 Calculate the first component in the likelihood function in (20) as

$$P(y_1|\mathcal{I}_0, \boldsymbol{\Theta}) = \sum_{k=1}^{N_x} q_{kj}\boldsymbol{\pi}_k(1|\mathcal{I}_0)$$

Step 7 Update the state probability vector as $\boldsymbol{\pi}(2|\mathcal{I}_1) = \boldsymbol{\pi}(1|\mathcal{I}_1)\,\mathbf{P}$

Step 8 Calculate the second component in the likelihood function in (20) as

$$P(y_2|\mathcal{I}_1, \boldsymbol{\Theta}) = \sum_{k=1}^{N_x} q_{kj}\boldsymbol{\pi}_k(2|\mathcal{I}_1)$$

Step 9 Revise the *ith* component of $\boldsymbol{\pi}\left(2|\mathcal{I}_1\right)$ to accommodate the observed value of y_t at $t=2$, as

$$\boldsymbol{\pi}\left(2|\mathcal{I}_2\right)=\frac{q_{ij}\boldsymbol{\pi}_i\left(2|\mathcal{I}_1\right)}{\sum_{k=1}^{N_x}q_{kj}\boldsymbol{\pi}_k\left(2|\mathcal{I}_1\right)},\quad i=1,2,\ldots,N_x$$

Step 10 Repeat steps 7 to 9 for $t=3,4,\ldots,T$.

References

[1] Al-Osh M.A. and Alzaid, A.A. (1987) First-order Integer Valued Autoregressive (INAR(1)) Process, *Journal of Time Series Analysis,* 8, 261-275.

[2] Bauwens, L. and Veredas, D. (2004) The Stochastic Conditional Duration Model: A Latent Variable Model for the Analysis of Financial Durations, *Journal of Econometrics,* 199, 381-412.

[3] Brown, R.G. (1959), Statistical Forecasting for Inventory Control, McGraw-Hill, New York.

[4] Chan, K.S. and Ledolter, J. (1995) Monte Carlo EM Estimation for Time Series Models Involving Counts, *Journal of the Americal Statistical Association,* 90, 242-252.

[5] Cox, D.R. (1981) Statistical Analysis of Time Series: Some Recent Developments, *Scandanavian Journal of Statistics,* 8, 93-115.

[6] Davis, R.A., Dunsmuir, W.T.M. and Wang, Y. (1999) Modelling Time Series of Counts, in Ghosh, S. ed. *Asymptotics, Nonparametrics and Time Series, A Tribute to Madan-Puri,* 63-113, Marcel Dekker.

[7] Davis, R.A., Dunsmuir, W.T.M. and Wang, Y. (2000) On Autocorrelation in a Poisson Regression Model, *Biometrika,* 87, 491-505.

[8] Davis, R.A., Dunsmuir, W.T.M. and Streett, S. (2003) Observation-driven Models for Poisson Counts, *Biometrika,* 90, 777-790.

[9] Durbin, J. and Koopman, S.J. (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*, 62, 3-56.

[10] Durbin, J. and Koopman, S.J. (2001), *Time Series Analysis by State Space Methods*, Oxford University Press.

[11] Engle, R.F. (1982) Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica,* 50, 987-1008.

[12] Engle, R.F. and Russell, J.R. (1998) Autoregressive Conditional Duration: A New Approach for Irregularly Spaced Transaction Data, *Econometrica*, 66, 987-1007.

[13] Freeland, K.R. and McCabe, B.P.M. (2004a) Forecasting Discrete Valued Low Count Time Series, *International Journal of Forecasting,* 20 427-434.

[14] Freeland, K.R. and McCabe, B.P.M. (2004b) Analysis of Low Count Time Series by Poisson Autoregression, *Journal of Time Series Analysis,* 25, 701-722.

[15] Frühwirth-Schnatter, S. and Wagner, H. (2006) Auxiliary Mixture Sampling for Parameter Driven Models of Time Series of Counts with Applications to State Space Modelling, *Biometrika,* 93, 827-841.

[16] Grunwald, G.K., Hamza, K. and Hyndman, R.J. (1997) Some Properties and Generalizations of Non-Negative Bayesian Time Series Models, *Journal of the Royal Statistical Society,* B, 59, 615-626.

[17] Harvey, A.C. (1991) Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press, Cambridge.

[18] Harvey, A.C. and Fernandez, C. (1989) Time Series Models for Counts or Qualitative Observations, *Journal of Business & Economics Statistics,* 7, 407-417.

[19] Harvey, A.C. and Durbin, J. (1986) The Effect of Seat Belt Legislation on British Road Casualties: A Case Study in Structral Time Series Modelling, *Journal of the Royal Statistical Society,* A, 149, 187-227.

[20] Heinen, A. (2003) Modelling Time Series Count Data: an Autoregressive Conditional Poisson Model, *Draft Paper, University of California, San Diego.*

[21] Jung, R.C., Kukuk, M. and Leisenfeld, R. (2006) Time Series of Count Data: Modeling, Estimation and Diagnostics, *Computational Statistics and Data Analysis,* 51, 2350-2364.

[22] Kim, S,. Shephard, N,. and Chib, S. (1998) Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models, *Review of Economic Studies*, 65, 361-393.

[23] McCabe, B.P.M. and Martin, G.M. (2005) Bayesian Predictions of Low Count Time Series, *International Journal of Forecasting*, 21, 315-330.

[24] McCabe, B.P.M., Martin, G.M. and Freeland, K.R. (2006) Testing for Low-order Dependence in Non-Gaussian Time Series Data, *Draft Paper, Monash University.*

[25] McDonald and Zucchini (1997) *Hidden Markov and Other Models for Discrete-valued Time Series* (Chapman and Hall)

[26] McKenzie, E. (1988) Some ARMA Models for Dependent Sequences of Poison Counts, *Advances in Applied Probability,* 20, 822-835.

[27] Ord, J.K., Koehler, A.B. and Snyder, R.D. (1997) Estimation and Prediction of a Class of Dynamic Nonlinear Statistical Models, *Journal of the American Statistical Association*, 92, 1621-1629.

[28] Shephard, N. (1995) Generalized Linear Autoregressions, *Working Paper, Nuffield College, Oxford.*

[29] Strickland, C.M., C.S. Forbes and Martin, G.M. (2006) Bayesian Analysis of the Stochastic Conditional Duration Model, *Computational Statistics and Data Analysis (Special Issue on Statistical Signal Extraction and Filtering)*, 50, 2247-2267.

[30] West, M., Harrison, J. and Migon, H.S. (1985) Dynamic Generalized Models and Bayesian Forecasting (with Discussion), *Journal of the American Statistical Association,* 80, 73-97.

[31] West, M., and Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models*, second edition, Springer-Verlag, New York.

[32] White, S.I., Clements, A. and Hurn, S. (2004) Discretized Non-Linear Filtering for Dynamic Latent Variable Models: with Application to Stochastic Volatility, *Draft Paper, Queensland University of Technology.*

[33] Zeger, S.L. (1988) A Regression Model for Time Series of Counts, *Biometrika,* 75, 621-629.

[34] Zu, R. and Joe, H. (2006) Modelling Count Time Series with Markov Processes Based on Binomial Thinning, *Journal of Time Series Analysis,* 27, 725-738.