Estimation and Testing in Integer-Valued AR(1) Models*

Kurt Brännäs
Department of Economics, University of Umeå
S-901 87 Umeå, Sweden
email: kurt.brannas@econ.umu.se
fax: +46-90-772302
phone: +46-90-786 6101

Umeå Economic Studies 335, 1994 (revised)

Abstract

The paper studies the integer-valued autoregressive model of order one and suggests a specification for panel data. Test statistics against under- or overdispersion within a generalized Poisson model are obtained. Predictors and prediction error variances are given for univariate and multivariate models. The small sample performance of maximum likelihood and new generalized method of moments estimators and tests are evaluated and compared. An empirical illustration based on the number of firms in sectors of the Swedish forest industry 1970–1992 is included.

Key Words: Poisson, generalized Poisson, estimation, testing, panel data, prediction, forest sector.

*I am grateful to Runar Brännlund for providing the data for the illustration and to Abdulhamid A. Alzaid and Per Johansson for comments. The financial support from the Swedish Research Council for the Humanities and Social Sciences is acknowledged.

1. Introduction

In this paper we study the integer-valued autoregressive model of order one [AR(1) model]. The Poisson AR(1) model was first introduced by McKenzie (1985) and later discussed by McKenzie (1988), Al-Osh and Alzaid (1987) and others. This pure time series model closely resembles the Gaussian AR(1) model in terms of the first two moments. In other respects the model is quite different, however. The Poisson distribution has a single parameter, which is equal to both the mean and the variance. If empirically, the variance exceeds the theoretically expected variance we talk of overdispersion. Correspondingly we may also talk of underdispersion when the empirical variance is smaller than the theoretically expected one. This issue has important empirical implications for interpretation and for obtaining an appropriate covariance matrix, which in turn is important, for instance, for testing purposes.

Regression models for time series count data are becoming increasingly often applied. Different approaches were initially suggested by Zeger (1988), Zeger and Qagish (1986) and Smith (1979). Recent work of Brännäs and Johansson (1992, 1994) and others suggest that these approaches are empirically feasible. Only in the regression approach of Zeger and Qagish (1986) is there an explicit lag structure. However, this is quite different from the one considered here. In the other regression models the autocorrelation structure is assumed to be due to a latent process. A different time series model is the so called DARMA model of Jacobs and Lewis (1983). In the conclusion to this paper we will make some comments on how the present AR(1) framework can accommodate explanatory variables as well.

In particular, we obtain generalized method of moment (GMM, Hansen, 1982) estimators for both the AR(1) Poisson and the generalized Poisson (Alzaid and Al-Osh, 1993) models. The generalized Poisson model has the desirable property of allowing for both under- and overdispersion. Lagrange multiplier (LM) and Wald type test statistics in the GMM domain are given against under- and/or overdispersion, when the alternative hypothesis is an AR(1) generalized Poisson model. The obtained estimators and test statistics are evaluated and compared by finite sample Monte Carlo experimentation. In addition, the new GMM estimators are compared to maximum likelihood (ML) and conditional least squares (CLS) estimators in the Poisson case. The latter two estimators were for this case given and partly evaluated by Al-Osh and Alzaid (1987). Predictors and prediction error variances are derived. For multivariate dependent AR(1) models for panel data corresponding measures are derived for both the Poisson and generalized Poisson cases.

The details of the basic models are outlined in Section 2. In Section 3 we consider estimation, while testing against under- and overdispersion is considered in Section 4. The design and results of small sample Monte Carlo experimentation are reported in Section 5. Section 6 provides predictors and prediction error variances. Section 7 introduces the panel data AR(1) model. An empirical illustration based on the number of Swedish forestry plants according to production type, 1972–1990, is reported in Section 8. Some of these series manifest underdispersion. A few concluding remarks are made at the end of the paper.

2. Models

Consider a stationary $\{y_t\}$ process with Poisson marginal densities. Both for simplicity reasons and for reasons of future empirical work we restrict our interest to the autoregressive model of order one [the AR(1) model]. At an arbitrary time t we write

$$y_t = \alpha \circ y_{t-1} + \epsilon_t, \quad t = 2, \ldots, T. \tag{1}$$

Here, the notation (cf. Steutel and van Harn, 1979) $\alpha \circ y = \sum_{i=1}^{y} u_i$, where u_i is a sequence of independent and identically distributed (i.i.d.) 0-1 random variables independent of y with $\Pr(u_i = 1) = \alpha$. The $\alpha \circ y$ operator represents binomial thinning of y, such that each of the y individuals either 'survives' ($u_i = 1$) with equal probability α or 'dies' ($u_i = 0$) with probability $1 - \alpha$. Under these assumptions it holds that

$$\begin{aligned} E(\alpha \circ y) &= E_y[yE(u_i|y)] = \alpha E(y) \\ V(\alpha \circ y) &= V_y[E(\alpha \circ y|y)] + E_y[V(\alpha \circ y|y)] = \alpha^2 V(y) + \alpha(1-\alpha)E(y). \end{aligned}$$

The ϵ_t is i.i.d. Poisson with mean $\lambda > 0$ and independent of y_{t-1}. For $\alpha \in (0,1)$ and y_1 discrete self-decomposable the AR(1) process is stationary (e.g., McKenzie, 1988).

The properties of the model have been detailed upon by Al-Osh and Alzaid (1987) and for a slightly different parametrization by McKenzie (1988). In the present case both the mean and the variance of the $\{y_t\}$ process is $\lambda/(1-\alpha)$. It is easy to verify that the autocorrelation at lag k is α^k, which obviously is restricted to be positive.

Extensions to more general ARMA models are discussed by McKenzie (1988) and others. The model (1) remains of the same form for other distributions such as the binomial, negative binomial or generalized Poisson (Alzaid and Al-Osh, 1993, McKenzie, 1986). The negative binomial allows for overdispersion while the generalized Poisson model has the attractive property of allowing for both under- and overdispersion.

Therefore, we consider the generalized Poisson (GP) [see Consul (1989) for a detailed treatment of the univariate GP model and its estimation] AR(1) model specification of Alzaid and Al-Osh (1993). Instead of binomial thinning we now employ a quasibinomial thinning operator, $\alpha \circ y_{t-1}$. By this the probability to retain an element is increasing in y_{t-1} for $\theta > 0$ and decreasing for $\theta < 0$. When y_{t-1} is GP distributed with parameters λ' and θ [GP(λ', θ); see eq. (9), below, for the density function] it follows that $\alpha \circ y_{t-1}$ is distributed as GP($\alpha\lambda', \theta$). When in addition ϵ_t is distributed as GP($(1-\alpha)\lambda', \theta$) it follows that y_t has a GP (λ', θ) distribution. The GP model has moments; $E(y_t) = \lambda'/(1-\theta)$, $V(y_t) = \lambda'/(1-\theta)^3$ and the autocorrelation at lag k is α^k. Underdispersion (over-) results for $\theta < 0$ ($\theta > 0$), while the Poisson case arises for $\theta = 0$.

To make the GP compatible with the previous Poisson specification we let the ϵ_t process have parameters λ and θ, i.e. $\lambda = (1-\alpha)\lambda'$. For the chosen specification the mean of the process is $\lambda/[(1-\alpha)(1-\theta)] = \lambda'/(1-\theta)$, which for $\theta = 0$ reduces to that of the Poisson model.

For the negative binomial AR(1) model some of the distributional features are treated by McKenzie (1986) and Al-Osh and Alzaid (1993). For a model specification compatible with the Poisson model in (1) with $E(\epsilon_t) = \lambda$ and $V(\epsilon_t) = \lambda + \sigma^2\lambda^2$. We obtain $E(y_t) = \lambda/(1-\alpha)$, $V(y_t) = [\lambda + \lambda(\alpha + \sigma^2\lambda)]/(1-\alpha^2)$ with α^k the autocorrelation at lag k.

3. Estimation

The estimation of α and λ in the stationary Poisson AR(1) case by conditional least squares (CLS) and maximum likelihood (ML) estimators was studied by Al-Osh and Alzaid (1987); see also Jin-Guan and Yuan (1991) and Ronning and Jung (1992). In this section we also introduce the new generalized method of moment (GMM) estimator and extend this estimator to the GP AR(1) model.

3.1 Conditional Least Squares

The conditional mean of y_t given y_{t-1} is for the Poisson model given by

$$E(y_t|y_{t-1}) = \alpha y_{t-1} + \lambda = g(\boldsymbol{\psi}_1), \tag{2}$$

where $\boldsymbol{\psi}_1 = (\alpha, \lambda)'$ is the vector of unknown parameters to be estimated. The CLS estimator minimizes the criterion function

$$Q(\boldsymbol{\psi}_1) = \sum_{t=2}^{T} [y_t - g(\boldsymbol{\psi}_1)]^2, \tag{3}$$

which yields the estimators

$$\begin{aligned} \hat{\alpha} &= \frac{\sum_{t=2}^{T} y_t y_{t-1} - (\sum_{t=2}^{T} y_t \sum_{t=2}^{T} y_{t-1})/T}{\sum_{t=2}^{T} y_{t-1}^2 - (\sum_{t=2}^{T} y_{t-1})^2/T} \\ \hat{\lambda} &= (\sum_{t=2}^{T} y_t - \hat{\alpha} \sum_{t=2}^{T} y_{t-1})/T. \end{aligned}$$

In the GP model the conditional mean equals $\alpha y_{t-1} + \lambda/(1-\theta)$. This reveals that λ and θ can not be estimated separately, unless additional information is incorporated for the CLS estimation problem. For the negative binomial model the conditional mean is of the form (2), indicating that σ^2 can not be estimated from the conditional mean.

3.2 Maximum Likelihood

For a fixed initial value y_1 the conditional log-likelihood function may be written

$$\ell_1(\boldsymbol{\psi}_1) = \sum_{t=2}^{T} \log \Pr(y_t|y_{t-1}), \tag{4}$$

where for the Poisson model (cf. Johnson and Kotz, 1970, ch. 11 and Al-Osh and Alzaid, 1987)

$$\Pr(y_t|y_{t-1}) = \exp(-\lambda) \sum_{i=0}^{m} \frac{\lambda^{y_t - i}}{(y_t - i)!} \binom{y_{t-1}}{i} \alpha^i (1-\alpha)^{y_{t-1}-i}, \quad t = 2, \ldots, T \tag{5}$$

is the conditional density for given y_{t-1} and where $m = \min(y_{t-1}, y_t)$. While the conditional ML estimator maximizes (4), the exact ML estimator maximizes

$$\ell(\boldsymbol{\psi}_1) = \ell_1(\boldsymbol{\psi}_1) + \log \Pr(y_1), \tag{6}$$

where

$$\Pr(y_1) = \left[\frac{\lambda}{1-\alpha}\right]^{y_1} \frac{1}{y_1!} \exp\left[-\frac{\lambda}{1-\alpha}\right], \tag{7}$$

is the steady state density.

For the GP model, with $E(\epsilon_t) = \lambda'/(1-\theta)$, the appropriate expressions (Alzaid and Al-Osh, 1993) are given by

$$\begin{aligned} \Pr(y_t|y_{t-1}) &= \sum_{i=0}^{m} \binom{y_{t-1}}{i} \frac{\alpha\overline{\lambda}'}{\lambda' + y_{t-1}\theta} \left[\frac{\alpha\lambda' + i\theta}{\lambda' + y_{t-1}\theta}\right]^{i-1} \left[\frac{\overline{\lambda}' + (y_{t-1} - i)\theta}{\lambda' + y_{t-1}\theta}\right]^{y_{t-1}-i-1} \\ &\quad \times \overline{\lambda}'[\overline{\lambda}' + \theta(y_t - i)]^{y_t - i - 1} \exp[-\overline{\lambda}' - \lambda'\theta(y_t - i)]/(y_t - i)! \qquad (8) \\ \Pr(y_1) &= \lambda'(\lambda' + \theta y_1)^{y_1 - 1} \exp[-(\lambda' + \theta y_1)]/y_1!, \qquad (9) \end{aligned}$$

where $\overline{\lambda}' = (1-\alpha)\lambda'$ and the parameter vector is $\boldsymbol{\psi}_2 = (\lambda', \alpha, \theta)$.[1] With $\theta = 0$ expressions (8)-(9) reduce to those in (5) and (7), when in the latter λ' is replaced by $\overline{\lambda}'$.

To obtain ML estimates using (6) a numerical maximization algorithm has to be employed. The conditional density (8) of the GP model is quite complex as will the derivatives required for ML estimation be. Wald and Lagrange multiplier tests are therefore analytically and numerically difficult to obtain. The same will hold true for the negative binomial model.

3.3 Generalized Method of Moments

In this subsection we consider the use of the GMM estimator of Hansen (1982) for our AR(1) purposes. Two approaches to GMM estimation can be considered. First, we may employ unconditional moment restrictions based on the stationary distribution to form the estimator. The second approach is based on conditional moment restrictions (e.g., Newey, 1985, Tauchen, 1986) and can be seen as an extension to the CLS estimator. Initial experimentation using unconditional moment restrictions suggested that the small sample properties are not at all favourable. Therefore, we only consider the conditional GMM estimation problem.

To avoid non-uniqueness due to a surplus of possibly informative moment restrictions as well as to 'increase efficiency' the GMM estimator is useful, though it may be numerically more involved than method of moment estimators. The GMM estimator minimizes a quadratic form

$$q = \mathbf{m}(\boldsymbol{\psi})'\hat{\mathbf{W}}^{-1}\mathbf{m}(\boldsymbol{\psi}), \tag{10}$$

where $\mathbf{m}(\boldsymbol{\psi})$ is the vector of moment restrictions. Subject to mild regularity conditions (e.g., MacKinnon and Davidson, 1993, ch. 17) the estimator of $\boldsymbol{\psi}$ is consistent and asymptotically normal for any symmetric and positive definite matrix $\hat{\mathbf{W}}$, such as the identity matrix $\mathbf{I}$. The estimator is efficient when $\hat{\mathbf{W}}$ is the asymptotic covariance matrix of $\mathbf{m}(\boldsymbol{\psi})$. To obtain $\hat{\mathbf{W}}$, q can in a first step be minimized using say the identity matrix $\mathbf{I}$ for $\hat{\mathbf{W}}$. For a second step the consistent estimates $\hat{\boldsymbol{\psi}}$ from step one are used to form $\hat{\mathbf{W}}$.

[1] For $\theta < 0$ the GP density in (9) is in fact truncated. Since (9) is the marginal density for any t, the truncation can be expressed as $\Pr(y_t) = 0$ for $y_t > k$. For $\lambda' > 0$, θ is restricted by $\max(-1, -\lambda'/k) < \theta \leq 1$ for $k > 4$. Consul (1989, ch. 2) gives additional details and a discussion about the small effect of neglecting the truncation.

For past observations $y_1, \ldots, y_{t-1}$, the conditional mean of y_t in the Poisson AR(1) model is given in (2) and the one-step-ahead prediction error $e_t = y_t - \alpha y_{t-1} - \lambda$ is the important part of (3), see also Section 6, below. The normal equations of (3) are empirical moment restrictions; $T^{-1}\sum_{t=2}^{T} e_t = 0$ for λ and $T^{-1}\sum_{t=2}^{T} y_{t-1}e_t = 0$ for α, with corresponding theoretical unconditional moments $E(e_t) = 0$ and $E(y_{t-1}e_t) = 0$. These unconditional moments are equal to conditional ones, since for any function $h(\cdot)$, $E(h(y_{t-1})e_t) = E_y[E(h(y_{t-1})e_t|y_{t-1}] = E_y h(y_{t-1})E(e_t|y_{t-1}) = E_y h(y_{t-1}) \cdot 0 = 0$. The CLS estimator can therefore be interpreted as a condtional GMM estimator. In addition, the numbers of unknowns and restrictions are equal, and $\hat{\mathbf{W}}$ is equal to the identity matrix of order two, $\mathbf{I}_2$.

Additional moment restrictions are available. For instance, it can be proved that both the conditional and unconditional expectations of $e_t e_{t-1}$ are equal to zero. From Alzaid and Al-Osh (1988) the general result for stationary integer-valued processes of order one $V(y_t|y_{t-1}) = E[(y_t - E(y_t|y_{t-1}))^2|y_{t-1}] = \alpha(1-\alpha)y_{t-1} + V(\epsilon_t)$ can be seen to be equal to $E(e_t^2|y_{t-1})$. This restriction proves to be most useful, in particular, for the GP model.[2] These and possibly other conditional moment restrictions can be used for GMM estimation.

The asymptotic covariance matrix $\hat{\mathbf{W}}$ can, in principle, be formed either from initial consistent estimators of the parameters and an analytical expression or from these estimators and the sample covariance matrix of the restrictions. The latter is based on the widely used consistent Newey and West (1987a) estimator

$$\hat{\mathbf{W}} = \hat{\mathbf{\Gamma}}_0 + \sum_{j=1}^{p}(1 - \frac{j}{p+1})[\hat{\mathbf{\Gamma}}_j + \hat{\mathbf{\Gamma}}_j'], \tag{11}$$

where

$$\hat{\mathbf{\Gamma}}_j = T^{-1}\sum_{t=j+1}^{T} \mathbf{m}_t(\hat{\boldsymbol{\psi}})'\mathbf{m}_{t-j}(\hat{\boldsymbol{\psi}}), \quad j = 0, 1, \ldots, p,$$

and p has to be given. The p corresponds to the order of an MA(p) process. Therefore, for the present AR(1) process we expect that p should be chosen large for large α values.

The estimated asymptotic covariance matrix of the GMM estimator based on $\hat{\mathbf{W}}$ is

$$Cov(\hat{\boldsymbol{\psi}}) = \frac{1}{T}[\hat{\mathbf{G}}'\hat{\mathbf{W}}^{-1}\mathbf{G}]^{-1}, \tag{12}$$

where the $\hat{\mathbf{G}}$ matrix has rows $\mathbf{G}^j = \partial \mathbf{m}_j(\boldsymbol{\psi})_j/\partial\boldsymbol{\psi}'$ evaluated at $\hat{\boldsymbol{\psi}}$.

As part of the Monte Carlo experiments reported below, we have an interest in studying the impact of the number of restrictions as well as the impact of $\hat{\mathbf{W}}$ versus $\mathbf{I}$. Even if $\hat{\mathbf{W}} = \mathbf{I}$ in (10) is enforced we need to have the asymptotic covariance matrix $\hat{\mathbf{W}}$ of the moment restriction vector to obtain the appropriate covariance matrix of the parameter estimates.

4. Specification Testing

In this section we consider specification testing against under- and overdispersion when under the null hypothesis we have an AR(1) Poisson model and under the alternative the generalized Poisson model. The test statistics are based on the GMM approach. The

[2] The dependence on y_{t-1} makes this model different from the Gaussian autoregressions with their constant variances.

Lagrange Multiplier (LM) test based on likelihood theory of $\mathrm{H}_0 : \theta = 0$ vs. $\mathrm{H}_A : \theta \neq 0$ is too complex analytically as well as numerically to be of empirical interest.

To test against under- or overdispersion using the GMM approach we may apply analogous techniques as used within likelihood theory (Newey and West, 1987b). A Lagrange multiplier (LM) or score statistic is analytically and numerically simple when under H_A the numbers of restrictions and unknown parameters are equal (e.g., Davidson and MacKinnnon, 1993, ch. 17). The LM statistic is then of the form

$$\mathrm{LM} = T\ \mathbf{m}'(\tilde{\boldsymbol{\psi}}_0)[\mathbf{W}(\tilde{\boldsymbol{\psi}}_0)]^{-1}\mathbf{m}(\tilde{\boldsymbol{\psi}}_0), \tag{13}$$

where $\tilde{\boldsymbol{\psi}}_0 = (\tilde{\alpha}, \tilde{\lambda}, \theta = 0)'$ is a GMM estimator of $\boldsymbol{\psi}_0$ under H_0 and $\mathbf{W}$ is estimated using these estimates. Asymptotically, LM is distributed as a $\chi^2(1)$ variate.

For a model estimated under H_A, i.e., under the unrestricted GP AR(1) model, the application of the Wald statistic is straightforward. We simply have to compare $\hat{\theta}^2/V(\hat{\theta})$, where $V(\hat{\theta})$ is the appropriate diagonal element in (12), to a $\chi^2(1)$ distribution.

5. Small Sample Performance

The CLS, exact ML and GMM estimators are compared for the Poisson AR(1) model in a small Monte Carlo experiment replicating and extending the study of Al-Osh and Alzaid (1987). To generate the data we assume $\alpha = 0.1, 0.3, 0.5, 0.7$ and 0.9 with $\lambda = 1$ and 5. Due to the specification of the model we then obtain the means of y_t as $10\lambda/9, 10\lambda/7, 2\lambda, 10\lambda/3, 10\lambda$ for $\alpha = 0.1, \ldots, 0.9$. To obtain a stationary series we set y_1 equal to the stationary mean $\lambda/(1-\alpha)$ and discard the first 150 generated observations. Two sample sizes, $T = 50$ and 200, are employed, and the number of replications is 1000 at each design point. For the exact ML estimator we employ reparametrizations such that $\lambda = \exp(\zeta)$ and $\alpha = 1/(1 + \exp(\eta))$ and maximize with respect to the unrestricted $\zeta = \log(\lambda)$ and $\eta = \log(\alpha^{-1} - 1)$. The GMM estimator is evaluated for the GP AR(1) model in the case of $\theta = 0$, i.e., the data is generated from a Poisson AR(1) model. The sizes of the LM and Wald statistics of $\mathrm{H}_0 : \theta = 0$ vs. $\mathrm{H}_A : \theta \neq 0$ are compared and evaluated.

The bias and MSE results for the estimators in the Poisson case are given in Table 1. Both the biases and MSEs of the CLS and ML estimators of α are small. The measures are smaller for larger sample sizes, and as expected, the exact ML estimator has smaller MSE than CLS. The bias of the CLS estimator of α is larger for $\lambda = 5$. The unrestricted CLS estimator yields a substantial fraction of $\hat{\alpha}$ estimates outside of the $[0, 1]$ interval for small and large values on α. As expected there is a high negative correlation between the estimators of α and λ for large α. On comparison with the results of Al-Osh and Alzaid (1987) the present study essentially confirms their conclusions. The MSE of the CLS estimator of λ increases with α and is larger for $\lambda = 5$, when the variance of the process is larger. The increases are much smaller for the ML estimator. With respect to λ, however, it is found that the performance of the CLS estimator varies with λ or equivalently with the mean or the variance of the $\{y_t\}$ process.

The bias and MSE results for the Poisson GMM estimator are based on $\mathbf{W} = \mathbf{I}_4$ as well as $\hat{\mathbf{W}}_4$. (The notation $\hat{\mathbf{W}}_p$ is used to indicate the number of terms p of the Newey and West (1987a) estimator in (11)). The four conditional moment restrictions are the two normal equations of the CLS estimator (the GMM reduces to the CLS estimator when $\mathbf{W} = \mathbf{I}_2$), $E(e_t^2|y_{t-1}) - \alpha(1-\alpha)y_{t-1} - \lambda = 0$ and $E(e_t e_{t-1}|y_{t-1}) = 0$. There is no gain of using $\hat{\mathbf{W}}_4$ for the smallest sample size $T = 50$. It is for the larger α-values that the GMM estimator has smaller MSE than CLS. For λ we again note that the GMM estimator based on $\mathbf{I}_4$ does, at least, equally well as the one based on $\hat{\mathbf{W}}_4$. GMM has substantially smaller MSEs for the largest α values.

In some additional experiments performed to study the robustness of the GMM estimator with respect to the choice of moment restrictions the conditional and unconditional prediction error variance restrictions were compared. The reported GMM estimator has mostly much smaller MSEs, but for $T = 50$ and α small the one based on the unconditional variance is better in the MSE sense.

In summary, the ML estimator has the smallest MSEs and for short time series there is a gain of using the GMM estimator as compared to CLS only when α is relatively large.

For the GP AR(1) model the conditional moment restriction due to the conditional variance proved crucial for getting around the problem with nonuniqueness arising from the ratio $\lambda/(1-\theta)$ in the conditional mean. Table 2 presents results for the case of $\mathbf{W} = \mathbf{I}_4$, $\hat{\mathbf{W}}_2$ and $\hat{\mathbf{W}}_4$ and $T = 50$ and 200, for the four conditional moment restrictions $T^{-1}\sum_{t=2}^{T} e_t = T^{-1}\sum_{t=2}^{T}[y_t - \alpha y_{t-1} - \lambda/(1-\theta)] = 0$, $T^{-1}\sum_{t=2}^{T} e_t y_{t-1} = 0$, $T^{-1}\sum_{t=2}^{T} e_t^2 - \alpha(1-\alpha)y_{t-1} - \lambda/(1-\theta)^3 = 0$ and $T^{-1}\sum_{t=2}^{T} e_t e_{t-1} = 0$.

On comparison, there appears to be little reason for using $\hat{\mathbf{W}}$ instead of $\mathbf{W} = \mathbf{I}_4$, whether $\hat{\mathbf{W}}$ is based on $p = 4$ in (11) or $p = 2$. A general conclusion seems to be that for large α values (or large $E(y_t)$) biases and MSEs of, in particular, λ are large in an absolute sense.

In Table 3 sizes are reported for the LM and Wald test statistics of $\mathrm{H}_0 : \theta = 0$ against a two-sided alternative for a nominal size of 0.05. For small T the obtained sizes are too large. Size properties improve as T increases. For the Wald statistic there is no clearcut indication of whether $p = 2$ or 4 is to be preferred.

6. Prediction

Consider the prediction of a future value y_{T+h} given that we have observed the series up through time T, i.e., $y_1, \ldots, y_T$ is observed. By repeated substitution we may write the future values of the process

$$y_{T+h} = \alpha^h \circ y_T + \sum_{i=1}^{h} \alpha^{h-i} \circ \epsilon_{T+i}, \qquad h = 1, 2, \ldots,$$

which should be interpreted as an equility in distribution.

From this we obtain the h-step ahead predictor for the Poisson model as

$$\begin{aligned} \hat{y}_{T+h|T} &= E(y_{T+h}|y_1, \ldots, y_T) = \alpha^h y_T + \lambda(1 + \alpha + \ldots + \alpha^{h-1}) \\ &= \alpha^h \left[y_T - \frac{\lambda}{1-\alpha} \right] + \frac{\lambda}{1-\alpha}, \end{aligned} \tag{14}$$

where in the final step the equality $(1 + \alpha + \ldots + \alpha^{h-1}) = (1-\alpha^h)/(1-\alpha)$ has been used. The term in brackets measures the deviation of the process from the mean at time T. As h goes to infinity and for $\alpha < 1$, the first part goes to zero and we hence find that the predictor approaches the mean of the process. As $\alpha \to 1$ the predictor approaches y_T, which is to be expected on comparison with a random walk model.

The prediction error is $e_{T+h} = y_{T+h} - \hat{y}_{T+h|T}$. It follows that $E(e_{T+h}) = 0$, for any $h > 0$, i.e., the predictor is unbiased. The prediction error variance at lead $h = 1$ is

$$\begin{aligned} V(e_{T+1}) &= E\left[(\alpha \circ y_T - \alpha y_T)^2\right] + E\left[(\epsilon_{T+1} - \lambda)^2\right] = \frac{\alpha(1-\alpha)\lambda}{1-\alpha} + \lambda \\ &= \frac{\lambda}{1-\alpha}\left[1-\alpha^2\right], \end{aligned} \tag{15}$$

where the results summarized in Section 2 are used to obtain the first part of the third expression. Proceeding in a similar way it can be shown that the prediction error variance at an arbitrary lead $h > 0$ can be written on the related form

$$V(e_{T+h}) = \frac{\lambda}{1-\alpha}\left[1-\alpha^{2h}\right]. \tag{16}$$

Note that the prediction error variance increases with the length of the forecast horizon h. As $h \to \infty$, the variance approaches an upper limit of $\lambda/(1-\alpha)$, which is the variance of the process.

The expressions for the GP AR(1) model with $E(\epsilon_t) = \lambda/(1-\theta)$ and $V(\epsilon_t) = \lambda/(1-\theta)^3$ are given by

$$\begin{aligned} \hat{y}_{T+h|T} &= \alpha^h\left[y_T - \frac{\lambda}{(1-\theta)(1-\alpha)}\right] + \frac{\lambda}{(1-\theta)(1-\alpha)} \\ V(e_{T+h}) &= \frac{\alpha\lambda}{(1-\theta)(1-\alpha)}(1-\alpha+\alpha^2-\alpha^3+\ldots-\alpha^{2h-1}) \\ &\quad + \frac{\lambda}{(1-\theta)^3}(1+\alpha^2+\ldots+\alpha^{2(h-1)}) \\ &= \frac{\alpha\lambda}{(1-\theta)(1-\alpha)}\frac{1-\alpha^{2h}}{1+\alpha} + \frac{\lambda}{(1-\theta)^3}\frac{1-\alpha^{2h-1}}{1-\alpha}. \end{aligned}$$

7. Panel Data

Consider a panel of M cross-section units and T time periods. We may write an M-variate AR(1) model for integer-valued data as

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix}_t = \begin{pmatrix} \alpha_1 & 0 & \ldots & 0 \\ 0 & \alpha_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \ldots & \alpha_M \end{pmatrix} \circ \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix}_{t-1} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_M \end{pmatrix}_t, \qquad t = 2, \ldots, T$$

or as

$$\mathbf{y}_t = \mathbf{A} \circ \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t. \tag{17}$$

The parameters are different for different cross-section units and we let the dependence between units be generated by

$$\boldsymbol{\epsilon}_t = \begin{pmatrix} \epsilon_1^* \\ \epsilon_2^* \\ \vdots \\ \epsilon_M^* \end{pmatrix}_t + \xi_t \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \boldsymbol{\epsilon}_t^* + \xi_t \mathbf{1}.$$

In the Poisson case the $\epsilon_{it}^*,\ i=1,\ldots,M,\ t=1,\ldots,T$, are independent Poisson variables with parameters $\lambda_i,\ i=1,\ldots,M$, and the scalar $\xi_t,\ t=1,\ldots,T$, is independently Poisson distributed with parameter δ (e.g., Johnson and Kotz, 1969, ch. 11). Let $\mathbf{\Gamma}_\epsilon(s) = E[\boldsymbol{\epsilon}_t - E(\boldsymbol{\epsilon}_t)][\boldsymbol{\epsilon}_{t+s} - E(\boldsymbol{\epsilon}_{t+s})]'$ be the autocovariance matrix at lag s. The assumptions imply $\mathbf{\Gamma}_\epsilon(s)=0,\ s\geq 1$, and

$$\mathbf{\Gamma}_\epsilon(0) = \begin{pmatrix} \lambda_1+\delta & \delta & \ldots & \delta \\ \delta & \lambda_2+\delta & \ldots & \delta \\ \vdots & \vdots & \ddots & \vdots \\ \delta & \delta & \ldots & \lambda_M+\delta \end{pmatrix}.$$

In addition to previous assumptions we add that $\mathbf{y}_{t-1}$ and ξ_t are independent. In the binomial thinning operators of

$$\mathbf{A}\circ\mathbf{y}_{t-1} = \left(\alpha_1\circ y_1 = \sum_{i=1}^{y_1} u_{i1},\ldots,\alpha_M\circ y_M = \sum_{i=1}^{y_M} u_{iM}\right)_{t-1}',$$

the u_{ij} are assumed to be i.i.d. $0-1$ random variables and independent of $\mathbf{y}_{t-1}$ and $\boldsymbol{\epsilon}_t,\ t=1,\ldots,T$.

The mean of the M-variate Poisson AR(1) process (17) is

$$E\left(\mathbf{y}_t\right) = \left(\mathbf{I}-\mathbf{A}\right)^{-1}\left(\boldsymbol{\lambda}+\delta\mathbf{1}\right),$$

where $\boldsymbol{\lambda}' = (\lambda_1,\ldots,\lambda_M)$. After some algebraic manipulation we obtain the covariance matrix of (17) as

$$Cov(\mathbf{y}_t) = \begin{pmatrix} (\lambda_1+\delta)/(1-\alpha_1) & \delta/\left(1-\alpha_1\alpha_2\right) & \ldots & \delta/\left(1-\alpha_1\alpha_M\right) \\ \delta/\left(1-\alpha_1\alpha_2\right) & (\lambda_2+\delta)/(1-\alpha_2) & \ldots & \delta/(1-\alpha_2\alpha_M) \\ \vdots & \vdots & \ddots & \vdots \\ \delta/\left(1-\alpha_1\alpha_M\right) & \delta/(1-\alpha_2\alpha_M) & \ldots & \left(\lambda_M+\delta\right)/\left(1-\alpha_M\right) \end{pmatrix}.$$

Obviously, it is possible to conceive of other and potentially more useful parametrizations for particular cases.

The h-steps ahead predictor of (17) takes the form

$$\hat{\mathbf{y}}_{T+h|T} = \mathbf{A}^h\mathbf{y}_T + \left(\mathbf{I}+\mathbf{A}+\ldots+\mathbf{A}^{h-1}\right)\left(\boldsymbol{\lambda}+\delta\mathbf{1}\right)$$

and the M-variate one-step-ahead prediction error is $\mathbf{e}_t = \mathbf{y}_t - \mathbf{A}\mathbf{y}_{t-1} - \boldsymbol{\lambda} - \delta\mathbf{1}$. The one-step ahead prediction error variance is

$$V(\mathbf{e}_{T+1}) = \begin{pmatrix} (\lambda_1+\delta)(1+\alpha_1) & \delta & \ldots & \delta \\ \delta & (\lambda_2+\delta)(1+\alpha_2) & \ldots & \delta \\ \vdots & \vdots & \ddots & \vdots \\ \delta & \delta & \ldots & (\lambda_M+\delta)(1+\alpha_M) \end{pmatrix},$$

which has a variance that is inflated by δ in comparison with the univariate case, cf. (15). The δ is also the covariance between one-step ahead prediction errors. For the general h-steps ahead prediction error covariance matrix we obtain, for $h>1$,

$$V(\mathbf{e}_{T+h}) = \left[\mathbf{A}^h(\mathbf{I}-\mathbf{A}^h)\right]E(\mathbf{y}_T) + \sum_{i=1}^{h}\left[\mathbf{A}^{h-i}\mathbf{\Lambda} + \delta\mathbf{A}^{h-i}\mathbf{1}\mathbf{1}'\mathbf{A}^{h-i}\right],$$

where $\mathbf{\Lambda}$ =diag($\boldsymbol{\lambda}$).

For this model type the conditional least squares (CLS) estimator is directly applicable, but $\lambda_i, i = 1, \ldots, M$, and δ cannot be separately estimated by this estimator. A CLS estimator $\hat{\boldsymbol{\psi}}' = (\text{diag } (\hat{\mathbf{A}})', (\hat{\boldsymbol{\lambda}} + \hat{\delta}\mathbf{1})')$ is obtained as

$$\hat{\boldsymbol{\psi}} = \arg\min_{\psi} \sum_{t=2}^{T} \mathbf{e}_t' \mathbf{e}_t,$$

while to estimate all parameters separately, for instance, $E(\mathbf{e}_t\mathbf{e}_t'|\mathbf{y}_{t-1})$ can be used to obtain a feasible conditional GMM estimator.

8. Empirical Illustration

To illustrate, we consider four short, annual industry series in the Swedish paper and pulp industry, 1972 – 1990. The series, Mechanical, Sulphate, Sulphite and Paper, represent the number of plants and are exhibited in Figure 1. For all series there appears to be slight negative trends. In terms of the Poisson AR(1) model this graphical evidence suggests that either there is a small α and a large λ (young plants) or a larger α but a small λ (old plants). The latter interpretation is the more reasonable one in view of the capital intensity of this industry. Also, there appears to be little variation in most series suggesting under- rather than overdispersion.[3] In view of this, the generalized Poisson model appears an interesting alternative. We anticipate negative estimates of θ for, at least, the Mechanical and Sulphate series. Note that $\theta < 0$ indicates that the probability of exit increases with the stock of plants y_{t-1}.

In model terms the number of entering firms in a given time period can be seen as arising from an unobservable and random number of potential entrants x_t, who decide to enter or not to enter. We may write $\epsilon_t = \sum_{i=1}^{x_t} z_{it}$. Here, $z_{it} = 1$ if the ith potential entrant chooses to enter and 0 otherwise. Let $\Pr(z_{it} = 1) = \delta$. Under an i.i.d. assumption on potential entrants and a Poisson (parameter η) assumption on x_t it holds that ϵ_t is Poisson distributed with parameter $\lambda = \eta\delta$. If at time $t-1$ there are no firms, the number of firms at time t will be ϵ_t. In the period from t to $t+1$ another ϵ_{t+1} firms will enter. On the other hand, the number of firms at time t, $y_t = \epsilon_t$, will be reduced by the number of exiting firms. If the probability of exiting is constant across firms and time periods, $1 - \alpha$, the number of remaining firms is denoted $\alpha \circ y_t = \sum_{i=1}^{y_t} u_{it}$. The u_{it} are again i.i.d. 0-1 random variables with $\Pr(u_{it} = 1) = \alpha$. This operation represents binomial thinning and yields a Poisson variable with mean $\alpha\mu$, where μ is the mean of the stationary $\{y_t\}$ process.

In Table 4 estimates of CLS, ML and GMM estimates of α and λ are given for both the Poisson and generalized Poisson models. The ML estimator is restricted to not fall outside the permissible parameter space. No parameter restrictions are employed for the CLS and GMM estimators. The $\hat{\lambda}$ is negative in two instances for CLS and in one instance for the GMM for the GP model. The LM and Wald statistics of $\theta = 0$ versus $\theta \neq 0$ in the GP AR(1) model do not reject the null hypothesis for any of the series. In terms of $\hat{\theta}$ it is negative and large only for the Mechanical production series.

[3] For Mechanical (mean = 22.5, standard deviation = 2.3), Sulphate (28.4, 3.5), Sulphite (17.2, 10.3) and Paper (58.9, 6.3).

Treating the series as a panel with different α_i and λ_i yields $\hat{\delta} = 0.148$ when parameters are estimated by CLS.

9. Concluding Remarks

It is a main attraction of the present model type that it can be extended in several ways. In other studies of firm entry or exits these are studied separately. The probability of exit $1 - \alpha$ can be made dependent on explanatory variables, using say, a logistic distribution function. Estimation of such parametric forms can then give an estimated time profile of past exits, when such are hard to observe directly. Obviously, the ex ante predictive performance may improve as well. In an analogous way the λ parameter characterizing entries may also be made, say, an exponential function of explanatory variables. Initial attempts in these directions for the above short time series suggested that these are feasible approaches.

In view of the results from the Monte Carlo experiments the series lengths in the empirical illustration seem short for strong inferences of parameters, etc. In a future study paralelling the one of Mayer and Chappell (1992) longer series will be used. Another alternative of gaining degrees of freedom is to use panel data. In this case a restrictive parametrization can often be estimated even with very short time series.

While the model type has been illustrated by firm stock data, its applicability is obviously of wider interest for the empirically oriented economist. Other fields that share features with the firm stock setting include population size with migratory movements and number of unemployees served by a small labour exchange agency. The strength of the model comes out when the observed integer-valued counts are small. For large counts one may expect conventional time series models to do equally well.

References

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.

Al-Osh, M.A. and Alzaid, A.A. (1993). Some Gamma Processes Based on the Dirichlet-Gamma Transformation. *Communications in Statistics: Stochastic Models* **9**, 123-143.

Alzaid, A.A. and Al-Osh, M.A. (1988). First-Order Integer-Valued Autoregressive (INAR (1)) Process: Distributional and Regression Properties. *Statistica Neerlandica* **42**, 53-61.

Alzaid, A.A. and Al-Osh, M.A. (1993). Some Autoregressive Moving Average Processes with Generalized Poisson Marginal Distributions. *Annals of the Institute of Mathematical Statistics* **45**, 223-232.

Brännäs, K. and Johansson, P. (1992). Estimation and Testing in a Model for Serially Correlated Count Data. In von Collani, E. and Göb, R. (eds.) *Proceedings of the Second Würzburg — Umeå Conference in Statistics.* University of Würzburg.

Brännäs, K. and Johansson, P. (1994). Time Series Count Data Regression. *Communications in Statistics: Theory and Methods* **23**, 2907-2925.

Consul, P.C. (1989). *Generalized Poisson Distribution: Properties and Applications* (Marcel Dekker, New York).

Davidson, R. and MacKinnon, J.G. (1993). *Estimation and Inference in Econometrics* (Oxford University Press, Oxford).

Hansen, L.P. (1982). Large Sample Properties of Generalized Method of Moment Estimators. *Econometrica* **50**, 1029-1054.

Jacobs, P.A. and Lewis, P.A.W. (1983). Stationary Discrete Autoregressive-Moving Average Time Series Generated by Mixtures. *Journal of Time Series Analysis* **4**, 19-36.

Jin-Guan, D. and Yuan, L. (1991). The Integer-Valued Autoregressive (INAR(p)) Model. *Journal of Time Series Analysis* **12**, 129-142.

Johnson, N.L. and Kotz, S. (1969). *Discrete Distributions* (Houghton Mifflin, Boston).

Mayer, W.J. and Chappell, W.F. (1992). Determinants of Entry and Exit: An Application of the Compounded Bivariate Poisson Distribution to U.S. Industries, 1972-1977. *Southern Economic Journal* **58**, 770-778.

McKenzie, E. (1985). Some Simple Models for Discrete Variate Time Series. *Water Resources Bulletin* **21**, 645-650.

McKenzie, E. (1986). Autoregressive Moving-Average Processes with Negative-Binomial and Geometric Marginal Distributions. *Advances in Applied Probability* **18**, 679-705.

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

Newey, W.K. (1985). Generalized Method of Moment Specification Testing. *Journal of Econometrics* **29**, 229-256.

Newey, W.K. and West, K.D. (1987a). A Simple, Positive Definite, Heteroskedasticity and Auto Correlation Consistent Covariance Matrix. *Econometrica* **55**, 703-708.

Newey, W.K. and West, K.D. (1987b). Hypothesis Testing with Efficient Method of Moments Estimation. *International Economic Review* **28**, 777-787.

Ronning, G. and Jung, R.C. (1992). Estimation of a First Order Autoregressive Process with Poisson Marginals for Count Data. In Fahrmeier, L. (ed.) *Advances in GLIM and Statistical Modelling* (Springer-Verlag, Berlin).

Smith, J.Q. (1979). A Generalization of the Bayesian Forecasting Model. *Journal of Royal Statistical Society* **B41**, 375-387.

Steutel, F.W. and van Harn, K. (1979). Discrete Analogues of Self-Decomposability and Stability. *The Annals of Probability* **7**, 893-899.

Tauchen, G.E. (1986). Statistical Properties of Generalized Method of Moments Estimators of Structural Parameters Obtained from Financial Data (with discussion). *Journal of Business and Economic Statistics* **4**, 397-424.

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

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

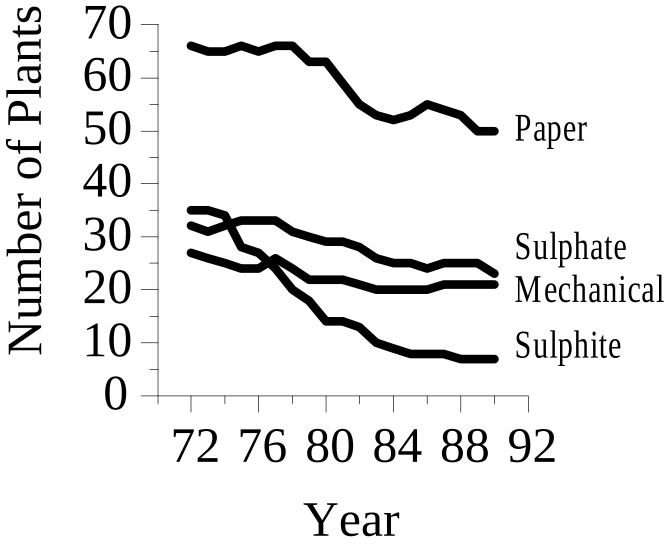

Figure 1:

Figure 1: Number of Mechanical, Sulphate, Sulphite and Paper plants in Sweden, 1972–1990.

Table 1: Biases and MSEs ($\times 10$) of ML, CLS and GMM (GMM_I for I_4 and GMM_W for $\hat{W}_4$) estimators, $T = 50$ and 200.

	$\lambda = 1$				$\lambda = 5$			
α	ML	CLS	GMM_I	GMM_W	ML	CLS	GMM_I	GMM_W
				Bias($\hat{\alpha}$), $T = 50$				
0.1	0.126	0.005	0.152	-0.291	0.074	0.704	0.251	-0.650
0.3	-0.277	-0.236	-0.057	-0.412	-0.147	0.630	0.224	-0.268
0.5	-0.210	-0.309	-0.028	-0.182	-0.187	0.478	0.107	-0.060
0.7	-0.134	-0.392	-0.049	-0.010	-0.068	0.493	0.068	0.144
0.9	-0.039	-0.283	-0.010	0.050	-0.016	0.388	0.014	0.071
				Bias($\hat{\alpha}$), $T = 200$				
0.1	-0.016	-0.025	0.012	-0.269	-0.022	0.164	0.095	-0.382
0.3	-0.025	-0.033	0.066	-0.183	-0.063	0.127	0.071	-0.214
0.5	-0.072	-0.072	0.004	-0.096	-0.067	0.111	0.013	-0.078
0.7	-0.047	-0.124	-0.022	-0.008	-0.015	0.109	0.012	0.034
0.9	-0.009	-0.100	-0.002	0.021	-0.008	0.092	0.010	0.012
				MSE($\hat{\alpha}$), $T = 50$				
0.1	0.139	0.190	0.260	0.281	0.133	0.201	0.345	0.351
0.3	0.186	0.191	0.188	0.316	0.179	0.179	0.209	0.333
0.5	0.121	0.161	0.111	0.172	0.110	0.117	0.107	0.182
0.7	0.051	0.123	0.050	0.065	0.042	0.067	0.042	0.057
0.9	0.006	0.037	0.006	0.007	0.004	0.019	0.005	0.006
				MSE($\hat{\alpha}$), $T = 200$				
0.1	0.045	0.050	0.057	0.066	0.041	0.048	0.067	0.078
0.3	0.043	0.052	0.046	0.065	0.041	0.046	0.049	0.072
0.5	0.031	0.044	0.029	0.041	0.025	0.036	0.030	0.031
0.7	0.012	0.029	0.013	0.014	0.010	0.024	0.011	0.011
0.9	0.001	0.011	0.001	0.001	0.001	0.005	0.001	0.001
				Bias($\hat{\lambda}$), $T = 50$				
0.1	-0.153	-0.222	-0.359	0.006	-0.362	-4.826	-1.710	3.552
0.3	0.350	0.068	-0.149	0.179	1.025	-5.732	-2.037	1.520
0.5	0.365	0.369	-0.236	-0.189	1.910	-5.848	-1.574	-0.011
0.7	0.248	0.885	-0.317	-0.472	1.056	-9.192	-1.525	-2.787
0.9	0.163	2.242	-0.401	-0.827	0.788	-20.115	-1.225	-3.907
				Bias($\hat{\lambda}$), $T = 200$				
0.1	0.004	-0.016	-0.071	0.185	0.138	-1.139	-0.728	2.043
0.3	0.051	-0.041	-0.199	0.084	0.352	-1.330	-0.786	1.330
0.5	0.094	0.059	-0.070	-0.029	0.589	-1.428	-0.251	0.556
0.7	0.111	0.309	-0.063	-0.134	0.238	-2.007	-0.327	-0.649
0.9	0.072	0.918	-0.078	-0.269	0.339	-4.769	-0.417	-0.876
				MSE($\hat{\lambda}$), $T = 50$				
0.1	0.366	0.418	0.634	0.621	4.934	7.398	12.917	11.853
0.3	0.552	0.509	0.674	0.844	10.122	10.303	12.590	18.756
0.5	0.567	0.690	0.673	0.790	11.543	12.096	12.069	18.759
0.7	0.564	1.245	0.661	0.755	11.950	19.387	13.084	16.682
0.9	0.545	3.036	0.601	0.744	11.086	47.918	11.433	14.398
				MSE($\hat{\lambda}$), $T = 200$				
0.1	0.097	0.100	0.154	0.126	1.529	1.723	2.592	2.581
0.3	0.122	0.141	0.163	0.171	2.273	2.553	2.886	3.771
0.5	0.143	0.196	0.182	0.182	2.687	3.691	3.313	3.266
0.7	0.142	0.320	0.182	0.175	2.972	6.833	3.270	3.175
0.9	0.139	1.078	0.162	0.164	2.757	13.578	2.787	3.067

Table 2: Biases and MSEs ($\times 10$) of GMM estimators of GP model with $W = I_4$ (top), $\hat{W}_4$ (middle) and $\hat{W}_2$, $T = 50, 200$ (bottom). The left part of the table corresponds to $\lambda = 1$ and the right part to $\lambda = 5$.

	$\lambda = 1$			$\lambda = 5$			$\lambda = 1$			$\lambda = 5$		
α	I_4	W_4	W_2	I_4	W_4	W_2	I_4	W_4	W_2	I_4	W_4	W_2
	Bias$(\hat{\alpha}), T = 50$						MSE$(\hat{\alpha}), T = 50$					
0.1	-0.18	-0.29	-0.30	-0.26	-0.36	-0.38	0.22	0.18	0.18	0.23	0.18	0.18
0.3	-0.37	-0.62	-0.64	-0.25	-0.51	-0.54	0.24	0.25	0.25	0.26	0.24	0.25
0.5	-0.42	-0.66	-0.67	-0.27	-0.56	-0.57	0.22	0.24	0.25	0.26	0.22	0.24
0.7	-0.50	-0.60	-0.62	-0.17	-0.52	-0.52	0.20	0.19	0.19	0.25	0.19	0.20
0.9	-0.31	-0.70	-0.67	-0.05	-0.42	-0.50	0.07	0.14	0.13	0.02	1.18	0.06
	Bias$(\hat{\alpha}), T = 200$						MSE$(\hat{\alpha}), T = 200$					
0.1	-0.06	-0.22	-0.22	-0.07	-0.21	-0.21	0.06	0.05	0.05	0.05	0.05	0.05
0.3	-0.08	-0.24	-0.23	-0.05	-0.23	-0.20	0.05	0.06	0.06	0.06	0.06	0.06
0.5	-0.12	-0.24	-0.23	-0.06	-0.18	-0.21	0.06	0.05	0.06	0.06	0.05	0.05
0.7	-0.11	-0.19	-0.18	-0.03	-0.18	-0.16	0.05	0.04	0.04	0.06	0.04	0.04
0.9	-0.10	-0.13	-0.13	-0.03	-0.18	-0.15	0.03	0.02	0.02	0.01	0.02	0.02
	Bias$(\hat{\lambda}), T = 50$						MSE$(\hat{\lambda}), T = 50$					
0.1	0.49	0.53	0.53	3.05	3.88	3.95	0.76	0.68	0.67	13.31	12.03	11.89
0.3	0.84	1.23	1.27	4.05	6.96	6.69	1.27	1.32	1.27	30.64	30.39	29.77
0.5	1.81	2.06	1.93	7.60	9.46	8.93	3.15	3.01	2.89	80.80	68.39	72.34
0.7	5.06	1.51	1.62	18.55	7.90	9.14	17.22	7.56	7.87	349.59	194.19	186.92
0.9	22.35	1.06	3.04	14.25	179.31	194.98	175.30	194.21	158.52	223.09	10680.9	12615.4
	Bias$(\hat{\lambda}), T = 200$						MSE$(\hat{\lambda}), T = 200$					
0.1	0.13	0.29	0.30	0.79	1.78	1.70	0.15	0.15	0.15	2.92	3.09	2.95
0.3	0.29	0.44	0.53	0.95	2.72	2.54	0.28	0.28	0.30	6.13	6.43	6.54
0.5	0.40	0.84	0.74	1.26	3.48	3.91	0.63	0.57	0.58	15.24	14.00	13.27
0.7	0.94	1.42	1.19	3.62	7.61	6.99	2.07	1.53	1.42	55.49	43.54	41.32
0.9	7.01	10.70	2.50	8.55	39.61	33.37	35.97	254.66	9.51	115.02	872.67	904.90
	Bias$(\hat{\theta}), T = 50$						MSE$(\hat{\theta}), T = 50$					
0.1	-0.28	-0.35	-0.35	-0.29	-0.38	-0.39	0.15	0.15	0.15	0.15	0.15	0.15
0.3	-0.30	-0.45	-0.47	-0.32	-0.60	-0.58	0.24	0.26	0.24	0.28	0.30	0.29
0.5	-0.65	-0.75	-0.64	-0.45	-0.55	-0.46	0.55	0.66	0.70	0.68	0.67	0.70
0.7	-1.86	0.02	-0.04	-1.00	0.08	-0.15	3.28	2.41	2.47	2.96	2.44	2.17
0.9	-9.06	1.03	0.15	-1.19	-16.01	-17.50	28.23	49.92	35.62	1.51	82.83	94.09
	Bias$(\hat{\theta}), T = 200$						MSE$(\hat{\theta}), T = 200$					
0.1	-0.08	-0.12	-0.12	-0.09	-0.13	-0.12	0.03	0.03	0.03	0.03	0.03	0.03
0.3	-0.15	-0.23	-0.26	-0.10	-0.23	-0.23	0.06	0.05	0.06	0.06	0.06	0.06
0.5	-0.14	-0.42	-0.33	-0.05	-0.27	-0.30	0.12	0.12	0.12	0.12	0.11	0.11
0.7	-0.28	-0.63	-0.51	-0.13	-0.66	-0.60	0.37	0.31	0.32	0.46	0.36	0.33
0.9	-2.39	-4.86	-0.82	-0.59	-3.49	-2.86	6.00	54.49	2.61	0.89	7.01	6.81

Table 3: Sizes of LM and Wald ($\hat{W}_4$ and $\hat{W}_2$) test statistics. Nominal size is 0.05.

	$\lambda = 1$			$\lambda = 5$		
α	LM	$\hat{W}_4$	$\hat{W}_2$	LM	$\hat{W}_4$	$\hat{W}_2$
			$T = 50$			
0.1	0.104	0.140	0.116	0.135	0.134	0.101
0.3	0.094	0.146	0.120	0.170	0.179	0.167
0.5	0.141	0.178	0.164	0.158	0.187	0.164
0.7	0.154	0.148	0.116	0.177	0.169	0.143
0.9	0.203	0.105	0.090	0.217	0.059	0.043
			$T = 200$			
0.1	0.080	0.072	0.066	0.101	0.076	0.073
0.3	0.086	0.090	0.099	0.097	0.088	0.096
0.5	0.094	0.110	0.119	0.082	0.120	0.099
0.7	0.089	0.081	0.074	0.105	0.120	0.101
0.9	0.104	0.024	0.033	0.124	0.011	0.009

Table 4: ML, CLS and GMM ($\hat{W}$) estimates of Poisson and generalized Poisson AR(1) models. GP-GMM and P-GMM correspond to generalized Poisson and Poisson, respectively.

Production	Estimator	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\theta}$	LM/Wald
	ML	0.98	0.51		
Mechanical	CLS	0.81	3.98		
	P-GMM	0.96	0.26		3.16
	GP-GMM	0.96	2.48	-4.40	-0.02
	ML	0.98	0.66		
Sulphate	CLS	1.00	-0.50		
	P-GMM	0.97	0.20		2.42
	GP-GMM	0.98	0.16	0.15	0.02
	ML	0.90	1.88		
Sulphite	CLS	0.92	-0.18		
	P-GMM	0.89	0.23		0.54
	GP-GMM	0.93	0.05	1.21	0.95
	ML	0.97	1.84		
Paper	CLS	0.98	0.13		
	P-GMM	0.97	0.86		0.58
	GP-GMM	0.99	-0.11	0.66	0.55