EXPLANATORY VARIABLES IN THE AR(1) COUNT DATA MODEL *

Kurt Brännäs
Department of Economics, Umeå University
S-901 87 Umeå, Sweden

Umeå Economic Studies 381, 1995 (revised)

Abstract

Explanatory variables are incorporated into the count data (integer-valued) autoregressive model of order one. The properties of the resulting model are studied for both univariate and panel data specifications. Weighted and unweighted conditional least squares and conditional generalized method of moment estimators are introduced and evaluated by Monte Carlo experimentation. The results indicate that the least squares estimators perform well for realistically short time series. Tests against time dependent parameters are obtained and their properties are evaluated. Tests based on least squares perform well in small samples. An empirical illustration is included.

Key Words: Count data model, Moments, Least squares, Generalized method of moments, Monte Carlo, Prediction, Survival function.

*An earlier version of this paper was presented at the World Congress of the Econometric Society, Tokyo, 1995. Comments by colleagues and Colin Cameron have improved the exposition. The financial support from The Swedish Research Council for the Humanities and Social Sciences is acknowledged.

1. Introduction

In this paper we study the first order autoregressive model [AR(1) model] for count data, when basic parameters depend on explanatory variables. For the basic count data model, the Poisson model, the constant parameter AR(1) model was introduced by McKenzie (1985) and later discussed by McKenzie (1988), Al-Osh and Alzaid (1987) and others. In terms of the first two moments, this model and the Gaussian are related. In other improtant respects the models are quite different, however. For instance, the marginal Poisson distribution of the process has only one parameter depending on basic process parameters, while the Gaussian has two.

Regression models for time series count data are becoming increasingly often applied. Different approaches were initially suggested by Zeger (1988), Zeger and Qagish (1988) and Smith (1979). Recent work of Brännäs and Johansson (1992, 1994) and others suggest that these approaches are empirically feasible. The only regression model with an explicit lag structure is that of Zeger and Qagish (1988). It is quite different from the one considered here, though. In the other regression models the autocorrelation structure is assumed to be due to a latent process. A different but pure time series model was given by Jacobs and Lewis (1983).

In this paper we study in more detail the consequences and adaptations that are required when explanatory variables are included in the count data AR(1) model (see also Berglund and Brännäs, 1995). We obtain new conditional least squares (CLS) and generalized method of moment (GMM, Hansen, 1982) estimators for the AR(1) model containing explanatory variables. Test statistics are derived for testing linear restrictions on parameters. The obtained estimators and test statistics are evaluated and compared by finite sample Monte Carlo experimentation. For the case of Poisson and time invariant parameters the CLS and maximum likelihood estimators were partly evaluated by Al-Osh and Alzaid (1987), Jin-Guan and Yuan (1991), Ronning and Jung (1992) and by Brännäs (1994) who also compared them to GMM.

Adopted predictors and prediction error variances extending previous results of Brännäs (1994) are derived. Multipliers and implied survival function are also given. For a multivariate dependent AR(1) Poisson model appropriate for panel data with explanatory variables, corresponding measures are derived and some comments on inference are made.

The model is formally introduced in Section 2. The estimators and test statistics are given in Section 3. The design and results of a finite sample Monte Carlo experiment are given in Section 4. Section 5 gives predictors, dynamic multipliers and implied survival function. A brief empirically based illustration is included in Section 6. Section 7 presents the panel data model. The final section concludes.

2. The Model

Consider a stationary $\{y_t\}$ process with Poisson marginal densities generated according to an autoregressive model of order one [the AR(1) model]. To start we let parameters be constant over time, i.e. they are not affected by explanatory variables. We write

$$y_t = \alpha \circ y_{t-1} + \epsilon_t, \qquad t = 2, \dots, T, \tag{1}$$

where $\alpha \circ y = \sum_{i=1}^{y} u_i$, with u_i a sequence of independent and identically distributed (i.i.d.) 0-1 random variables independent of y and 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 $E(\alpha \circ y) = E_y[yE(u_i|y)] = \alpha E(y)$ and that $V(\alpha \circ y) = V_y[E(\alpha \circ y|y)] + E_y[V(\alpha \circ y|y)] = V_y(\alpha y) + E_y[\alpha(1-\alpha)y] = \alpha^2 V(y) + \alpha(1-\alpha)E(y)$.

The ϵ_t is assumed 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 with time invariant α and λ have been detailed upon by Al-Osh and Alzaid (1987) and for a slightly different paramet- rization 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, Brännäs, 1994, Berglund and Brännäs, 1994). The negative binomial allows for overdispersion, while the generalized Poisson model allows for both under- and overdispersion.

To make the model more attractive for econometric application, we need to extend the model such that explanatory variables are made part of the model. As discussed by Berglund and Brännäs (1994) in the context of stock of plant time series data, the α parameter represents the survival probability of a plant, which, for instance, could depend on the business cycle phase. The λ represents the mean entry of plants and may, e.g., depend on market characteristics that obviously vary over time. Hence, we introduce explanatory variables through these parameters, as $\alpha_t \in (0,1)$ and $\lambda_t > 0$. Obviously, there are a number of potential candidates satisfying these range conditions. We assume two quite convenient and in other corresponding situations widely adopted specifications; the logistic distribution function, i.e.

$$\alpha_t = 1/[1 + \exp(\mathbf{x}_t \boldsymbol{\beta})] \tag{2}$$

and the exponential function

$$\lambda_t = \exp(\mathbf{z}_t \boldsymbol{\gamma}). \tag{3}$$

The explanatory variable vectors $\mathbf{x}_t$ and $\mathbf{z}_t$ are treated as fixed and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the corresponding vectors of unknown parameters.

The full model can now be written

$$y_t = \alpha_t \circ y_{t-1} + \epsilon_t, \qquad t = 1, \ldots, T. \tag{4}$$

The following moment relations can be shown to hold true for the time varying model (4):

$$E(y_t) = \alpha_t E(y_{t-1}) + \lambda_t \tag{5}$$

$$V(y_t) = \alpha_t^2 V(y_{t-1}) + \alpha_t(1-\alpha_t)E(y_{t-1}) + \lambda_t \tag{6}$$

$$\gamma(t,k) = \left[\prod_{i=0}^{k-1} \alpha_{t-i}\right] V(y_{t-k}), \quad k = 1, 2, \ldots \tag{7}$$

$$\rho(t,k) = \gamma(t,k) / \left[V(y_t)V(y_{t-k})\right]^{\frac{1}{2}}$$

$$= \left[\prod_{i=0}^{k-1} \alpha_{t-i}\right] \left[\frac{V(y_{t-k})}{V(y_t)}\right]^{\frac{1}{2}}, \quad k = 1, 2, \ldots, \tag{8}$$

where $\gamma(t,k)$ and $\rho(t,k)$ are the autocovariance and autocorrelation functions at time t and lag k, respectively. In this time dependent case the autocorrelation depends on the position in the time series through both α_t and λ_t. With variances approximately equal we expect that $\rho(t,k) > \rho(t,k+1)$ for any t and $k = 1, 2, \ldots$. As can be expected, the moment expressions (5)-(8) reduce to those of the time invariant case when the α_t and λ_t parameters are time invariant. For identical initial specifications of $E(y_1)$ and $V(y_1)$ all subsequent expected values and variances are equal. The marginal distribution of y_t is Poisson with parameter $E(y_t) = V(y_t)$.

In Figures 1–2 we illustrate the time series and autocorrelation properties. Figure 1 gives $\rho(t,1)$ for a series generated according to the design used in the Monte Carlo experiment of Section 4, below. In this particular case there is substantial variation in $\rho(t,1) = \alpha_t \left[V(y_{t-1})/V(y_t)\right]^{\frac{1}{2}}$. Not surprisingly, there is a strong covariation between the autocorrelation and both the observed series and the mean function of the process. Figure 2 plots $\rho(t,k)$, for $k = 1, 2, 3$, and demonstrates the expected size relationship between autocorrelations.

3. Estimation and Testing

In this section we consider the estimation of unknown model parameters by conditional least squares (CLS) and generalised method of moments (GMM) estimators for models with the same moment structure as Poisson. We also give test statistics for linear restrictions. While, maximum likelihood estimation may be feasible for the Poisson model, it is quite intractable for some other distributions, such as the generalized Poisson (Alzaid and Al-Osh, 1993, Brännäs, 1994) and therefore it is not considered here.

3.1 Conditional Least Squares

The one-step ahead prediction error is of the form

$$e_t = y_t - \alpha_t y_{t-1} - \lambda_t,$$

where α_t and λ_t are as in (2)-(3).

The CLS estimator of $\boldsymbol{\psi}' = (\boldsymbol{\beta}', \boldsymbol{\gamma}')$ minimizes the sum of squared prediction errors $Q = \sum_{t=2}^{T} e_t^2$. As in many nonlinear least squares problems the application of the Gauss-Newton algorithm to minimize Q is straightforward. At the iterative step $k+1$ the algorithm gives the new estimate $\boldsymbol{\psi}^{k+1}$ according to

$$\boldsymbol{\psi}^{k+1} = \boldsymbol{\psi}^k + \left[\sum_{t=2}^{T} \frac{\partial e_t}{\partial \boldsymbol{\psi}} \cdot \frac{\partial e_t}{\partial \boldsymbol{\psi}'}\right]_{|k}^{-1} \left[\sum_{t=2}^{T} e_t \frac{\partial e_t}{\partial \boldsymbol{\psi}}\right]_{|k},$$

where the gradient and the approximation to the Hessian matrix are evaluated at the previous estimate $\boldsymbol{\psi}^k$. The required first order derivatives are given by

$$\begin{aligned} \partial e_t/\partial \boldsymbol{\beta} &= y_{t-1}\mathbf{x}_t' \Lambda_t/(1+\Lambda_t)^2 \\ \partial e_t/\partial \boldsymbol{\gamma} &= -\mathbf{z}_t' \lambda_t, \end{aligned}$$

where $\Lambda_t = \exp(\mathbf{x}_t\boldsymbol{\beta})$.

On convergence, the conventional covariance matrix of the estimator $\hat{\boldsymbol{\psi}}$ is estimated as

$$Cov(\hat{\boldsymbol{\psi}}) = \mathcal{F}^{-1} = \left[\sum_{t=2}^{T} \frac{\partial e_t}{\partial \boldsymbol{\psi}} \cdot \frac{\partial e_t}{\partial \boldsymbol{\psi}'}\right]^{-1},$$

which is to be evaluated at the estimator $\hat{\boldsymbol{\psi}}$.

Note that the CLS estimator is appropriate for any model satisfying the same predictor expression, i.e. the Poisson assumption is not required. For instance, the generalized Poisson model can be estimated by this approach at the price of not enabling separate estimation of a parameter characterizing this particular distribution from a constant term in λ_t. Consistency and asymptotic normality follows from, for instance, Wooldridge (1994, Theorems 4.3 and 4.4).

As is obvious from the variance of y_t, given in (6), the process is heteroskedastic whenever the parameters are time dependent. The same holds true for the variance of the prediction error. This will have a biasing effect on the covariance matrix estimator.

A simple approach is to use the CLS estimator $\hat{\boldsymbol{\psi}}$, but to correct the covariance matrix estimator for heteroskedasticity. An estimator of the asymptotic covariance matrix is of the form

$$Cov(\hat{\boldsymbol{\psi}}) = \mathcal{F}^{-1}\mathcal{J}\mathcal{F}^{-1},$$

where

$$\mathcal{J} = \sum_{t=2}^{T} e_t^2 \frac{\partial e_t}{\partial \boldsymbol{\psi}} \cdot \frac{\partial e_t}{\partial \boldsymbol{\psi}'}$$

is evaluated at $\hat{\boldsymbol{\psi}}$. The conditional variance of the prediction error is $E(e_t^2|y_{t-1}) = \alpha_t(1-\alpha_t)y_{t-1} + \lambda_t$. With estimated parameters this can be used instead of e_t^2 in the $\mathcal{J}$ matrix of the corrected $Cov(\hat{\boldsymbol{\psi}})$.

Alternatively, we may attempt to find a more efficient estimator from the weighted criterion function

$$Q = \sum_{t=2}^{T} \frac{e_t^2}{E(e_t^2|y_{t-1})}, \tag{9}$$

where both the numerator and the denominator are functions of the unknown $\boldsymbol{\psi}$ vector.

A simpler weighted CLS estimator starts from CLS estimation to obtain an estimate of $E(e_t^2|y_{t-1})$. In a second step the weighted CLS estimator minimizes Q in (9) to obtain $\hat{\boldsymbol{\psi}}$ estimates with the $E(e_t^2|y_{t-1})$ estimator treated as known. The gradient vector is given by $\sum_{t=2}^{T} e_t c_t \ (\partial e_t/\partial \boldsymbol{\psi})$ and the approximate Hessian also used to estimate the asymptotic covariance matrix of $\hat{\boldsymbol{\psi}}$ is given by

$$Cov(\hat{\boldsymbol{\psi}}) = \left[\sum_{t=2}^{T} c_t \frac{\partial e_t}{\partial \boldsymbol{\psi}} \cdot \frac{\partial e_t}{\partial \boldsymbol{\psi}'}\right]^{-1},$$

where $c_t = 1/E(e_t^2|y_{t-1})$ is evaluated at the first step estimates. The consistency and asymptotic normality of the estimator follows directly (e.g., Wooldridge, 1994, Theorems 4.3 and 4.4).

Consider next the testing of linear restrictions on the parameter vector $\boldsymbol{\psi}$ of the Poisson AR(1) model. Important special cases include testing hypotheses about individual parameters as well as hypotheses of time invariant α and λ parameters (or equivalently that all slope parameters in $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are equal to zero).

For the CLS or the weighted CLS estimators testing is straightforward by performing Wald type tests based on the corresponding appropriate covariance matrix estimator. We formulate r linear restrictions $\mathbf{R}\boldsymbol{\psi} = \mathbf{0}_r$ and use the asymptotically $\chi^2(r)$ distributed test statistic

$$\mathrm{W} = \hat{\boldsymbol{\psi}}'\mathbf{R}'[\mathbf{R}Cov(\hat{\boldsymbol{\psi}})\mathbf{R}']^{-1}\mathbf{R}\hat{\boldsymbol{\psi}} \sim \chi^2(r).$$

3.2 Conditional GMM

In this subsection we consider the GMM estimator first introduced by Hansen (1982). The estimator is consistent and efficient under certain conditions. In particular, we employ the conditional GMM estimator introduced by Newey (1985) and Tauchen (1986). The conditioning set consists of the histories of previous values (the predetermined variables) $y_{t-1},\ y_{t-2},\ \ldots$ and of $\mathbf{x}_t,\ \mathbf{x}_{t-1},\ \ldots$ and $\mathbf{z}_t,\ \mathbf{z}_{t-1},\ \ldots$. The conditional GMM estimator for time invariant Poisson and generalized Poisson AR(1) models was studied by Brännäs (1994).

The conditional 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 conditional 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 GMM 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 estimator $\hat{\boldsymbol{\psi}}$ from step one is used to form $\hat{\mathbf{W}}$ after which the q in (9) is minimized.

For past observations $y_1, \ldots, y_{t-1}$, the conditional mean of y_t in the AR(1) model is $\alpha_t y_{t-1} + \lambda_t$ and the one-step-ahead prediction error is $e_t = y_t - \alpha_t y_{t-1} - \lambda_t$. The normal equations for CLS correspond to empirical moment restrictions: $T^{-1}\sum_{t=2}^{T} e_t(\partial e_t/\partial\boldsymbol{\beta}) = \mathbf{0}$ for $\boldsymbol{\beta}$ and $T^{-1}\sum_{t=2}^{T} e_t(\partial e_t/\partial\boldsymbol{\gamma}) = \mathbf{0}$ for $\boldsymbol{\gamma}$, with corresponding theoretical, conditional moments $E[e_t(\partial e_t/\partial\boldsymbol{\beta})|y_{t-1}] = \mathbf{0}$ and $E[e_t(\partial e_t/\partial\boldsymbol{\gamma})|y_{t-1}] = \mathbf{0}$. It follows that the CLS estimator can be interpreted as a conditional GMM estimator with equal numbers (k) of unknown parameters and restrictions and with $\hat{\mathbf{W}}$ equal to the identity matrix of order k, $\mathbf{I}_k$.

Additional conditional moment restrictions are available. For instance, Alzaid and Al-Osh (1988) give a general result that we adapt into $V(y_t|y_{t-1}) = E(e_t^2|y_{t-1}) = \alpha_t(1 - \alpha_t)y_{t-1} + \lambda_t$. In addition, $E(e_t e_{t-1}|y_{t-1}) = 0$ can be employed. We recognize that there may be other higher order conditional moment restrictions that could be used for GMM estimation.

The estimation of $\mathbf{W}$ is most easily based on the consistent Newey and West (1987a) estimator

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

where $\hat{\boldsymbol{\Gamma}}_j = T^{-1}\sum_{t=j+1}^{T}\mathbf{m}_t(\hat{\boldsymbol{\psi}})'\mathbf{m}_{t-j}(\hat{\boldsymbol{\psi}}),\ j = 0, 1, \ldots, p$, and p has to be given.

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

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

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

To test the time invariance 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 MacKinnon, 1993, ch. 17). In general, the LM statistic is 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 $\tilde{\boldsymbol{\psi}}_0$ is a GMM estimator of $\boldsymbol{\psi}_0$ under H_0 and $\mathbf{W}$ is estimated using these estimates. We may use the CLS estimator of β_0 obtained from $\alpha = 1/[1+\exp(\beta_0)]$ and of γ_0 from $\lambda = \exp(\gamma_0)$ under H_0 and the moment restrictions of the full CLS estimator under H_A. The restriction vector under H_A is $\mathbf{m}(\boldsymbol{\psi}_0)' = (\tilde{\alpha}(1-\tilde{\alpha})[\sum_{t=2}^T e_t y_{t-1} : \sum_{t=2}^T e_t \mathbf{x}_t y_{t-1}] : -\tilde{\lambda}[\sum_{t=2}^T e_t : \sum_{t=2}^T e_t \mathbf{z}_t])$. Asymptotically, LM in (13) is distributed as a $\chi^2(r)$ variate, where r is the number of restrictions imposed on $\boldsymbol{\psi}$.

For a model estimated under H_A, i.e. under the time varying specification with α_t and λ_t, the application of the Wald statistic with linear restriction $\mathbf{R}\boldsymbol{\psi} = \mathbf{0}_r$ is again straightforward. We get

$$\mathrm{W} = T\hat{\boldsymbol{\psi}}'\mathbf{R}'[\mathbf{R}[\mathbf{G}'\hat{\mathbf{W}}^{-1}\mathbf{G}]^{-1}\mathbf{R}']^{-1}\mathbf{R}\hat{\boldsymbol{\psi}} \sim \chi^2(r).$$

4. Small Sample Performance

The small sample Monte Carlo experiment aims at providing a partial comparison and evaluation of the properties of the different versions of the CLS estimator and the conditional GMM estimator in the Poisson case. In addition, we evaluate the size and power properties of linear restriction tests.

4.1 Design

The following basic design with respect to the α_t and λ_t parameters is employed:

$$\alpha_t = \frac{1}{1+\exp(\beta_0 + \beta_1 x_t)} \qquad \text{and} \qquad \lambda_t = \exp(\gamma_0 + \gamma_1 z_t).$$

The scalar x_t and z_t variables are generated from Gaussian AR(1) processes with unit variance, white noise innovations and with AR(1) parameters 0.7 and 0.8, respectively. Both variables are treated as fixed over replications. The parameters are set such that α_t takes approximate average values 0.7 or 0.9 ($\beta_0 = -1$ or -2, with β_1 free to be varied); for λ_t the mean is varied between approximate average levels 1 and 5 by setting $\gamma_0 = 1$ and γ_1 between -0.3 and 0.3. Two sample sizes are employed, $T = 50$ and 200, and the number of replications is 1000. In generating the data an initial set of 150 observations in each replication is discarded.

For both unweighted and weighted [weights $E(e_t^2|y_{t-1})$ estimated in a first step by CLS] CLS estimation Gauss-Newton algorithms are used, while for conditional GMM a simplex algorithm is used. The GMM estimator is obtained in two steps: (i) CLS estimates are used to get a Newey and West (1987a) estimator of $\mathbf{W}$, with $p = 4$, and (ii) the full GMM estimator with 6 restrictions for q is minimized. The restrictions beyond those corresponding to CLS are the conditional prediction error variance and the conditional lag one autocovariance of the prediction error.

4.2 Results

The biases and MSEs for some selected values on the parameters are given in Tables A – C. The overall picture is quite clear – for realistically short time series, there is no gain in bias and MSE properties from using the more elaborate GMM estimator instead of either of the CLS estimators. The relative performance of GMM improves for $T = 200$ and is closer to or even better than CLS for even larger sample sizes. No doubt, GMM will become the more efficient one of the two for very long time series.

There is a substantial improvement in MSE performance in most cases when the weighted CLS estimator is used instead of the unweighted CLS. The difference between estimators is smaller for the larger sample size of $T = 200$. This also implies that the GMM estimator is inferior to the weighted CLS estimator in the MSE sense. With respect to bias we find no large difference between the two CLS estimators, while both have smaller bias than GMM. Again, differences get smaller for $T = 200$.

The size and power properties for a two-sided t-test of $\mathrm{H}_0 : \gamma_1 = 0$ are summarized in Figure 3. The test is based on the CLS estimator and is evaluated at a nominal significance level of 5 per cent (assuming normality). The figure provides a comparison of the effect of using e_t^2 or $E(e_t^2|y_{t-1})$ in the $\mathcal{J}$ matrix of the corrected covariance matrix estimator. While there is hardly any notable difference between the two alternatives for the larger sample size of $T = 200$, there are considerable differences in favour of $E(e_t^2|y_{t-1})$ for $T = 50$. The sizes of this alternative test are not significantly different from the nominal 0.05 for either sample size. With $\beta_0 = -2$ and a higher average of about 0.9 for α_t, the size is significantly too high for $T = 50$ and close to being insignificantly different from the nominal size for $T = 200$ for the better $E(e_t^2|y_{t-1})$ variant.

The size and power properties of the Wald tests based on CLS with the $E(e_t^2|y_{t-1})$ corrected covariance matrix estimator and the WCLS procedure are comparable for $\beta_1 = -1$, while there are differences for $\beta_1 = -2$, cf. Figure 4. In this latter case when the average of α_t is higher, the weighted CLS estimator results in improved size properties. For the Wald test based on the GMM estimator, the sizes are significantly too high for both sample sizes (0.25 for $T = 50$ and 0.12 for $T = 200$).

In Figure 5 we give sizes and powers for the testing of time invariance, i.e. of the joint test of $\beta_1 = 0$ and $\gamma_1 = 0$, based on a Wald test for the CLS estimator and the $E(e_t^2|y_{t-1})$ corrected covariance matrix estimator and an LM test for the GMM estimator. In the illustrative special case of the figure, sizes are not significantly different from the nominal sizes of 0.05. As expected the power is increasing with sample size. While the LM test is less powerful for $T = 50$, powers are almost identical for $T = 200$.

In summary, we find that the best overall estimator and test performance is obtained with the weighted CLS estimator and associated Wald test statistics. Computationally, this estimator is simple to code and fast. The GMM estimator was found to diverge in 1–2 replications out of 1000 in each cell for $T = 50$.

5. Prediction, Multipliers and Survival

In this section we consider the prediction of a future value y_{T+h} given that we have observed the series and associated explanatory variables up through time T, i.e. $y_1, \dots, y_T$ is observed. The parameters are treated as known. We also give some expressions for dynamic multipliers given changes in $\mathbf{x}$ and/or in $\mathbf{z}$. Finally, we outline the implied survival function and some measures based on this.

5.1 Prediction

By repeated substitution we may write the future values of the process

$$y_{T+h} = \left[\prod_{i=1}^{h} \alpha_{T+i}\right] \circ y_T + \sum_{i=1}^{h} \left[\prod_{j=i+1}^{h} \alpha_{T+j}\right] \circ \epsilon_{T+i}, \qquad h = 1, 2, \dots, \tag{14}$$

where in this and subsequent expressions $\alpha_j = 0, j > h$, and the equality should be interpreted as one of equality in distribution. To predict we obviously need to know the future values of explanatory variables $\mathbf{x}_{T+j}$ and $\mathbf{z}_{T+j}$ for $j = 1, \dots, h$.

From (14) we obtain the h-step ahead predictor

$$\begin{aligned} \hat{y}_{T+h|T} &= E(y_{T+h}|y_1, \dots, y_T) \\ &= \left[\prod_{i=1}^{h} \alpha_{T+i}\right] y_T + \sum_{i=1}^{h} \lambda_{T+i} \left[\prod_{j=i+1}^{h} \alpha_{T+j}\right]. \end{aligned} \tag{15}$$

As h gets larger the products in (14) become smaller and the impact of y_T is diminishing. The second term is of the form $\lambda_{T+h} + \lambda_{T+h-1}\alpha_{T+h} + \lambda_{T+h-2}\alpha_{T+h}\alpha_{T+h-1} + \dots$ Hence, the effects of more remote λ_t's decline approximately in a geometric way. For the time invariant case Brännäs (1994) obtains the predictor $\alpha^h y_T + \lambda(1-\alpha^h)/(1-\alpha)$, which approaches the mean of the process as $h \to \infty$. 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 $h = 1$ takes the form

$$\begin{aligned} V(e_{T+1}) &= E\left[(\alpha_{T+1} \circ y_T - \alpha_{T+1} y_T)^2\right] + E\left[(\epsilon_{T+1} - \lambda_{T+1})^2\right] \\ &= \alpha_{T+1}(1-\alpha_{T+1})E(y_T) + \lambda_{T+1}. \end{aligned} \tag{16}$$

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 form

$$V(e_{T+h}) = \prod_{i=1}^{h} \alpha_{T+i}\left[1-\alpha_{T+i}\right] E(y_T) + \sum_{i=1}^{h} \lambda_{T+i} \left[\prod_{j=i+1}^{h} \alpha_{T+j}\right]. \tag{17}$$

For the time invariant case the prediction error variance $(1-\alpha^{2h})V(y_T)$ increases with the length of the forecast horizon h and approaches the variance of the process (Brännäs, 1994). In view of (17), it is not possible to draw a related general conclusions, as the variances depend on time dependent α_t and λ_t.

5.2 *Multipliers*

Consider dynamic multipliers in terms of the expected value of the process, $E(y_t)$. By repeated substitution in (5) we obtain

$$E(y_t) = \prod_{i=s}^{t} \alpha_i E(y_{s-1}) + \sum_{i=s}^{t} \lambda_i \left[\prod_{j=i+1}^{t} \alpha_j \right],$$

where $\alpha_j = 0$ for $j > t$.

To simplify assume that $\mathbf{x}_t$ and $\mathbf{z}_t$ do not contain common variables. General expressions for the delayed multipliers for variables x_{ks} and z_{ks} for some $s < t$ can after some manipulation be written on the forms

$$\begin{aligned} \frac{\partial E(y_t)}{\partial x_{ks}} &= -\beta_k e^{\mathbf{x}_s \boldsymbol{\beta}} \alpha_s \prod_{i=s}^{t} \alpha_i E(y_{s-1}) \\ \frac{\partial E(y_t)}{\partial z_{ks}} &= \gamma_k \lambda_s \left[\prod_{j=s+1}^{t} \alpha_j \right]. \end{aligned}$$

To implement the multiplier with respect to x_{ks} we may replace $E(y_{s-1})$ with y_{s-1}. In fact, this is what arises when taking the derivative on the predictor in (15). Intermediate run multipliers are obtained by summing the delayed multipliers. We get

$$\begin{aligned} \sum_{i=s}^{t} \frac{\partial E(y_t)}{\partial x_{ks}} &= -\beta_k e^{\mathbf{x}_s \boldsymbol{\beta}} \alpha_s \prod_{i=s}^{t} \alpha_i E(y_{s-1}) \\ &\quad -\beta_k e^{\mathbf{x}_{s+1} \boldsymbol{\beta}} \alpha_{s+1} \left[\prod_{i=s}^{t} \alpha_i E(y_{s-1}) + \lambda_s \left[\prod_{j=s+1}^{t} \alpha_j \right] \right] \\ &\quad \ldots \\ &\quad -\beta_k e^{\mathbf{x}_t \boldsymbol{\beta}} \alpha_t \left[\prod_{i=s}^{t} \alpha_i E(y_{s-1}) + \lambda_s \prod_{j=s+1}^{t} \alpha_j + \ldots + \lambda_{t-1} \alpha_t \right] \\ \sum_{i=s}^{t} \frac{\partial E(y_t)}{\partial z_{ks}} &= \gamma_k \sum_{i=s}^{t} \lambda_i \left[\prod_{j=i+1}^{t} \alpha_j \right]. \end{aligned}$$

The long run multipliers are obtained by letting $t \to \infty$ in the sums of the intermediate multipliers.

5.3 *Survival Function*

The variables u_i in the binomial thinning operator associated with (1) and (4) are independent over time. Hence, a component u_i in y can be viewed as new at some time s and to have survived with probability α_{s+1} to time $s+1$, with probability $\alpha_{s+1}\alpha_{s+2}$ to time $s+2$, and so on. The resulting survival function, $\bar{F}(\cdot)$, is

$$\bar{F}(s+k) = \begin{cases} 1, & k = 0 \\ \prod_{i=s+1}^{s+k} \alpha_i, & k = 1, 2, \ldots \end{cases}$$

With estimated parameters survival functions can therefore readily be calculated. In addition, based on the survival function measures such as the mean, median or mean residual survival time can be calculated.

The mean survival time of a component which is new at time s is given by

$$m(s) = 1 + \sum_{k=1}^{\infty} \prod_{i=s+1}^{s+k} \alpha_i.$$

The median survival time is located in the time interval $(d, d+1)$ for which $\bar{F}(d) > 0.5$ and $\bar{F}(d+1) < 0.5$. The mean residual time starting at s but after $d > 1$ is $\sum_{i=d}^{\infty} \bar{F}(i)/\bar{F}(d)$.

6. Empirical Illustration

Consider as an illustration of some of the introduced techniques the number of Swedish mechanical paper and pulp mills 1921–1981, cf. Figure 6. This industrial production technology is obviously on its way out and new production capacity is created in more high tech plants.

Table 1 gives parameter estimates for a simple model, where the industrial gross profit margin and GNP are used as explanatory variables. The fit (based on weighted CLS estimates) of the model is exhibited in Figure 6 and is quite good. In Figure 7 the estimated survival probability (for surviving one year) is plotted together with the estimated mean entry. Evidently, there are negative trends in both measures and in addition we note a negative correlation between the two. This is not surprising in view of the relative constancy of the expected value of the process, cf. (5). Both measures appear to provide pictures of business cycles.

In Figures 8 and 9 we give the implied survival functions starting at 1921 and 1931 and mean residual life times. The survival functions are quite different indicating a smaller survival probability over the first few years for mills started in 1931. The mean residual life times for mills started in 1921 and 1931 follow the same trajectory. There is a through around 1930 and a peak in mean residual time around 1940. After this the mean residual life time is declining. The implied mean life times are 6.9 years for a new mill in 1921 and 5.7 years for a mill started in 1931. A Wald test of time invariant α and λ rejects the hypothesis ($p = 0$).

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

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix}_t = \begin{pmatrix} \alpha_{1t} & 0 & \dots & 0 \\ 0 & \alpha_{2t} & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \alpha_{Mt} \end{pmatrix} \circ \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix}_{t-1} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_M \end{pmatrix}_t$$

or

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

for $t = 2, \ldots, T$. The parameters are different for different cross-section units as well as over time 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 ϵ_{it}^* are independent Poisson variables with parameters $\lambda_{it},\ i = 1, \ldots, M,\ t = 1, \ldots, T$. The scalar $\xi_t,\ t = 1, \ldots, T$, is independently Poisson distributed with parameter δ (e.g., Johnson and Kotz, 1969, ch. 11). Note, that without larger difficulties δ can be allowed to be time dependent and specified as, for instance, $\delta_t = \exp(\mathbf{w}_t \boldsymbol{\theta})$. Let $\boldsymbol{\Gamma}_\epsilon(t,s) = E[\boldsymbol{\epsilon}_t - E(\boldsymbol{\epsilon}_t)][\boldsymbol{\epsilon}_{t+s} - E(\boldsymbol{\epsilon}_{t+s})]'$ be the autocovariance matrix at lag s. The assumptions imply $\boldsymbol{\Gamma}_\epsilon(t,s) = 0,\ s \geq 1$, and

$$\boldsymbol{\Gamma}_\epsilon(t,0) = \begin{pmatrix} \lambda_{1t} + \delta & \delta & \ldots & \delta \\ \delta & \lambda_{2t} + \delta & \ldots & \delta \\ \vdots & \vdots & \ddots & \vdots \\ \delta & \delta & \ldots & \lambda_{Mt} + \delta \end{pmatrix}.$$

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

$$\mathbf{A}_t \circ \mathbf{y}_{t-1} = \left(\alpha_{1t} \circ y_{1,t-1} = \sum_{i=1}^{y_{1,t-1}} u_{i1}, \ldots, \alpha_{Mt} \circ y_{M,t-1} = \sum_{i=1}^{y_{M,t-1}} u_{iM} \right)',$$

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 (18) is

$$E(\mathbf{y}_t) = \mathbf{A}_t E(\mathbf{y}_{t-1}) + \boldsymbol{\lambda}_t + \delta \mathbf{1},$$

where $\boldsymbol{\lambda}_t' = (\lambda_{1t}, \ldots, \lambda_{Mt})$ and $\mathbf{1}$ an M column vector of ones. After some algebraic manipulation of off-diagonal elements we can write the covariance matrix of (18) as

$$Cov(\mathbf{y}_t) = \mathbf{A}_t Cov(\mathbf{y}_{t-1}) \mathbf{A}_t + \boldsymbol{\Gamma}_\epsilon(t,0).$$

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

The h-steps ahead predictor of (18) takes the form

$$\hat{\mathbf{y}}_{T+h|T} = \prod_{i=1}^{h} \mathbf{A}_{T+i} \mathbf{y}_T + \sum_{i=1}^{h} (\boldsymbol{\lambda}_{T+i} + \delta \mathbf{1}) \prod_{j=i+1}^{h} \mathbf{A}_{T+j}$$

with $\mathbf{A}_{T+h+1} = \mathbf{I}$. The M-variate one-step-ahead prediction error

$$\mathbf{e}_{T+h} = \mathbf{y}_{T+h} - \mathbf{A}_{T+h} \mathbf{y}_{T+h-1} - \boldsymbol{\lambda}_{T+h} - \delta \mathbf{1}$$

has zero mean and unconditional prediction error variance

$$V(\mathbf{e}_{T+1}) = \mathbf{A}_{T+1}(\mathbf{I} - \mathbf{A}_{T+1}) \mathrm{diag}\{E(\mathbf{y}_T)\} + \mathrm{diag}\{\boldsymbol{\lambda}_{T+1}\} + \delta \mathbf{1}\mathbf{1}',$$

which has a variance that is inflated by δ in comparison with the univariate case. The δ is also the covariance between one-step ahead prediction errors.

The conditional one-step-ahead prediction error variance is useful for estimation purposes and is given by

$$E(\mathbf{e}_t'\mathbf{e}_t|\mathbf{y}_{t-1}) = \mathbf{A}_t(\mathbf{I}-\mathbf{A}_t)\mathbf{y}_{t-1} + \mathrm{diag}\{\boldsymbol{\lambda}_t\} + \delta\mathbf{1}\mathbf{1}'.$$

For this panel data model the conditional least squares (CLS) estimator minimizing $\sum_{t=2}^{T}\mathbf{e}_t'\mathbf{e}_t$ is directly applicable. In this case too, the covariance matrix estimator need to be adapted for both heteroskedasticity and for the cross-covariances. Under time invariant α_i and λ_i, $i=1,\ldots,M$, however, the λ_i and δ cannot be separately estimated. GMM estimation is applicable with similar types of conditional moment restrictions as in the univariate case. In addition, we may use the sample cross-covariance between cross-section units as a useful restriction for δ.

8. Conclusions

In principle, it is straightforward to apply the estimators and test statistics to other distributions than the Poisson. As long as the conditional mean structure is not changed no changes need to be made. When the conditional mean contains additional parameters one would expect that additional and/or changed conditional moment restrictions need to be used to separately estimate additional parameters.

The CLS estimators behave relatively better than the GMM estimator for realistically short time series lengths. There is some efficiency gain from using weighting to correct for heteroskedasticity for CLS. The weights are determined by a general expression for the conditional prediction error. Therefore, we can expect the estimator to be robust to distributional misspecifications as long as the predictor is well specified.

For the estimation of the full panel data model related estimators can be considered. Whether additional conditional moment restrictions are required for the estimation of the δ parameter is an open issue. We will return to this in forthcoming empirical work. We note that with a potentially large number of cross section units, it will be necessary to look into feasible numerical solutions and into, say, sequential and conditional GMM estimation.

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.

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.

Berglund, E. and Brännäs, K. (1995). Entry and Exit of Plants: A Study based on Swedish Panel Count Data. In Brännäs, K. and Stenlund, H. (eds.) *Proceedings of the Second Umeå – Würzburg Conference in Statistics* (Umeå University, Umeå).

Brännäs, K. (1994). Estimation and Testing in Integer–Valued AR(1) Models. Umeå Economic Studies 335 (revised).

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 Third Würzburg – Umeå Conference in Statistics* (University of Würzburg, Würzburg).

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

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

Gallant, A.R. (1987). *Nonlinear Statistical Models* (Wiley, New York).

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).

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, New York).

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

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.

Wooldridge, J.M. (1994). Estimation and Inference for Dependent Processes. In Engle, R.F. and McFadden, D.L. (eds.) *Handbook of Econometrics* **4** (North-Holland, Amsterdam).

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.

Table 1: Estimation results (t-values in parantheses) and variable definitions.

Variable/Parameter	CLS	Weighted CLS
	Survival Probability	
Constant (β_1)	3.6048	1.0810
	(5.36)	(0.02)
Gross Profit Margin (β_2)	-0.0550	-0.0261
(1950-72=100)	(6.42)	(82.69)
	Mean Entry	
Constant (γ_1)	5.0507	4.6197
	(8.32)	(14608)
Gross Profit Margin (γ_2)	-0.0375	-0.0261
(1950-72=100)	(6.08)	(82.41)
GNP (γ_3)	-0.0012	-0.0017
(1900=100)	(3.28)	(5.41)

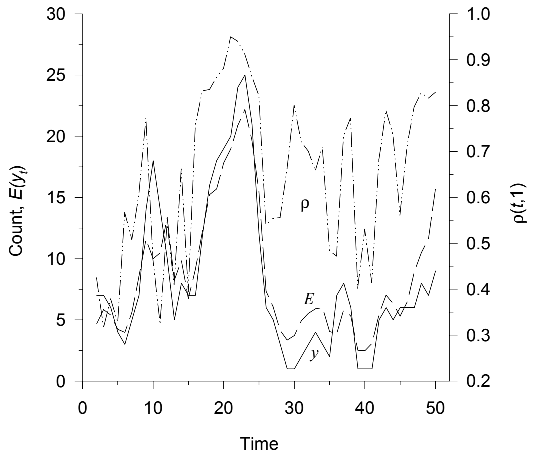

Figure 1: Series y_t (y, solid line), $E(y_t)$ (E, long dashed line) and $\rho(t,1)$ (ρ, dash dot dot line) for scalar x_t and z_t (see Section 4 for explanations, note that $\beta_0 = -1$, $\beta_1 = 1, \gamma_0 = 1$ and $\gamma_1 = 1$ are used in generating the data).

Table A: Bias of CLS and GMM estimators for true values $\beta_0 = -1$ and -2, and $\beta_1 = \gamma_0 = 1$.

True	β_0		β_1		γ_0		γ_1	
γ_1	CLS	GMM	CLS	GMM	CLS	GMM	CLS	GMM
				$\beta_0 = -1$; $T = 50$				
-0.3	0.0524	-0.1325	0.0631	0.1289	0.0327	-0.0907	0.0011	-0.0477
-0.2	0.0572	-0.1007	0.0561	0.1262	0.0334	-0.0760	0.0015	-0.0354
-0.1	0.0717	-0.1894	0.0650	0.1605	0.0422	-0.1301	0.0037	-0.0295
0.0	0.0350	-0.1140	0.0658	0.1041	0.0322	-0.1041	-0.0044	-0.0303
0.1	0.0554	-0.1561	0.0633	0.1070	0.0304	-0.1175	0.0029	-0.0300
0.2	0.0210	-0.1618	0.0584	0.0753	0.0209	-0.1048	-0.0021	-0.0369
0.3	0.0074	-0.1588	0.0553	0.0646	0.0107	-0.1389	-0.0002	-0.0311
				$\beta_0 = -1$; $T = 200$				
-0.3	0.0075	-0.0254	0.0102	0.0198	0.0006	-0.0223	-0.0019	0.0011
-0.2	0.0072	-0.0279	0.0104	0.0216	0.0025	-0.0226	0.0001	0.0034
-0.1	0.0093	-0.0332	0.0090	0.0228	0.0028	-0.0280	-0.0006	0.0025
0.0	0.0041	-0.0351	0.0090	0.0139	0.0025	-0.0311	-0.0008	0.0014
0.1	0.0161	-0.0300	0.0034	0.0141	0.0079	-0.0302	-0.0002	-0.0019
0.2	0.0163	-0.0328	0.0049	0.0134	0.0083	-0.0289	-0.0001	-0.0075
0.3	0.0056	-0.0403	0.0096	0.0253	0.0022	-0.0317	-0.0011	-0.0097
				$\beta_0 = -2$; $T = 50$				
-0.3	0.0140	-0.2454	0.0373	0.1400	0.0341	-0.1134	0.0021	-0.0487
-0.2	0.0447	-0.2343	0.0269	0.1298	0.0513	-0.1085	0.0047	-0.0409
-0.1	0.0542	-0.2402	0.0231	0.1374	0.0522	-0.0996	-0.0006	-0.0401
0.0	0.0327	-0.2443	0.0379	0.1501	0.0555	-0.1012	0.0043	-0.0343
0.1	0.0554	-0.2007	0.0254	0.1032	0.0594	-0.1074	-0.0034	-0.0361
0.2	-0.0007	-0.1628	0.0326	0.0674	0.0333	-0.0901	-0.0044	-0.0414
0.3	-0.0115	-0.1902	0.0364	0.0783	0.0206	-0.0958	-0.0036	-0.0343
				$\beta_0 = -2$; $T = 200$				
-0.3	-0.0022	-0.0361	0.0083	0.0186	0.0002	-0.0209	0.0009	0.0020
-0.2	0.0082	-0.0320	0.0037	0.0155	0.0043	-0.0265	-0.0011	-0.0010
-0.1	0.0066	-0.0383	0.0024	0.0186	0.0051	-0.0305	0.0011	0.0007
0.0	0.0096	-0.0361	-0.0001	0.0158	0.0062	-0.0386	0.0001	-0.0015
0.1	0.0066	-0.0494	0.0022	0.0191	0.0058	-0.0450	-0.0009	-0.0016
0.2	-0.0025	-0.0383	0.0097	0.0197	0.0013	-0.0316	-0.0008	-0.0086
0.3	0.0017	-0.0313	0.0097	0.0139	0.0018	-0.0296	-0.0001	-0.0103

Table B: MSE of CLS and GMM estimators for true values $\beta_0 = -1$ and -2, and $\beta_1 = \gamma_0 = 1$.

True	β_0		β_1		γ_0		γ_1	
γ_1	CLS	GMM	CLS	GMM	CLS	GMM	CLS	GMM
				$\beta_0 = -1$; $T = 50$				
-0.3	0.3301	0.3109	0.1367	0.1551	0.0389	0.0667	0.0071	0.0165
-0.2	0.3036	0.3464	0.1044	0.1942	0.0388	0.0579	0.0068	0.0158
-0.1	0.3506	0.3766	0.1314	1.0306	0.0400	0.5459	0.0060	0.0591
0.0	0.2699	0.3003	0.0856	0.1165	0.0389	0.0627	0.0050	0.0161
0.1	0.2311	0.2242	0.0786	0.1261	0.0403	0.0599	0.0043	0.0169
0.2	0.2129	0.2035	0.0761	0.0850	0.0403	0.0590	0.0040	0.0158
0.3	0.2602	0.1914	0.0934	0.1034	0.0458	0.7893	0.0040	0.0432
				$\beta_0 = -1$; $T = 200$				
-0.3	0.0165	0.0173	0.0100	0.0120	0.0083	0.0116	0.0016	0.0028
-0.2	0.0205	0.0187	0.0104	0.0124	0.0083	0.0109	0.0014	0.0025
-0.1	0.0231	0.0201	0.0120	0.0130	0.0079	0.0099	0.0014	0.0026
0.0	0.0258	0.0218	0.0125	0.0128	0.0084	0.0106	0.0013	0.0028
0.1	0.0270	0.0227	0.0120	0.0142	0.0081	0.0104	0.0011	0.0029
0.2	0.0258	0.0229	0.0124	0.0146	0.0088	0.0101	0.0013	0.0035
0.3	0.0257	0.0257	0.0123	0.0202	0.0092	0.0105	0.0013	0.0036
				$\beta_0 = -2$; $T = 50$				
-0.3	0.4089	0.4191	0.1420	0.1606	0.0534	0.0695	0.0094	0.0190
-0.2	0.4356	0.4083	0.1335	0.1600	0.0524	0.0576	0.0087	0.0187
-0.1	0.4392	0.3681	0.1362	0.1457	0.0560	0.0512	0.0073	0.0176
0.0	0.4835	0.4058	0.1451	0.1639	0.0533	0.0523	0.0053	0.0165
0.1	0.3599	0.3073	0.1086	0.1228	0.0484	0.0551	0.0046	0.0160
0.2	0.3506	0.2250	0.1153	0.0866	0.0447	0.0518	0.0043	0.0165
0.3	0.3017	0.2396	0.1126	0.0976	0.0464	0.0635	0.0046	0.0164
				$\beta_0 = -2$; $T = 200$				
-0.3	0.0185	0.0187	0.0078	0.0086	0.0099	0.0118	0.0019	0.0027
-0.2	0.0207	0.0181	0.0080	0.0085	0.0090	0.0115	0.0017	0.0030
-0.1	0.0235	0.0200	0.0092	0.0094	0.0083	0.0113	0.0015	0.0032
0.0	0.0259	0.0213	0.0096	0.0089	0.0085	0.0113	0.0013	0.0032
0.1	0.0264	0.0242	0.0097	0.0103	0.0082	0.0117	0.0012	0.0032
0.2	0.0264	0.0251	0.0096	0.0110	0.0088	0.0098	0.0013	0.0037
0.3	0.0280	0.0241	0.0110	0.0130	0.0103	0.0097	0.0016	0.0040

Table C: Bias and MSE of weighted CLS estimator for true values $\beta_0 = -1$ and -2, and $\beta_1 = \gamma_0 = 1$.

True	β_0		β_1		γ_0		γ_1	
γ_1	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
				$\beta_0 = -1; T = 50$				
-0.3	0.0682	0.2608	0.0404	0.1077	0.0379	0.0324	0.0048	0.0054
-0.2	0.0772	0.2199	0.0348	0.0675	0.0382	0.0341	0.0037	0.0054
-0.1	0.0872	0.2831	0.0495	0.1000	0.0456	0.0382	0.0050	0.0053
0.0	0.0583	0.2453	0.0528	0.0770	0.0407	0.0370	-0.0037	0.0046
0.1	0.0670	0.2150	0.0549	0.0692	0.0343	0.0388	0.0033	0.0041
0.2	0.0477	0.1796	0.0455	0.0593	0.0301	0.0360	-0.0016	0.0034
0.3	0.0233	0.1790	0.0437	0.0613	0.0164	0.0363	-0.0007	0.0031
				$\beta_0 = -1; T = 200$				
-0.3	0.0132	0.0164	0.0066	0.0089	0.0048	0.0073	-0.0005	0.0013
-0.2	0.0088	0.0185	0.0035	0.0093	0.0025	0.0075	-0.0002	0.0012
-0.1	0.0115	0.0201	0.0050	0.0106	0.0030	0.0069	-0.0011	0.0012
0.0	0.0128	0.0219	0.0065	0.0106	0.0076	0.0071	-0.0002	0.0011
0.1	0.0109	0.0239	0.0067	0.0112	0.0050	0.0075	-0.0016	0.0011
0.2	0.0118	0.0230	0.0010	0.0108	0.0082	0.0075	-0.0010	0.0011
0.3	0.0102	0.0183	0.0085	0.0098	0.0044	0.0069	0.0000	0.0010
				$\beta_0 = -2; T = 50$				
-0.3	0.0427	0.2850	0.0156	0.0911	0.0409	0.0482	0.0063	0.0077
-0.2	0.0730	0.3400	0.0076	0.0972	0.0583	0.0495	0.0073	0.0071
-0.1	0.0729	0.3726	0.0109	0.1108	0.0575	0.0535	0.0002	0.0062
0.0	0.0545	0.3842	0.0244	0.1104	0.0589	0.0507	0.0044	0.0049
0.1	0.0715	0.3205	0.0141	0.0941	0.0617	0.0454	-0.0024	0.0044
0.2	0.0358	0.2441	0.0112	0.0762	0.0406	0.0392	-0.0039	0.0038
0.3	0.0183	0.2358	0.0206	0.0863	0.0299	0.0401	-0.0039	0.0038
				$\beta_0 = -2; T = 200$				
-0.3	-0.0034	0.0164	0.0063	0.0066	-0.0013	0.0081	0.0005	0.0015
-0.2	0.0069	0.0184	0.0028	0.0072	0.0036	0.0077	-0.0008	0.0014
-0.1	0.0094	0.0204	0.0006	0.0082	0.0069	0.0071	0.0018	0.0012
0.0	0.0077	0.0234	0.0007	0.0089	0.0054	0.0076	0.0005	0.0012
0.1	0.0091	0.0229	-0.0006	0.0089	0.0066	0.0071	-0.0014	0.0010
0.2	0.0010	0.0224	0.0066	0.0083	0.0030	0.0075	-0.0007	0.0011
0.3	0.0042	0.0226	0.0051	0.0091	0.0021	0.0081	0.0000	0.0011

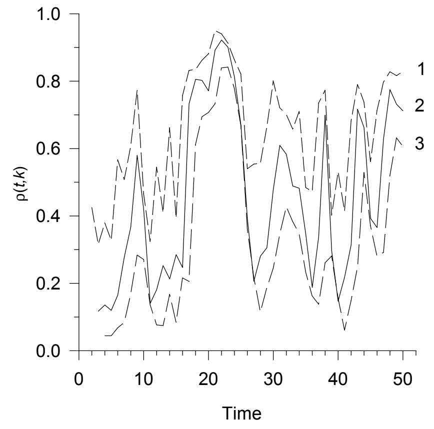

Figure 2: Autocorrelations $\rho(t,k), k = 1, 2, 3$ (see Figure 1 for explanations).

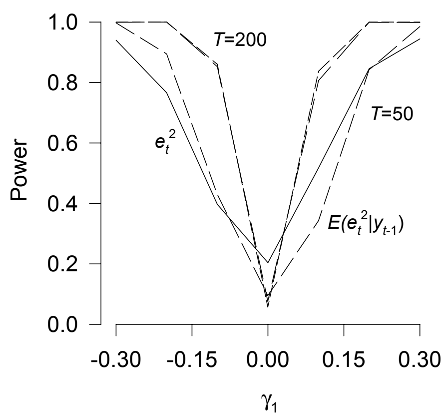

Figure 3: Power of t-test of $\mathrm{H}_0 : \gamma_1 = 0$ based on CLS with corrected covariance matrices based on e_t^2 and $E(e_t^2|y_{t-1})$ ($\beta_0 = -1, \beta_1 = \gamma_0 = 1$; nominal size 0.05; $T = 50, 200$).

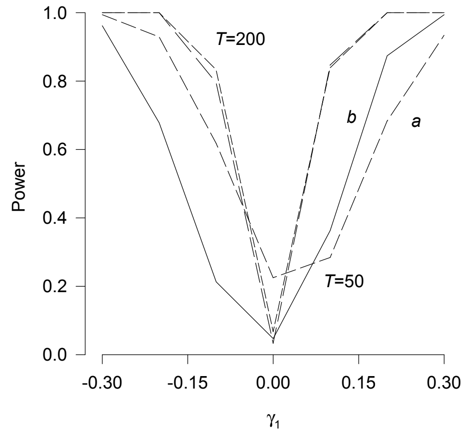

Figure 4: Sizes and powers of Wald test statistics for $\mathrm{H}_0 : \gamma_1 = 0$ based on CLS (a) and weighted CLS (b) when $\beta_1 = -2$.

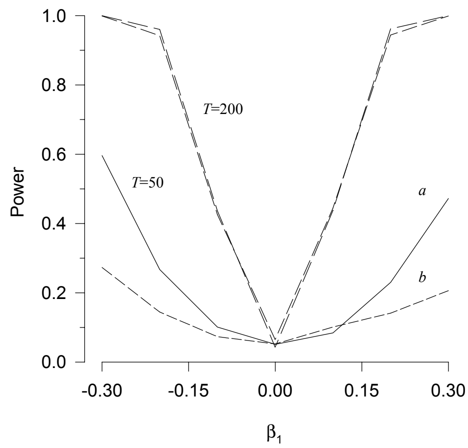

Figure 5: Sizes and powers of time invariance test based on CLS and a Wald test (in a) and on GMM and an LM test (in b) ($\beta_0 = -1, \gamma_0 = 1, \gamma_1 = 0$, nominal size 0.05, $T = 50, 200$).

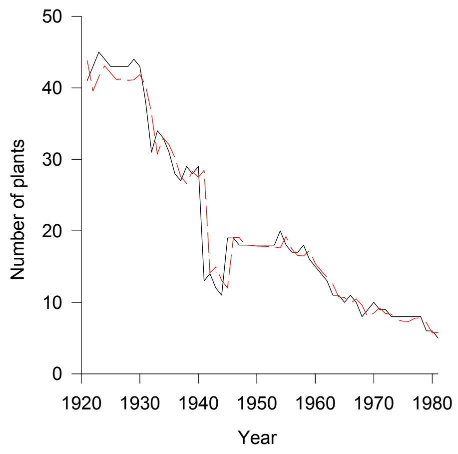

Figure 6: The number of Swedish mechanical paper and pulp plants (solid line) and fitted values (dashed line).

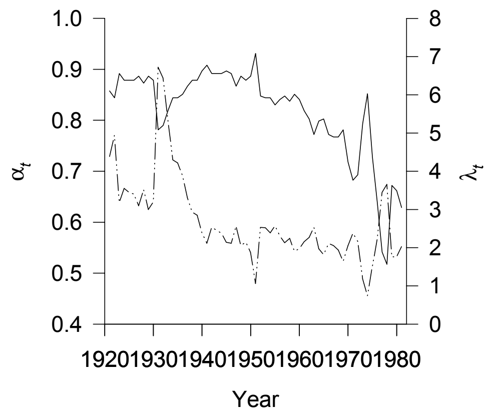

Figure 7: Estimated survival probability (α_t, solid line) and mean entry (λ_t, dash dot dot line).

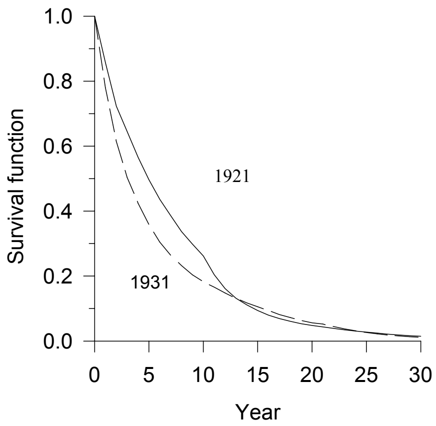

Figure 8: Survival function for new mills starting in 1921 and 1931.

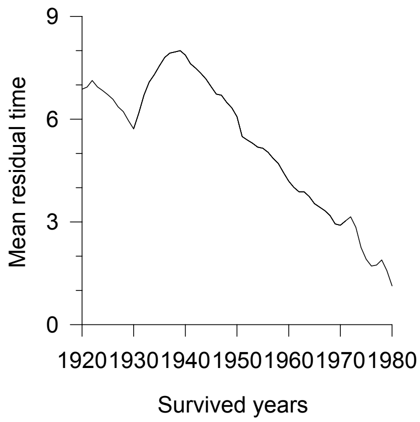

Figure 9: Mean residual life times for new mills starting in 1921.