Integer-Valued Moving Average Modelling of the Number of Transactions in Stocks*

Kurt Brännäs and Shahiduzzaman Quoreshi
Department of Economics & USBE, Umeå University
SE-901 87 Umeå, Sweden

Abstract

The integer-valued moving average model is advanced to model the number of transactions in intra-day data of stocks. The conditional mean and variance properties are discussed and model extensions to include, e.g., explanatory variables are offered. Least squares and generalized method of moment estimators are presented. In a small Monte Carlo study a feasible least squares estimator comes out as the best choice. Empirically we find support for the use of long-lag moving average models in a Swedish stock series. There is evidence of asymmetric effects of news about prices on the number of transactions.

Key Words: Count data, Intra-day, High frequency, Time series, Estimation, Finance.
JEL Classification: C13, C22, C25, C51, G12, G14.

Umeå Economic Studies 637, 2004

1. Introduction

The paper focuses on modelling time series of the number of intra-day transactions in stocks. Such count data time series typically take on small numbers even for frequently traded stocks if the counts are recorded in short time intervals of, for instance, one minute length. There is an obvious connection between the current count data model and the conditional duration model of, e.g., Engle and Russell (1998) in the sense that long durations in a time interval correspond to a small count and vice versa. Hence, a main use of the count data models discussed here is also one of measuring reaction times to shocks or news. In this context straightforward use of the Box-Jenkins methodology for identifying parsimonious time series models raises fundamental questions about the resulting model specification when we wish to adhere to the integer-valued or count data nature of the number of transactions variable.

Previous models for the number of transactions or related variables within the intra-day financial arena have departed from conventional count data regression models or from extended model versions (e.g., Brännäs and Brännäs, 2004; Heinen and Rengifo, 2003). Here, we consider a different approach and start from an integer-valued model corresponding to the conventional ARMA class of Box and Jenkins (1970). An important difference between the continuous variable ARMA model and its corresponding integer-valued version (INARMA) is that the latter contains parameters that are interpreted as probabilities and then take on values in narrower intervals than do the parameters of the ARMA model (e.g., McKenzie, 1986; Al-Osh and Alzaid, 1991; Joe, 1996; Jørgensen and Song, 1998; McKenzie, 2003). This then may make model identification of an appropriate model by correlation methods (Box and Jenkins, 1970) less suitable. While for the ARMA class, specification searches aim at models that leave no serial correlation and satisfy stationarity and invertibility criteria, the INARMA specification should additionally have each parameter estimate of lagged variables in the unit interval.

In this paper the empirical results indicate that long-lag INMA models satisfy such restrictions, while mixed INARMA models do not. The INMA model class has been studied by, e.g., Al-Osh and Alzaid (1988), McKenzie (1988) and Brännäs and Hall (2001). As far as we are aware the only published, empirical application is due to Brännäs, Hellström and Nordström (2002), who estimated a nonlinear INMA(1) model for tourism demand. This model had time dependent parameters that were functions of explanatory variables. The present empirical study is focused on a stock transaction series (Ericsson) registered at the order driven Stockholmsbörsen stock exchange and emphasizes different specification issues.

The INMA model is introduced in Section 2, where we also give some moment results and discuss conditional heteroskedasticity. General expressions for conditional and unconditional moments are obtained. Extensions to include explanatory variables in the conditional mean and a more flexible conditional heteroskedasticity specification are also discussed. Section 3 discusses estimation of unknown parameters and gives least squares and GMM estimators for

1

the model class. The section also considers aspects of model evaluation and forecasting. A small Monte Carlo experiment to study some key characteristics of the estimators and of forecasts is included. Section 4 contains the empirical results for the stocks series. The final section offers some concluding comments.

2. MODEL

The single thing that most visibly makes the integer-valued MA (INMA) model different from its continuous variable MA counterpart is that multiplication of variables with real valued parameters is no longer a viable operation, when the result is to be integer-valued. Multiplication is therefore replaced by the binomial thinning operator

$$\alpha \circ u = \sum_{i=1}^{u} v_i,$$

where $\{v_i\}_{i=1}^{u}$ is an iid sequence of $0 - 1$ random variables, such that $\Pr(v_i = 1) = \alpha = 1 - \Pr(v_i = 0)$. Conditionally on the integer-valued u, $\alpha \circ u$ is binomially distributed with $E(\alpha \circ u|u) = \alpha u$ and $V(\alpha \circ u|u) = \alpha(1 - \alpha)u$. Unconditionally it holds that $E(\alpha \circ u) = \alpha\lambda$, where $E(u) = \lambda$, and $V(\alpha \circ u) = \alpha^2\sigma^2 + \alpha(1 - \alpha)\lambda$, where $V(u) = \sigma^2$. Obviously, $\alpha \circ u \in [0, u]$.

Employing this binomial thinning operator, the INMA(∞) model can be written

$$y_t = \sum_{i=0}^{\infty} \beta_i \circ u_{t-i}, \tag{1}$$

with mostly $\beta_0 = 1$.[1] The $\{u_t\}$ is an iid sequence of non-negative and integer-valued random variables with, as above, $E(u) = \lambda$ and $V(u) = \sigma^2$.

McKenzie (1988), Joe (1996), Jørgensen and Song (1998) and others stress exact distributional results for y_t, while we emphasize only the first two conditional and unconditional moments of the model. One reason for our choice will become apparent below when we discuss and introduce more flexible conditional mean and heteroskedasticity specifications for y_t than implied by (1). As a consequence exact maximum likelihood (ML) estimation is beyond reach though otherwise a desirable candidate for estimation. We could potentially use ML estimation by directly specifying a conditional density for y_t given its history Y_{t-1}; cf. the conditional duration model approach of Engle and Russell (1998).

The finite-lag INMA(q) model

$$y_t = u_t + \beta_1 \circ u_{t-1} + \ldots + \beta_q \circ u_{t-q} \tag{2}$$

was introduced by McKenzie (1986). Brännäs and Hall (2001) discuss model generalizations and interpretations resulting from different thinning operator structures, and an empirical study and approaches to estimation are reported by Brännäs et al. (2002).

---

[1] The INMA(∞) can, e.g., be obtained from the INAR(1) $y_t = \alpha \circ y_{t-1} + \varepsilon_t$ by repeated substitution: $y_t = \alpha^t \circ y_0 + \sum_{i=1}^{t} \alpha^{t-i} \circ \varepsilon_i$. Assuming $\alpha \in [0, 1)$ and t large gives that $\beta_i = \alpha^i$.
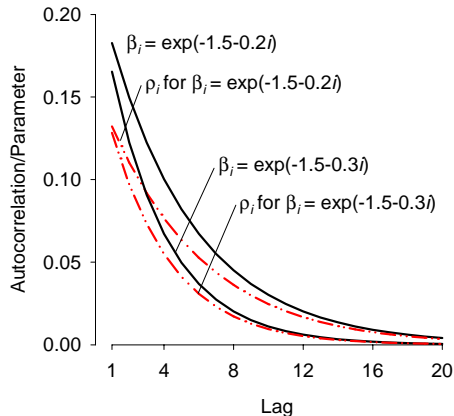
Figure 1: Autocorrelation functions in the Poisson case (dot-dashed lines) and β_i parameters (solid lines) when parameters are as in the Monte Carlo study, below.

For the INMA(∞) model in (1), with independence between and within thinning operations,[2] and with $\{u_t\}$ an iid Poisson sequence with $\sigma^2 = \lambda$, and $\beta_0 = 1$, the moment expressions are:

$$E(y_t) \;=\; V(y_t) = \lambda\left(1 + \sum_{i=1}^{\infty}\beta_i\right) \tag{3a}$$

$$\rho_k \;=\; \lambda\left(\beta_k + \sum_{i=1}^{\infty}\beta_i\beta_{k+i}\right)/V(y_t), \qquad k \geq 1. \tag{3b}$$

It is obvious from these moments that they only generate positive values and that $\sum_{i=0}^{\infty}\beta_i < \infty$ is required for $\{y_t\}$ to be a stationary sequence. Assuming instead, e.g., an iid distribution with mean λ and variance σ^2 changes the variance and the autocorrelation function to

$$V(y_t) \;=\; \lambda\sum_{i=1}^{\infty}\beta_i(1-\beta_i) + \sigma^2\left(1 + \sum_{i=1}^{\infty}\beta_i^2\right) \tag{3c}$$

$$\rho_k \;=\; \sigma^2\left(\beta_k + \sum_{i=1}^{\infty}\beta_i\beta_{k+i}\right)/V(y_t), \qquad k \geq 1, \tag{3d}$$

while leaving the mean unaffected and as in (3a).

In case the lag length q is finite, summing to infinity is replaced by summing to q in (3c)-(3d) and to $q - k$ in the numerator of (3d). In that case $\rho_k = 0$,

---

[2] Pairs of thinning operations of the type $\theta_i \circ u_t$ and $\theta_j \circ u_t$, for $i \neq j$, are independent (McKenzie, 1988). Assumptions of this type can be relaxed (cf. Brännäs and Hall, 2001).

for $k > q$. Figure 1 gives an illustration of two autocorrelation functions when $\beta_i, i \geq 1$, are set as in the Monte Carlo experiment, below.

The conditional moments for the INMA(∞) model are:

$$E(y_t|Y_{t-1}) = \lambda + \sum_{i=1}^{\infty} \beta_i u_{t-i} \tag{4a}$$

$$V(y_t|Y_{t-1}) = \sigma^2 + \sum_{i=1}^{\infty} \beta_i(1 - \beta_i)u_{t-i}, \tag{4b}$$

where Y_{t-1} is the information set available at time $t-1$.

As the conditional variance varies with $u_{t-i}, i \geq 1$, there is conditional heteroskedasticity of a moving average type. We can call this property MACH(∞) and for finite q, MACH(q) (cf. the notation ARCH(p)). Note that the response to lagged u_{t-i}s is weaker in the conditional variance (4b) than in (4a) as $\beta_i \in [0, 1]$. The relative size of the two moments will largely depend on the sizes of λ and σ^2. Note also that even though a Poisson distribution for u_t implies a Poisson distributed y_t, the conditional distribution of y_t given its past is not the Poisson, as indicated, e.g., by the difference between (4a) and (4b).

As measures of reaction times to shocks/news in the $\{u_t\}$ sequence we may use the mean lag $\sum_{i=0}^{\infty} i\beta_i/w$, where $w = \sum_{i=0}^{\infty} \beta_i$. Alternatively, we may use the median lag, which is the smallest k such that $\sum_{i=0}^{k} \beta_i/w \geq 0.5$.

2.1 Extensions

There are two obvious extensions to the model that appear of high empirical relevance. First, we may let λ become time-varying and a function of explanatory variables. The natural specification is

$$\lambda_t = \exp(\mathbf{x}_t\boldsymbol{\theta}) \geq 0, \tag{5}$$

where in $\mathbf{x}_t$ we may include k variables related to previous prices, etc. A consequence of the time-varying λ_t is that moment expressions become time dependent, but the additional difficulty with respect to estimation is marginal.

Another obvious extension in order to obtain more flexible conditional variance specifications in (4b) is to let σ^2 become time dependent. We may let σ_t^2 depend on past values on σ_t^2, u_t and explanatory variables in, e.g., the following exponential way

$$\sigma_t^2 = \exp\Big[\phi_0 + \phi_1 \ln \sigma_{t-1}^2 + \ldots + \phi_P \ln \sigma_{t-P}^2$$
$$+ \gamma_1(u_{t-1} - \lambda)^2 + \ldots + \gamma_Q(u_{t-Q} - \lambda)^2 + \mathbf{x}_t\boldsymbol{\alpha}\Big] \tag{6}$$

(cf. Nelson, 1991). There could also be additional contributions, or at least different ones, to the conditional variance if, e.g., the different thinning operations were dependent (cf. Brännäs and Hall, 2001; Brännäs and Hellström, 2001).

It is also possible to let β_i be, e.g., a logistic function of explanatory variables (Brännäs et al., 2002) and to reduce the number of β_is by specifying a

distributed lag distribution of, e.g., the form $\beta_i = \delta_0 \exp(-\delta_1 i)$ with $\delta_0 \in (0, 1]$ and $\delta_1 \geq 0$ for $i \geq 1$.

3. ESTIMATION

In this section we discuss approaches to the estimation of the unknown parameters of the conditional mean and variance functions. Both the conditional mean and the conditional variance will contain time dependent λ_t and σ_t^2 specifications, respectively, unless otherwise stated. As we do not assume a full density specification the proposed estimators can be viewed as semiparametric ones.

The conditional least squares (CLS) estimator is focused on the unknown parameters of the conditional mean function and requires an additional step to estimate the unknown parameters of the conditional variance expression, in particular in its σ_t^2 part. Typically, CLS estimates, of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)'$ and in case λ_t is time invariant λ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$ when λ_t is time-varying, are used and kept fixed when estimating the variance function. In a feasible generalized least squares (FGLS) estimator, these two steps facilitate GLS estimation of the conditional mean function in a final step. Note that for small β_i parameters we may expect the conditional variance to be almost constant if $\sigma_t^2 = \sigma^2$ holds (cf. (4b)). In such an instance we expect the CLS and FGLS estimators of the parameters to be numerically close.

By a GMM estimator (Hansen, 1982), all parameters can be estimated jointly. For the GMM estimator weighting is with respect to moment conditions and not with respect to individual observations as in FGLS. We may anticipate better performance of the FGLS than of the GMM estimator for the parameters of the conditional mean function (Brännäs, 1995). Brännäs and Hall (2001) found the CLS estimator to have weaker bias/mean square error (MSE) performance than a GMM estimator based on probability generating functions. This type of GMM estimator will not be considered here as it is computationally more intricate and currently rests on arbitrarily setting values on the argument of the generating function.

Common to the considered estimators is their reliance on the prediction error

$$e_{1t} = y_t - E(y_t | Y_{t-1}). \tag{7}$$

Moment conditions are created with e_{1t} as the residual and instruments depending on the particular model specification. This corresponds to the normal equations of the CLS estimator, whose solution also makes up the first step of the FGLS estimator. Alternatively, we may use the properties $E(e_{1t}) = 0$ and $E(e_{1t} e_{1t-i}) = 0, i \geq 1$. The CLS estimator minimizes the criterion function $S_{CLS} = \sum_{t=r}^{T} e_{1t}^2$, where $r = q + 1$ and T is the time series length, with respect to the unknown parameter vector $\boldsymbol{\psi}' = (\lambda, \boldsymbol{\beta}')$ or $\boldsymbol{\psi}' = (\boldsymbol{\theta}', \boldsymbol{\beta}')$. To calculate the e_{1t} sequence we consider

$$e_{1t} = u_t - \lambda_t + \sum_{i=1}^{\infty} (\beta_i \circ u_{t-i} - \beta_i u_{t-i}) \tag{8}$$

5

and advocate using the expected value zero of the final term in (8) rather than using some randomization device.

Obviously, there are alternative estimators, such as Durbin's (1959) estimator extended to handle λ, for the estimation of $\boldsymbol{\psi}$. A recent summary of least squares and related estimators (interpreted as GMM) of low order MA(q) models utilizing dual AR representations is given in Harris (1999). Method of moment (or GMM) estimation based on the unconditional first and second moments requires the solution of a system of nonlinear equations. Hence, a simplicity argument does not apply and moreover the properties of the estimator are in the MA(q) model not very satisfactory. When standard software based on the assumption $E(u_t) = \lambda = 0$ is used to CLS estimate an INMA(q) model, the obtained estimate of the constant term is an estimate of $\lambda \sum_{i=0}^{q} \beta_i$. Given estimates of β_i it is hence possible to obtain an estimate of λ manually.

For the second step, normal equations based on the conditional variance prediction error

$$e_{2t} = (y_t - E(y_t|Y_{t-1}))^2 - V(y_t|Y_{t-1}) \tag{9}$$

are used for FGLS estimation, and incorporated as moment conditions for GMM estimation. Here too the conditional variance specification should have a say on the choice of instrumental variables.

For FGLS $S_2 = \sum_{t=s}^{T} e_{2t}^2$, where $s = \max(q, P, Q) + 1$, is minimized with respect to the parameters of the σ_t^2 function, i.e. $\boldsymbol{\omega}' = (\sigma^2, \phi_0, \ldots, \phi_P, \gamma_1, \ldots, \gamma_Q, \boldsymbol{\alpha}')$ and with the CLS estimates $\hat{\boldsymbol{\psi}}$ and $\{\hat{u}_t\}$ kept fixed. In case σ^2 is time invariant an obvious estimator is of the simple form

$$\hat{\sigma}^2 = (T - s)^{-1} \sum_{t=s}^{T} \left[\hat{e}_{1t}^2 - \sum_{i=1}^{q} \hat{\beta}_i (1 - \hat{\beta}_i) \hat{u}_{t-i} \right].$$

For the third step of FGLS, minimizing the criterion

$$S_{FGLS} = \sum_{t=s}^{T} e_{1t}^2 / \hat{V}(y_t|Y_{t-1})$$

with $\hat{V}$ taken as given, gives the FGLS estimates of the parameter vector $\boldsymbol{\psi}$ of the conditional mean function. The covariance matrix estimators are:

$$Cov(\hat{\boldsymbol{\psi}}_{CLS}) = \left(\sum_{t=r}^{T} \frac{\partial e_{1t}}{\partial \boldsymbol{\psi}} \frac{\partial e_{1t}}{\partial \boldsymbol{\psi}'} \right)^{-1} = \mathbf{A}^{-1}$$

$$Cov(\hat{\boldsymbol{\psi}}_{FGLS}) = \left(\sum_{t=r}^{T} \hat{V}^{-1}(y_t|Y_{t-1}) \frac{\partial^2 e_{1t}}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} \right)^{-1} = \mathbf{B}^{-1}.$$

A robust estimator for the CLS estimator is of the form $\mathbf{A}^{-1} \mathbf{B}_c \mathbf{A}^{-1}$, where $\mathbf{B}_c$ is as $\mathbf{B}$ with $\hat{V}^{-1}(y_t|Y_{t-1})$ replaced by $\hat{e}_{1t}^2$.

Using a finite maximum lag q for a true INMA(∞) model can be expected to have a biasing effect on the estimator of the constant term λ of the conditional mean function of the INMA(q) model. The conditional expectation of

6

the difference between the infinite and finite models is $\sum_{i=q+1}^{\infty} \beta_i u_{t-i}$, which has expectation $\lambda \sum_{i=q+1}^{\infty} \beta_i$. For large q the latter sum will be close to zero as we expect $\beta_i, i \geq q+1$, to approach zero for large q. An estimate of the constant term can then be expected to be only moderately too large. In analogy with a linear regression model and the OLS estimator the conventional analysis of the consequences of omitted variables, i.e. $u_{t-q-1}, \ldots, u_{t-\infty}$, suggests that we should expect a positive bias in the estimates of β_i for small q. This is so, since all $\beta_i > 0$ and the covariance between included and incorrectly excluded lags of u_t is always λ^2 as $E(u_t u_{t-j}) = \lambda^2, j \neq 0$, under independence. By a related argument, we expect no additional bias if q is chosen larger than some true value q^*. An immediate consequence of these intuitive arguments is that q should, at least, initially be chosen large. Subsequent testing could later be used to reduce the initial q.

The GMM criterion function

$$q = \mathbf{m}' \hat{\mathbf{W}}^{-1} \mathbf{m}$$

has the vector of moment conditions $\mathbf{m}$ depending on the specification and is minimized with respect to $\boldsymbol{\eta}' = (\boldsymbol{\psi}', \boldsymbol{\omega}')$. The moment conditions corresponding to the conditional mean are collected into

$$\mathbf{m}_1 = (T - m)^{-1} \sum_{t=m+1}^{T} \mathbf{m}_{1t},$$

where $m \geq q$. In the Monte Carlo study below we use $e_{1t}, e_{1t}e_{1t-1}, \ldots, e_{1t}e_{1t-m}$ for $\mathbf{m}_1$, while for the empirical results we use the first order condition of the CLS estimator, i.e. $e_{1t} \partial e_{1t} / \partial \psi_k = 0$, to give the conditions.

For the conditional variance the moment conditions are formed from $E(e_{1t}^2) - \sigma_t^2 - \sum_{i=1}^{q} \beta_i (1 - \beta_i) u_{t-i}$ with instruments from the collection of $\sigma_{t-i}^2, e_{1t-i}^2$ and x_{kt}. The conditions are collected into a vector $\mathbf{m}_2$. Finally, $\mathbf{m} = (\mathbf{m}_1', \mathbf{m}_2')'$. As a consistent estimator of the weight matrix $\mathbf{W}$ we use

$$\hat{\boldsymbol{\Gamma}} = (T - m)^{-1} \sum_{t=m+1}^{T} \mathbf{m}_t \mathbf{m}_t'.$$

The covariance matrix of the parameter estimator is, when $\mathbf{W}$ is set equal to an identity matrix, estimated by $Cov(\hat{\boldsymbol{\eta}}) = (T-m)^{-1}(\hat{\mathbf{G}}'\hat{\mathbf{G}})^{-1}\hat{\mathbf{G}}'\hat{\mathbf{T}}\hat{\mathbf{G}}(\hat{\mathbf{G}}'\hat{\mathbf{G}})^{-1}$, where $\hat{\mathbf{G}} = \partial \hat{\mathbf{m}} / \partial \hat{\boldsymbol{\eta}}'$. When the numbers of moment conditions and parameters are equal; $Cov(\hat{\boldsymbol{\eta}}) = (T-m)^{-1}\hat{\mathbf{G}}^{-1}\hat{\boldsymbol{\Gamma}}(\hat{\mathbf{G}}')^{-1}$.

3.1   Model Evaluation

To test against serial correlation in standardized residuals $\hat{e}_{1t} / \hat{V}^{1/2}(y_t | Y_{t-1})$ and squared standardized residuals we may use the Ljung-Box test statistic $LB_K = T(T+2) \sum_{i=1}^{K} r_i^2 / (T - i)$, where r_i is the lag i autocorrelation of the

standardized residual. Under homoskedasticity and independence the test statistic is asymptotically $\chi^2(K)$ distributed. Davis and Dunsmuir (2000) and Lobato, Nankervis and Savin (2001) recently considered corrections to the Ljung-Box statistic when heteroskedasticity and serial correlation are present.

<div align="center">

3.2   Forecasting

</div>

For an INMA(∞) model the forecasts $\mu_{T+h|T} = E(y_{T+h}|Y_T), h \geq 1$, are

$$\mu_{T+h|T} = \lambda \sum_{i=0}^{h-1} \beta_i + \sum_{i=h}^{\infty} \beta_i u_{T+h-i}, \qquad h \geq 1.$$

The limiting value of the forecast as $h \to \infty$ is $\lambda \sum_{i=0}^{\infty} \beta_i$, which is the mean of the process. For finite q the sum to infinity in the second term is replaced by summing up to q for $h \leq q$. For $h > q$, the forecast is again equal to the mean of the process, i.e. $\lambda \sum_{i=0}^{q} \beta_i$.

The variance of the forecast error $e_{T+h} = y_{T+h} - \mu_{T+h|T}$ is for the INMA(∞) model with known parameters

$$s_{T+h|T} = \sigma^2 \sum_{i=0}^{h-1} \beta_i^2 + \lambda \sum_{i=1}^{\infty} \beta_i(1 - \beta_i).$$

For $h \leq q$, the forecast error variance for the INMA(q) model is $s_{T+h|T} = \sigma^2 \sum_{i=0}^{h-1} \beta_i^2 + \lambda \sum_{i=1}^{q} \beta_i(1 - \beta_i)$.

Obviously, the uncertainty in estimated parameters will increase these variances. To obtain expressions for such variances we could consider various approximations to $V(e_{T+h}) = E_{\hat{\eta}}(s_{T+h|T,\hat{\eta}}) + V_{\hat{\eta}}(\mu_{T+h|T,\hat{\eta}})$. Using first order Taylor expansions of the two terms around the true parameter vector $\boldsymbol{\eta}$ we get the approximative variance

$$V(e_{T+h}) \approx s_{T+h|h} + \mathbf{g}' \left[E(\hat{\boldsymbol{\eta}}) - \boldsymbol{\eta} \right] + \mathbf{h}' Cov(\hat{\boldsymbol{\eta}}) \mathbf{h},$$

where $\mathbf{g} = \partial s_{T+h|h}/\partial \boldsymbol{\eta}$ and $\mathbf{h} = \partial \mu_{T+h|h}/\partial \boldsymbol{\eta}$. Most often a consistent estimator $(E(\hat{\boldsymbol{\eta}}) - \boldsymbol{\eta} = \mathbf{0})$ and a large sample $(Cov(\hat{\boldsymbol{\eta}}) \approx \mathbf{0})$ are assumed in which case $s_{T+h|T}$ evaluated at estimates is the expression to employ.

<div align="center">

3.3   A Small Monte Carlo Experiment

</div>

In this small Monte Carlo experiment we study the bias and MSE properties of the CLS, FGLS and GMM estimators for finite-lag specifications, when data is generated according to an infinite-lag INMA model. In addition, we study the serial correlation properties of estimated models by the Ljung-Box statistic as well as the properties of forecasts one and two steps ahead.

The data generating process is as in (1), with $\beta_i = \exp(\gamma_0 + \gamma_1 i), i \geq 1$, and $\beta_0 = 1$. The $\{u_t\}$ sequence is generated as Poisson with parameter λ, so

<div align="center">

8

</div>

that $\sigma^2 = \lambda$ is time invariant in the conditional variance (4b).[3] We set $\lambda = 5$, $\gamma_0 = -1.5, \gamma_1 = -0.1, -0.2, -0.3$ and -0.4, and $T = 1000$ and 10 000. The number of replications is 1000 in each design point. In generating the data the first 50 observations are discarded, which appears safe as β_i is effectively zero at lag 50 for the used γ_0 and γ_1 combinations.

For the estimators we choose $q = 10, 20$ and 30. By the q choices we will effectively study under as well as overparameterized model version.[4] We use a simplex algorithm to minimize the criterion function of each estimator. For the GMM estimator we set $\mathbf{W}$ equal to the identity matrix. The Ljung-Box statistic is based on 10 autocorrelations, and the forecast horizon is $h = 2$.

We report bias and MSE results for β_i after accumulation over $i = 1, \ldots, 10$ for all employed q values. The full results are summarized in Table A1 of the Appendix, while Figure 2 contains the results for the CLS estimator. Starting with the CLS estimator we find that as sample size increases there is a decline in MSE throughout. For the bias there is a decline only for the most overparameterized case of $q = 30$. Biases and MSEs drop as q increases. Biases are negative for $q = 10$ and 20, with the exception of the $\gamma_1 = -0.1$ case for $q = 10$. This is the most underparameterized model. For $q = 10$ the absolute bias drops from 14.6 percent (or on average 1.46 percent for the individual parameter) for $\gamma_1 = -0.1$ to 4.4 percent and less for $\gamma_1 \leq -0.2$. The FGLS and GMM estimators also focus on the σ^2 parameter, albeit in different ways. The biases and MSEs of the GMM estimator are in most cases the poorest. The FGLS estimator has smaller biases and MSEs than the CLS estimator for $q \leq 20$ and all estimators are quite close for $q = 30$. Setting $\sigma^2 = \lambda$ and letting all other parts of the conditional variance be known improves on the performance of the GMM estimator though not by much. Note that the use of a weight matrix different from the identity matrix may change these outcomes.

In some additional experiments with $e_{1t}\partial e_{1t}/\partial \psi_k = 0$ instead of $e_{1t}e_{1t-k}$ in forming the $\mathbf{m}_1$ part of the moment conditions for the GMM estimator it was found that the MSE is larger for $\gamma_1 = -0.1$ but smaller for other γ_1-values, in particular for $T = 10\,000$. The bias is larger is larger for $\gamma_1 = -0.1$ and -0.2 and smaller for the other γ_1-values. For $q = 30$ the large number of parameters makes this a very slow estimator, at least, for simulations.

In summary, the FGLS estimator comes out as the best estimator of $\sum_{i=1}^{10} \beta_i$. However, the CLS estimator which is the simplest to use of the three considered estimators is not far behind. It is also clear that q should be chosen large.

The ability of the Ljung-Box statistic to detect remaining serial correlation is studied by counting the number of replications exceeding a critical value of $\chi^2_{0.95}(10) = 18.307$. The results are summarized in Table A2. For $q = 10$ and 30 and $T = 10\,000$ the statistic indicates remaining serial correlation in a large fraction of instances. For $q = 20$ indications are strong only for $\gamma_1 \leq$

---

[3] The experiments are performed using Fortran codes. Poisson random deviates are generated by the POIDEV function (Press et al., 1992), while the binomial thinning is performed by the BNLDEV function.

[4] $\sum_{i=0}^{10} \beta_i = 2.34$ and $\beta_k < 0.01$ for $k \geq 32$ for $\gamma_1 = -0.1$, the sum is 1.87 for $k \geq 16$ and $\gamma_1 = -0.2, 1.61$ for $k \geq 11$ and $\gamma_1 = -0.3$, and 1.45 for $k \geq 8$ and $\gamma_1 = -0.4$.
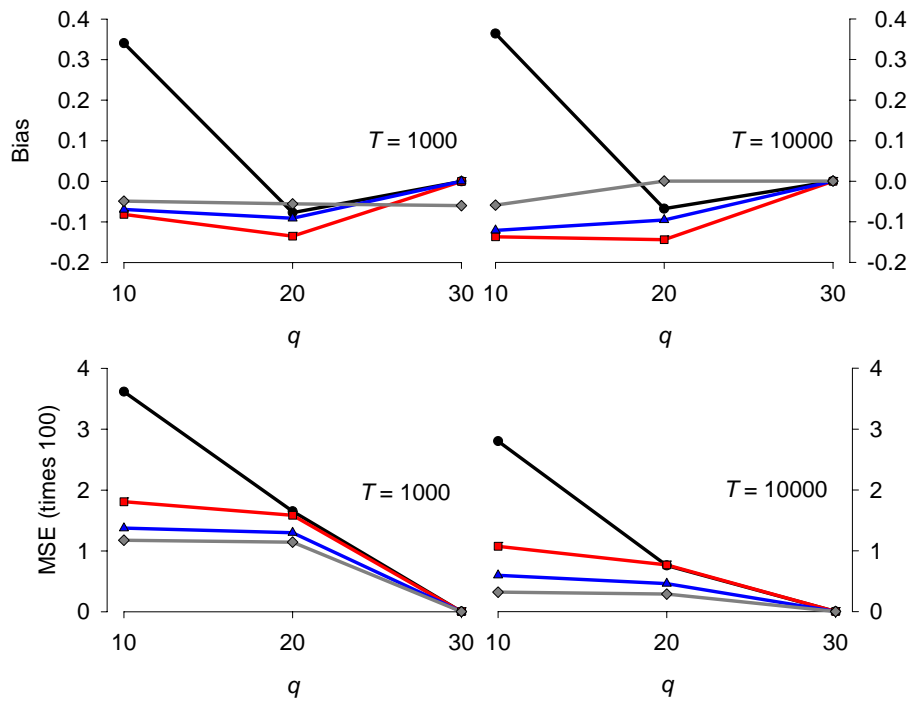
Figure 2: Bias and MSE properties in the Monte Carlo experiment of the CLS estimator of $\sum_{i=1}^{10} \beta_i$ for $T = 1000$ and $10\,000$ and different values of the lag length q in the adopted INMA models. The circle marker indicates that $\gamma_1 = -0.1$ is used in generating the data, square ($\gamma_1 = -0.2$), triangle ($\gamma_1 = -0.3$) and diamond ($\gamma_1 = -0.4$).
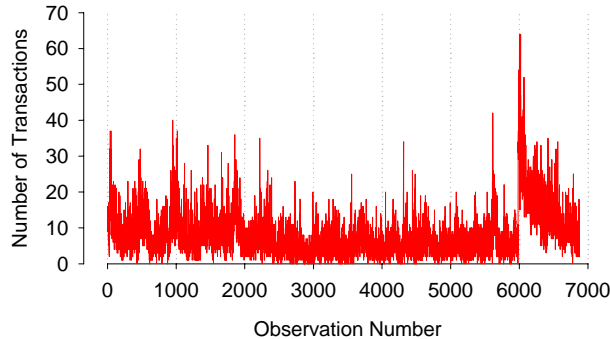
Figure 3: Time series plots for Ericsson.

-0.2. Hence, both under and overparameterization give rise to detectable serial correlation.

The forecasting performance is reported in Table A3. Both in terms of bias (or mean error) and MSE the FGLS estimator performs better than the other two estimators. In addition, the GMM performance is weaker than that of the CLS estimator. For $q = 30$ the differences between the performances of estimators are small. If the alternative GMM estimator is used the performance improves and brings GMM close to FGLS.

4. DATA AND DESCRIPTIVES

The time series for the number of transactions per minute in Ericsson B, in the period July 2 – 22, 2002, are displayed in Figure 3. There are frequent zero frequencies and hence the use of a count data modelling approach is called for. In producing the graph and for the analyses of this paper we have deleted all trading before 0935 (trading opens at 0930) and after 1714 (order book is closed at 1720). The reason for these deletions is that our main focus is on ordinary transactions and the first few minutes are likely to be subject to a different mechanism with considerably higher trading frequencies. The final minutes of the trading day have practically no trading. The basic data were downloaded from the Ecovision system and later filtered by the authors. Due to a technical problem in capturing transactions the first captured minute of July 19 is 0959. There are altogether 6875 observations for the Ericsson series. Descriptive statistics and a histogram for the series are given in Figure 4.

Autocorrelation functions for the series and its first difference are given in Figure 5. For the level series, the function indicates long memory. The autocorrelations after lag one of the first difference series are practically zero. The partial autocorrelation functions die out gradually for both the level and the difference series. Taken together the functions for the first difference series sig-

11
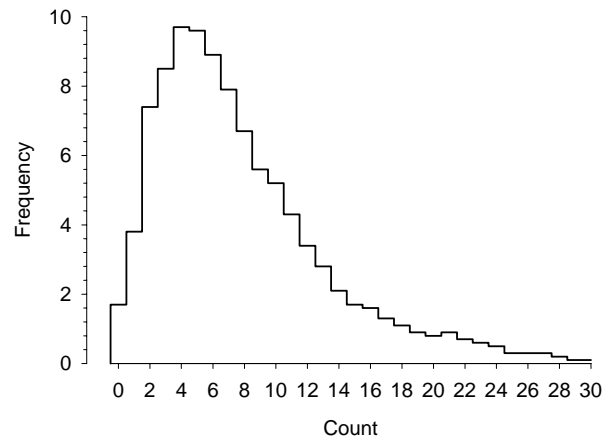
Figure 4: Histogram (in percent) for Ericsson (mean 8.13, variance 40.4, maximum 64).
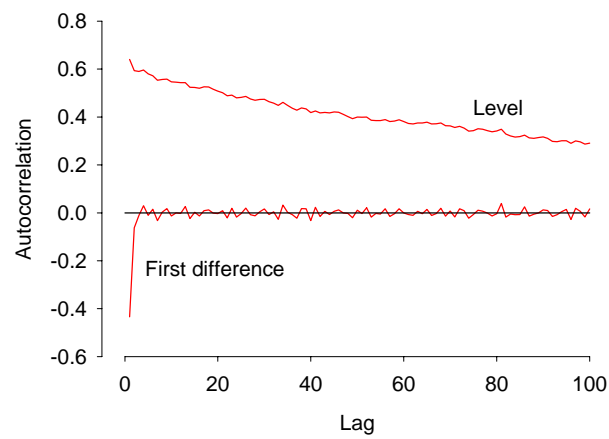
Figure 5: Autocorrelation functions from lag one for the Ericsson series and its first differences.

12

Table 1: ARMA estimation results (standard errors in parentheses) for the Ericsson series.

$$\hat{y}_t = \underset{(0.002)}{0.991}\,\hat{y}_{t-1} + \underset{(0.870)}{8.179} - \underset{(0.012)}{0.757}\,\hat{u}_{t-1} - \underset{(0.012)}{0.088}\,\hat{u}_{t-2}$$

$$s^2 = 18.25, LB_{20} = 28.16\ (p = 0.11)$$

$$\nabla\hat{y}_t = \underset{(0.014)}{0.110}\,\nabla\hat{y}_{t-1} - \underset{(0.007)}{0.001} - \underset{(0.007)}{0.874}\,\hat{u}_{t-1}$$

$$s^2 = 18.32, LB_{20} = 29.79\ (p = 0.07)$$

Note: Models estimated in SPSS.

nal that a model for such a series should include a MA(1) component. The autocorrelations for the level series suggest that a low order AR-part is required together with a low order MA-part. With respect to the mean pattern over the day there is more trading during the first two hours than later.

When specifying and estimating INARMA models according to the conventional Box-Jenkins methodology, in both level and first difference forms, it is obvious that, at least, the $\hat{\beta}_1$ estimates are large and have negative signs and that the INAR(1) parameter is suspiciously close to one (cf. Table 1).[5] Hence there is a violation of the probability interpretation of β_i. It also appears that pure but higher order INAR models are not successful in eliminating serial correlation. Therefore, there is empirical justification for considering INMA models with long lags even though such models are less parsimoniously parameterized.

For the purpose of evaluating the impact of explanatory variables on λ_t and σ_t^2 we also have access to price, volume and spread. These are recorded at the last transaction of the previous minute, i.e. at $t-1$. In case there is no transaction in the previous minute(s), we use the most recent previous notations. Instead of including variables in levels we use first differences to reflect news: $\nabla p_t = p_{t-1} - p_{t-2}$ reflects the news (innovation when p_t follows a random walk) value in the price, $\nabla v_t = v_{t-1} - v_{t-2}$ in volume and $\nabla s_t = s_{t-1} - s_{t-2}$ in the spread. For volume we divide through by 10 000 000.

5.   Empirical Results

The empirical results are presented in terms of finite-lag INMA(q) models. Estimation is carried out by the conditional and feasible generalized least squares (CLS(FGLS)) estimators for the standard INMA(q) model as well as when a time-varying λ_t is allowed for. GMM estimation is employed for full models including conditional heteroskedasticity and time-varying λ_t specifications.

AIC and SBIC criteria are used to find the lag length q, allowing for no

---

[5]Note that for a count data INAR(1) model with a unit root the observed sequence of observations can not decline. Adding a MA part to the INAR(1) does not alter this feature. As is obvious from Figure 3 there are ups and downs in the present time series, so that a unit root can not logically be supported by the data.

Table 2: Estimation results for INMA(q) models for Ericsson (s.e. times 100).

	$q = 47$						$q = 50$				
Lag	$\hat{\beta}_i$	s.e.	Lag	$\hat{\beta}_i$	s.e.	Lag	$\hat{\beta}_i$	s.e.	Lag	$\hat{\beta}_i$	s.e.
1	.2407	.28	26	.1146	.30	1	.2378	.28	26	.1102	.30
2	.1509	.29	27	.0919	.30	2	.1466	.29	27	.0899	.30
3	.1539	.29	28	.0917	.30	3	.1520	.29	28	.0891	.30
4	.1776	.29	29	.1025	.30	4	.1734	.29	29	.0991	.30
5	.1545	.30	30	.1135	.30	5	.1498	.30	30	.1114	.30
6	.1559	.30	31	.0905	.30	6	.1512	.30	31	.0869	.30
7	.1175	.30	32	.0901	.30	7	.1141	.30	32	.0863	.30
8	.1310	.30	33	.0737	.30	8	.1260	.30	33	.0696	.30
9	.1414	.30	34	.1060	.30	9	.1382	.30	34	.1031	.30
10	.1365	.30	35	.0874	.30	10	.1309	.30	35	.0855	.30
11	.1377	.30	36	.0683	.30	11	.1321	.30	36	.0676	.30
12	.1358	.30	37	.0476	.30	12	.1320	.30	37	.0494	.30
13	.1457	.30	38	.0669	.30	13	.1426	.30	38	.0665	.30
14	.1120	.30	39	.0775	.30	14	.1090	.30	39	.0780	.30
15	.1200	.30	40	.0407	.30	15	.1139	.30	40	.0380	.30
16	.1203	.30	41	.0578	.30	16	.1147	.30	41	.0545	.30
17	.1333	.30	42	.0313	.30	17	.1269	.30	42	.0293	.30
18	.1470	.30	43	.0412	.29	18	.1412	.30	43	.0395	.30
19	.1393	.30	44	.0368	.29	19	.1347	.30	44	.0369	.30
20	.1252	.30	45	.0528	.29	20	.1200	.31	45	.0540	.29
21	.1245	.30	46	.0583	.29	21	.1198	.30	46	.0633	.29
22	.0990	.30	47	.0289	.28	22	.0938	.30	47	.0324	.29
23	.1127	.30				23	.1090	.30	48	.0200	.29
24	.0936	.30				24	.0899	.30	49	.0029	.29
25	.1020	.30				25	.0974	.30	50	.0179	.28
λ	1.341	1.67					1.362	1.72			
LB_{20}	14.37						14.80				
LB_{20}^2	81.74						84.06				

Note: The Ljung-Box statistics LB_{20} and LB_{20}^2 are obtained with $\hat{\sigma}^2 = \hat{\lambda}$.

gaps in the β_i sequence. For Ericsson AIC is minimized at $q = 50$, while SBIC indicates an order $q = 47$. Both criterion functions are quite flat indicating some uncertainty with respect to a true q-value.[6] CLS estimation results for pure INMA models are presented in Table 2. Though differences between estimates are quite small, the smaller q-value gives estimates that are larger for low lags and smaller for large lags. The standard errors of the estimates (based on a numerical derivative version of $\mathbf{A}^{-1}$) are throughout small and for both q-values individual hypotheses of $\beta_i = 0, i = 1, \ldots, q$, are rejected throughout. Note also that $\hat{\beta}_i$ estimates are larger than zero throughout even if an unrestricted estimator is used. The confidence intervals at the same lag overlap in most instances. For $q = 47$, the $\hat{\beta}_i$ estimates give a mean lag of 15.8 minutes and a median lag of 14 minutes, while for $q = 50$ the mean lag is 16.0 minutes and the median 14 minutes. Hence, for the measurement of reaction time the q-choice does not matter much in this case.

For both models $R^2 = 0.54$, while the fit of models containing an INAR(1) parameter (cf. Table 1) is better than for pure INMA models. Note also that there are no strong correlations between estimates in this case. Table 2 suggests that the $\hat{\beta}_i$s are roughly linear in i. A linear regression gives $\tilde{\beta}_i = 0.175 - 0.0029i$ ($R^2 = 0.84$) for $q = 47$ and $\tilde{\beta}_i = 0.174 - 0.0029i$ ($R^2 = 0.85$) for $q = 50$. The goodness-of-fit improves further if $\hat{\beta}_1$ is dropped for these regressions.

There is no remaining serial correlation in the standardized residual, for either q-value, when for the conditional variance we use that of the Poisson distributed INMA(q), i.e. we set $\hat{\sigma}^2 = \hat{\lambda}$. However, the squared standardized residuals indicate remaining conditional heteroskedasticity for both models. The largest autocorrelation coefficient is 0.057 for the squared and -0.018 for the standardized residual. As a first step of attempting to find a remedy for the squared residual problem, a time invariant σ^2 is estimated in a second step of the FGLS estimator. For both q-values the $\hat{\sigma}^2$s are substantially larger than the corresponding $\hat{\lambda}$ estimates. Using $\hat{\sigma}^2$ instead of $\hat{\lambda}$ gives roughly the same LB_{20} statistic, while for the squared standardized residual values increase substantially. Estimating models with $\beta_i = \kappa_0 + \kappa_1 i$, where κ_0 and κ_1 are the unknown parameters, leads to more severe serial correlation.

Next, we consider the impact of explanatory variables on λ_t and on σ_t^2. Table 3 reports FGLS and CLS estimates for INMA(q) with λ_t specifications for $q = 50$. The FGLS estimates are obtained by using the CLS estimates $\hat{e}_{1t}$ and $\sum_{i=1}^q \hat{\beta}_i(1 - \hat{\beta}_i)\hat{u}_{t-i}$ to estimate a σ_t^2 model. Corresponding to the $\hat{\lambda}_t$ specification we have $\hat{\sigma}_t^2 = \exp(1.17 + 0.03 \ln \hat{\sigma}_{t-1}^2 - 3.70\nabla p_t + 0.08\nabla p_t^+ + 8.58\nabla s_t)$, $R^2 = 0.10$. This suggests that a negative price change increases volatility while a positive change reduces volatility. A negative spread change lowers volatility while a widening spread increases volatility. A dummy variable for trading before 1101 AM had no significant effect even though trading is more frequent in the early hours of the day.

On comparison with Table 2, the CLS $\hat{\beta}_i$ estimates are marginally smaller for

---

[6] In some experimentation with an AstraZeneca series lower order model representations ($q = 18$ and 30) are found.

Table 3: FGLS and CLS estimation results for INMA(50) with exponential λ_t models for Ericsson (s.e. times 100). For the σ_t^2 specification, see the text.

	FGLS						CLS				
Lag	$\hat{\beta}_i$	s.e.	Lag	$\hat{\beta}_i$	s.e.	Lag	$\hat{\beta}_i$	s.e.	Lag	$\hat{\beta}_i$	s.e.
1	.2257	.80	26	.1035	.84	1	.2312	.28	26	.1069	.30
2	.1481	.82	27	.0839	.85	2	.1412	.29	27	.0854	.30
3	.1466	.82	28	.0868	.85	3	.1511	.29	28	.0853	.30
4	.1641	.83	29	.0941	.85	4	.1701	.29	29	.0943	.30
5	.1502	.83	30	.1110	.85	5	.1463	.30	30	.1078	.30
6	.1468	.83	31	.0914	.85	6	.1462	.30	31	.0845	.30
7	.1124	.84	32	.0844	.85	7	.1116	.30	32	.0826	.30
8	.1207	.84	33	.0631	.85	8	.1215	.30	33	.0666	.30
9	.1301	.84	34	.0993	.85	9	.1295	.30	34	.1004	.30
10	.1301	.84	35	.0855	.85	10	.1255	.30	35	.0835	.30
11	.1296	.84	36	.0594	.85	11	.1273	.30	36	.0625	.30
12	.1251	.85	37	.0417	.84	12	.1283	.30	37	.0436	.30
13	.1435	.85	38	.0607	.84	13	.1401	.30	38	.0606	.30
14	.1034	.85	39	.0689	.84	14	.1012	.30	39	.0718	.30
15	.1159	.85	40	.0302	.84	15	.1126	.30	40	.0333	.30
16	.1096	.86	41	.0500	.84	16	.1080	.30	41	.0540	.30
17	.1254	.85	42	.0316	.84	17	.1252	.30	42	.0281	.30
18	.1402	.85	43	.0390	.84	18	.1374	.30	43	.0381	.30
19	.1317	.86	44	.0331	.84	19	.1292	.30	44	.0328	.30
20	.1160	.86	45	.0548	.84	20	.1173	.31	45	.0500	.29
21	.1193	.85	46	.0616	.82	21	.1172	.30	46	.0599	.29
22	.0910	.85	47	.0308	.82	22	.0930	.30	47	.0301	.29
23	.0991	.85	48	.0172	.81	23	.1057	.30	48	.0174	.29
24	.0847	.85	49	.0051	.81	24	.0883	.30	49	-.0001	.29
25	.0927	.85	50	.0235	.80	25	.0973	.30	50	.0158	.28
θ_0	.2483	31.54					-.5802	4.01			
λ_{t-1}^*	.3281	3.07					-.0151	1.33			
∇p_t	-3.892	142.6					-7.276	51.59			
∇p_t^+	-4.154	314.4					13.60	18.79			
∇s_t	11.128	54.85					4.987	35.40			
1_t	.2420	2.86					.3613	1.30			
LB_{20}	43.95						20.66				
LB_{20}^2	12.60						33.28				

Note: $\lambda_t = \exp(\theta_0 + \theta_1 \lambda_{t-1}^* + \theta_2 \nabla p_t + \theta_3 \nabla p_t^+ + \theta_4 \nabla s_t + \theta_5 1_t)$, where $\lambda_{t-1}^* = \ln \lambda_{t-1}$, $\nabla p_t^+ = 0$ for $\nabla p_t \leq 0$ and $\nabla p_t^+ = \nabla p_t$ for positive news, and $1_t = 1(t \leq 1100)$.

all lags. In addition, the FGLS estimates are sometimes smaller than the CLS estimates. For CLS the lag 49 estimate has a negative sign but is insignificant. It remains insignificant and small when estimated by the FGLS estimator the sign is correct.

In the λ_t function the lagged mean level, λ_{t-1}, has a rather small but significant effect when estimated by the FGLS estimator, while it is insignificant in the CLS estimated model. In terms of the CLS estimates there are significant asymmetric but not very different effects for the price change variable; with a tick size of 0.1 SEK we expect an enhancing average effect of 0.63 for a positive and 0.73 for a negative one tick change. The asymmetry is insignificant for FGLS but the corresponding estimated effects are larger and equal to -0.81 and 0.39, respectively. News about spread increase the frequencies. These are expected signs when compared to duration models for the same underlying data (Brännäs and Simonsen, 2003). The current effects are more significant as for the duration data only the news about prices came out with a significant effect. To account, at least, partly for seasonal within days effects we included a dummy variable $\mathbf{1}(t \leq 1100)$ which takes value one for transactions before 1101 AM and zero otherwise. The estimated effect is positive and significant for both estimators.

There is no practical change in the serial correlation properties for the CLS estimated model with $\hat{\sigma}_t^2 = \hat{\lambda}_t$, though the Ljung-Box statistic for the squared residuals is much smaller and not very far from a p-value of 0.05. In this case the largest autocorrelations are -0.038 and 0.032 (squared standardized residual) for $q = 50$. The effect of news in volume also come out significantly but implies substantial serial correlation in both $\hat{e}_{1t}/\hat{V}^{1/2}(y_t|Y_{t-1})$ and its square. When λ_t is changed to have a linear form there are no serial correlation problems. Unfortunately, $\hat{\lambda}_t$ is then negative for some 30 percent of the observations. Obviously, this is a logically unappealing feature. For the FGLS estimated model we note that conditional heteroskedasticity no longer appears a problem, while the standardized residual now signals trouble. The largest autocorrelation coefficients are 0.042 and 0.021 (squared standarized residual).

Full models including σ_t^2 of the exponential type in (6) have been estimated by GMM. In each instance there is no serial correlation in the standardized residual, but there is serial correlation in the squared standardized residuals of about the same magnitude as the CLS of Table 3.

6. CONCLUDING REMARKS

The suggested integer-valued moving average model has relatively straightforward moment properties and estimating the unknown parameters by well-known techniques is relatively simple. In addition, both the conditional least squares and feasible least squares estimators are readily available in many standard statistical packages and have the good statistical properties.

The current paper focused on modelling a time series of the intra-day number of transactions per time unit using the integer-valued moving average model

class. In its practical implementation for the time series of the number of transactions in Ericsson B, we found both promising and less advantageous features of the model. With the CLS estimator it was relatively easy to model the conditional mean in a satisfactory way in terms of both interpretation and residual properties. It was more difficult to obtain satisfactory squared residual properties for the conditional variance specifications that were tried. The FGLS estimator reversed this picture and we suggest that more empirical research is needed on the interplay between the conditional mean and heteroskedasticity specifications for count data. Depending on research interest the conditional variance parameters are or are not of particular interest. For studying reaction times to shocks or news it is the conditional mean that matters, in much the same way as for conditional duration models. In addition, the conditional variance has no direct ties to, e.g., risk measures included in, e.g., option values or portfolios.

Table A1: Bias and MSE properties in the Monte Carlo experiment of estimators of $\sum_{i=1}^{10} \beta_i$ when the model is estimated as INMA(q), for $q = 10, 20$ and 30. The results are based on 1000 replications, and the bias is multiplied by 100 and the MSE by 100 000.

| | $q = 10$ | | | | $q = 20$ | | | | $q = 30$ | | | |
| | 1000 | | 10000 | | 1000 | | 10000 | | 1000 | | 10000 | |
γ_1	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
						CLS						
-0.1	34.073	3615.148	36.390	2800.233	-7.714	1650.969	-6.778	757.487	0.011	0.016	0.007	0.008
-0.2	-8.169	1807.777	-13.723	1073.082	-13.525	1586.333	-14.437	766.568	0.011	0.017	0.009	0.011
-0.3	-6.948	1374.710	-12.111	597.983	-9.124	1300.164	-9.569	460.360	0.016	0.019	0.015	0.015
-0.4	-4.867	1175.993	-5.883	319.916	-5.562	1142.259	-6.033	286.904	0.020	0.021	0.019	0.018
						FGLS						
-0.1	35.365	3816.869	37.921	3059.045	-6.767	1646.517	-5.520	743.225	0.014	0.016	0.007	0.008
-0.2	-7.803	1792.105	-11.186	986.607	-13.399	1585.772	-14.098	755.514	0.011	0.017	0.012	0.012
-0.3	-7.529	1393.705	-12.157	597.938	-8.926	1294.023	-9.370	456.195	0.013	0.020	0.017	0.017
-0.4	-4.387	1163.534	-6.404	327.260	-5.583	1137.357	-5.917	285.665	0.021	0.022	0.020	0.019
						GMM						
-0.1	21.976	2714.975	21.707	2041.617	-15.382	1770.271	-17.114	1046.574	0.002	0.017	-0.004	0.010
-0.2	-12.272	1829.799	-19.636	1255.089	-14.906	1667.040	-17.836	934.515	0.011	0.018	0.011	0.015
-0.3	-8.401	1407.453	-12.394	653.398	-9.081	1308.663	-10.594	524.195	0.016	0.020	0.014	0.016
-0.4	-5.232	1216.969	-7.897	400.151	-6.001	1142.354	-6.721	329.992	0.020	0.023	0.017	0.017

Table A2: Power properties of the Ljung-Box statistic at 10 degrees of freedom and nominal size 0.05.

	$q = 10$		$q = 20$		$q = 30$	
γ_1	1000	10000	1000	10000	1000	10000
			CLS			
-0.1	0.951	0.997	0.005	1.000	0.881	1.000
-0.2	0.001	0.612	0.000	0.450	0.473	1.000
-0.3	0.000	0.240	0.000	0.003	0.215	1.000
-0.4	0.000	0.018	0.000	0.000	0.118	0.921
			FGLS			
-0.1	0.945	1.000	0.005	1.000	0.881	1.000
-0.2	0.003	0.681	0.000	0.504	0.473	1.000
-0.3	0.000	0.243	0.000	0.002	0.216	1.000
-0.4	0.000	0.014	0.000	0.000	0.118	0.921
			GMM			
-0.1	0.901	1.000	0.011	1.000	0.881	1.000
-0.2	0.014	0.546	0.001	0.402	0.472	1.000
-0.3	0.000	0.273	0.000	0.062	0.216	1.000
-0.4	0.000	0.102	0.000	0.015	0.119	0.920

Table A3: Bias and MSE properties of forecasts $\mu_{T+h|T}$, $h = 1, 2$.

	q = 10								q = 20								q = 30							
	T = 1000				T = 10000				T = 1000				T = 10000				T = 1000				T = 10000			
	Bias(h)		MSE(h)		Bias(h)		MSE(h)		Bias(h)		MSE(h)		Bias(h)		MSE(h)		Bias(h)		MSE(h)		Bias(h)		MSE(h)	
γ_1	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
												CLS												
-0.1	0.38	0.48	15.30	16.03	0.34	0.40	15.62	17.35	0.33	0.44	15.14	15.93	0.27	0.28	15.21	16.48	0.27	0.36	14.82	15.04	0.26	0.27	15.24	16.44
-0.2	0.13	0.16	9.85	10.26	0.19	0.00	9.70	10.41	0.09	0.15	10.02	10.33	0.17	-0.02	9.71	10.40	0.09	0.09	9.92	10.08	0.15	-0.05	9.76	10.38
-0.3	0.02	0.12	7.91	8.48	0.14	0.02	8.20	8.31	0.00	0.13	7.96	8.52	0.14	0.02	8.20	8.29	-0.01	0.09	7.90	8.45	0.12	0.01	8.22	8.29
-0.4	0.01	0.13	7.30	7.56	0.13	0.04	7.09	7.45	-0.00	0.14	7.33	7.63	0.13	0.04	7.13	7.46	-0.00	0.12	7.25	7.55	0.12	0.03	7.12	7.44
												FGLS												
-0.1	0.31	0.42	15.30	15.91	0.31	0.36	15.91	17.25	0.31	0.40	15.08	15.95	0.23	0.23	15.25	16.47	0.28	0.36	14.82	15.04	0.26	0.27	15.24	16.44
-0.2	0.12	0.15	9.79	10.17	0.18	-0.01	9.78	10.39	0.08	0.14	10.04	10.32	0.16	-0.03	9.72	10.38	0.09	0.09	9.92	10.08	0.15	-0.05	9.76	10.38
-0.3	0.01	0.11	7.89	8.46	0.14	0.02	8.17	8.32	-0.00	0.13	7.98	8.53	0.13	0.02	8.21	8.29	-0.01	0.09	7.90	8.45	0.12	0.01	8.22	8.29
-0.4	0.01	0.13	7.29	7.55	0.13	0.04	7.09	7.47	-0.01	0.14	7.36	7.63	0.12	0.03	7.12	7.47	-0.00	0.12	7.25	7.55	0.12	0.03	7.12	7.44
												GMM												
-0.1	0.64	0.84	15.07	15.77	0.59	0.75	15.73	17.22	0.57	0.70	14.93	15.81	0.55	0.60	15.11	16.50	0.28	0.36	14.82	15.04	0.26	0.27	15.24	16.44
-0.2	0.27	0.34	9.83	10.24	0.27	0.10	9.70	10.41	0.16	0.24	9.98	10.36	0.32	0.15	9.68	10.43	0.09	0.09	9.92	10.08	0.15	-0.04	9.76	10.38
-0.3	0.08	0.20	7.91	8.49	0.18	0.07	8.21	8.39	0.02	0.16	7.97	8.55	0.19	0.09	8.24	8.32	-0.01	0.09	7.90	8.45	0.12	0.01	8.22	8.29
-0.4	0.05	0.17	7.31	7.53	0.15	0.06	7.13	7.49	0.02	0.17	7.33	7.67	0.15	0.07	7.14	7.49	-0.00	0.12	7.25	7.55	0.12	0.03	7.12	7.44

<center>REFERENCES</center>

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control* **19**, 716-723.

Al-Osh M. and Alzaid A. (1988). Integer-Valued Moving Average (INMA) Process. *Statistical Papers* **29**, 281-300.

Al-Osh M. and Alzaid A. (1991). Binomial Autoregressive Moving Average Models. *Communications in Statistics: Stochastic Models* **7**, 261-282.

Box, G.E.P. and Jenkins, G.M. (1970). *Time Series Analysis: Forecasting and Control.* San Francisco: Holden-Day.

Brännäs, K. (1995). Explanatory Variables in the AR(1) Count Data Model. *Umeå Economic Studies* **381**.

Brännäs, K. and Brännäs, E. (2004). Conditional Variance in Count Data Regression. To appear in *Communications in Statistics: Theory and Methods* **33**.

Brännäs, K. and Hall, A. (2001). Estimation in Integer-Valued Moving Average Models. *Applied Stochastic Models in Business and Industry* **17**, 277-291.

Brännäs, K. and Hellström, J. (2001). Generalizations to the Integer-Valued AR(1) Model. *Econometric Reviews* **20**, 425-443.

Brännäs, K., Hellström, J. and Nordström, J. (2002). A New Approach to Modelling and Forecasting Monthly Guest Nights in Hotels. *International Journal of Forecasting* **18**, 19-30.

Brännäs, K. and Simonsen, O. (2003). Discretized Time and Conditional Duration Modelling for Stock Transaction Data. *Umeå Economic Studies* **610**.

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

Durbin, J. (1959). Efficient Estimation of Parameters in Moving-Average Models. *Biometrika* **46**, 306-316.

Engle, R.F. and Russell, J.R. (1998). Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica* **66**, 1127-1162.

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

Harris, D. (1999). GMM Estimation of Time Series Models. In Mátyás, L. (ed.) *Generalized Method of Moments Estimation.* Cambridge: Cambridge University Press.

Heinen, A. and Rengifo, E. (2003). Multivariate Modelling of Time Series Count Data: An Autoregressive Conditional Poisson Model. CORE Discussion Paper 2003/25, Université Catholique de Louvain.

Joe, H. (1996). Time Series Models with Univariate Margins in the Convolution-Closed Infinitely Divisible Class. *Journal of Applied Probability* **33**, 664-677.

Jørgensen, B. and Song, P.X-K. (1998). Stationary Time Series Models with Exponential Dispersion Model Margins. *Journal of Applied Probability* **35**, 78-92.

<center>22</center>

Lobato, I., Nankervis, J.C. and Savin, N.E. (2001). Testing for Autocorrelation using a Modified Box-Pierce Q Test. *International Economic Review* **42**, 187-205.

McKenzie E. (1986). Autoregressive-Moving-Average Processes with Negative Binomial and Geometric Marginal Distribution. *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.

McKenzie, E. (2003). Discrete Variate Time Series. In Shanbhag, D.N. and Rao, C.R. (eds.) *Handbook of Statistics*, Volume 21, pp. 573-606. Amsterdam: Elsevier Sciences.

Nelson, D. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica* **59**, 347-370.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1992). *Numerical Recipes*. 2nd ed., Cambridge: Cambridge University Press.