The Asymmetric Count Data Moving Average Model

Kurt Brännäs*

Department of Economics
Umeå School of Business and Economics, Umeå University

Abstract

This note defines the asymmetric count data, first order moving average model and gives some of its basic properties. A brief account of conditional least squares estimation of unknown parameters is also given.

Keywords: Moments, Dual autoregression, Invertibility, Nonlinear, Least squares estimation
JEL: C22, C25, C51

This note gives some basic results for the integer-valued or count data time series model corresponding to Wecker's (1981) asymmetric moving average (asMA) model for continuous variables. Wecker's model has in the asMA(1) case the form $y_t = u_t + \theta^+ u_{t-1}^+ + \theta^- u_{t-1}^-$, with $u^+ = \max(0, u)$ and $u^- = \min(u, 0)$. Hence, differently signed shocks may have different effects on y_t. We may alternatively write the model as $y_t = u_t + \theta^- u_{t-1} + (\theta^+ - \theta^-) I_{t-1} u_{t-1}$, where $I_{t-1} = I(u_{t-1} > 0)$ is the indicator or Heaviside function taking value 1 when the argument is true and 0 otherwise.

The basic count data MA or INMA model and its interpretation and extensions have been discussed by Al-Osh and Alzaid (1988), McKenzie (1988), Brännäs and Hall (2001) and others. In this model, multiplication is replaced by a thinning operation in order to generate integers. For asymmetry in the sense of Wecker the threshold of 0 in the indicator function of the asMA(1) model needs to be replaced by some $k > 0$ since $u_t \geq 0$ for every t, in the count data setting.

We first consider the first order asymmetric integer-valued moving average or asINMA(1) model and briefly discuss two alternative specifications. We also briefly discuss the least squares estimation of unknown model parameters.

*Umeå Economic Studies 841, 2012; email: kurt.brannas@econ.umu.se

1 Model and Properties

The most obvious way of formulating the count data, asymmetric MA(1) or asINMA(1) model is as

$$y_t = u_t + \theta^+ \circ I_{t-1}u_{t-1} + \theta^- \circ (1 - I_{t-1})\, u_{t-1}, \qquad t = 1, 2, \ldots, T \tag{1}$$

where $I_{t-1} = I(u_{t-1} > k)$ is the indicator function. In (1), $\{u_t \geq 0\}$ is a sequence of i.i.d. integer-valued random variables. The discussion is based on a Poisson assumption for u_t. Hence, we assume $E(u_t) = V(u_t) = \lambda$. The binomial thinning operations are defined as $\theta \circ u = \sum_{i=1}^{u} v_i$, where $\{v_i\}$ is a sequence of i.i.d. 0-1 random variables, with $\Pr(v_i = 1) = \theta = 1 - \Pr(v_i = 0)$. The random u is assumed independent of $\{v_i\}$. In addition, thinning operations over time are independent. The parameters of the asINMA model are probabilities and naturally restricted to $[0, 1]$ intervals. When $I_{t-1} = 0$, there is a positive thinning probability for $I_{t-1}u_{t-1} = 0$. With $\theta^+ = \theta^- = \theta$ the model reduces to a standard INMA(1) model $y_t = u_t + \theta \circ u_{t-1}$.

The conditional expectation of (1) is

$$E(y_t|Y_{t-1}) = \lambda + \theta^- u_{t-1} + \delta I_{t-1}u_{t-1}, \tag{2}$$

where $\delta = \theta^+ - \theta^-$, and Y_{t-1} is the observed information up to and including time $t - 1$. Unconditionally we get

$$E(y_t) = \lambda + \theta^- \lambda + \delta(\lambda - e_1), \tag{3}$$

where e_1 is obtained from $e_m = \sum_{i=0}^{k} i^m \lambda^i e^{-\lambda}/i!$ with $m = 1$ in the Poisson case (cf. the Appendix for additional details). If $k \to 0$, e_1 tends to zero and then (3) corresponds to an INMA(1) model with parameter θ^+. As $k \to \infty$, e_1 tends to λ and now the limiting model is an INMA(1) with parameter θ^-.

The conditional and unconditional variances are

$$\begin{aligned} V(y_t|Y_{t-1}) &= \lambda + \delta I_{t-1}u_{t-1} - \kappa I_{t-1}u_{t-1} + \theta^-(1 - \theta^-)u_{t-1} \qquad (4)\\ V(y_t) &= \lambda + \theta^-(1 - \theta^-)\lambda + (\delta + \theta^{-^2})\lambda - (\delta + \kappa + 2\delta\theta^+\lambda)e_1 - \delta^2 e_1^2 - \kappa e_2, \qquad (5) \end{aligned}$$

where $\kappa = \theta^{+^2} - \theta^{-^2}$ The conditional variance also varies asymmetrically with respect to lagged u_{t-1}, while for INMA there is no asymmetry. The first order autocovariance is

$$\begin{aligned} Cov(y_t y_{t-1}) &= \lambda^2 + 2\theta^+\lambda - \delta(\lambda e_1 + e_2) + (\theta^+\lambda - \delta e_1)^2 - E^2(y_t) \\ &= \lambda\left[2\theta^+(1 - \lambda) - \delta e_2\right] \qquad (6) \end{aligned}$$

and the autocorrelations for all higher order lags equal zero.

Next, we consider the dual autoregressive representation of the asINMA(1) in (1). At $t = 1$ and given $u_0 = 0$, $y_1 = u_1$ and then $y_2 = u_2 + \theta^+ \circ I_1 y_1 + \theta^- \circ (1 - I_1)y_1$, where equality is in

Table 1: Parameter values and probabilities for the first three lags in the dual autoregression of an asINMA(1) model at $t = 4$ with $\theta^+ = 0.8$ and $\theta^- = 0.4$ and probabilities $p = \Pr(I_t = 1) = 0.75$ and $1 - p = 0.25$.

Lag	Values/Probabilities				Expectation
1	0.8	0.4			
	0.75	0.25			0.70
2	0.64	0.16	0.32		
	0.563	0.063	0.375		0.49
3	0.512	0.064	0.256	0.128	
	0.422	0.016	0.422	0.141	0.34

distribution. For $t = 4$ we give the expression in terms of expectations

$$\begin{aligned} E(y_4) = {} & \lambda + \theta^+ E(I_3 y_3) + \theta^- E(\bar{I}_3 y_3) \\ & -\theta^{+^2} E(I_3 I_2 y_2) - \theta^{-^2} E(\bar{I}_3 \bar{I}_2 y_2) - \theta^- \theta^+ E(\bar{I}_3 I_2 y_2) - \theta^+ \theta^- E(I_3 \bar{I}_2 y_2) \\ & +\theta^{+^3} E(I_3 I_2 I_1 y_1) + \theta^{-^3} E(\bar{I}_3 \bar{I}_2 \bar{I}_1 y_1) \\ & +\theta^{+^2} \theta^- E(I_3 I_2 \bar{I}_1 y_1) + \theta^{+^2} \theta^- E(I_3 \bar{I}_2 I_1 y_1) + \theta^{+^2} \theta^- E(\bar{I}_3 I_2 I_1 y_1) \\ & +\theta^+ \theta^{-^2} E(I_3 \bar{I}_2 \bar{I}_1 y_1) + \theta^+ \theta^{-^2} E(\bar{I}_3 I_2 \bar{I}_1 y_1) + \theta^+ \theta^{-^2} E(\bar{I}_3 \bar{I}_2 I_1 y_1) \end{aligned} \tag{7}$$

where $\bar{I}_j = 1 - I_j$. Depending on the outcomes of u_1, u_2 and u_3 there is in this $t = 4$ case 2^3 different autoregressive models. For instance, for all $I_t = 1$ we get $y_4 = u_4 + \theta^+ \circ y_3 - \theta^{+^2} \circ y_2 + \theta^{+^3} \circ y_1$, while for only $u_2 \leq k$, $y_4 = u_4 + \theta^+ \circ y_3 - \theta^+ \theta^- \circ y_2 + \theta^{+^2} \theta^- \circ y_1$.

The parameters in (7) arise with binomial probabilities $\Pr(j) = [m!/(j!(m-j)!)]\, p^j (1-p)^{m-j}$ where m is the lag and $j = 0, ..., m$. The number of possible parameter values increases with the lag but the associated probabilities get smaller at a fast rate if both θ^+ and θ^- are smaller than one. Note also that the parameters at even lags have negative signs. The expected value at lag m is given by $\sum_{j=0}^{m} \Pr(j)(\theta^+)^j (\theta^-)^{m-j}$. Table 1 gives an example. Related results for the extension of the asMA(1) model are considered by Brännäs and Ohlsson (1999).

Since θ^+ and θ^- are both in $[0, 1]$ intervals, we conjecture that for a stationary dual AR model to exist, at least, one of the θ's needs to be smaller than one, subject also to $\Pr(I_t = 1) \in (0, 1)$. When this is the case we say that the asINMA(1) is invertible. De Gooijer and Brännäs (1995) found that for a conventional asMA(1) model $|\theta^+||\theta^-| < 1$ is a necessary condition, using a numerical invertibility approach and assuming normality. This approach gives a larger invertibility region than the originally proposed $|\theta^+| < 1$ and $|\theta^-| < 1$ region of Wecker (1981).

An alternative model specification to (1) is

$$y_t = u_t + \theta^+ I_{t-1} \circ u_{t-1} + \theta^- (1 - I_{t-1}) \circ u_{t-1}. \qquad t = 1, 2, \ldots, T \tag{8}$$

The indicator variable effectively switches the probabilities θ^+ and θ^- to zero depending on the outcome of the discrete u_{t-1} variable. The model can be said to be a special case of a random coefficient INMA(1) model. In model (8) we have thinning $\theta I \circ u = \sum_{i=1}^{u} w_i$, with $w_i = 1$ if $v_i = 1$ and $I = 1$ and $w_i = 0$ otherwise, so that $\Pr(w_i = 1) = \theta p$ with $p = \Pr(I = 1)$. In (1), the thinning is $\theta \circ Iu = \theta \circ u$ for $I = 1$ and equal to zero otherwise. Practically, there appears to be little difference between the two model versions, but it remains to study whether the thinning operations are different or equal.

The probability generating function of $x_0 = \theta \circ u$ conditional on u is $P_0(t|u) = E(t^x|u) = [\theta t + (1-\theta)]^u$ which corresponds to the binomial distribution and unconditionally $P_0(t) = E(t^x) = \exp[\lambda\theta(t-1)]$, which tells us that x_0 is Poisson distributed with mean $\lambda\theta$. For the standard INMA(1) model we then obtain the generating function for y_t as $P_y(t) = \exp[\lambda(1+\theta)(t-1)]$, which again corresponds to a Poisson distribution.

Using $E(t^x) = E(t^x|u, I=0)\Pr(u \leq k) + E(t^x|u, I=1)\Pr(u > k)$ we find for the thinning operation in (1), i.e. $x_1 = \theta \circ Iu$, and for the $x_2 = \theta I \circ u$ of (8) that $P_1(t) = E(t^{x_1}) = (1-p) + pE(t^{x_1}|u > k) = P_2(t) = E(t^{x_2})$. From this follows equality in distribution for the thinning operations and then also for the unconditional distributions of y_t in (1) and (8).

Obviously, the indicator function $I(u_t > k)$ has a key role in the model. If we were willing to consider an alternative such as $I'(x_t > k) = I'_t$ for some exogenous or predetermined variable x_t (independent of u_t) , moment results would be easy to obtain. For instance, the expected value is then $E(y_t) = \lambda + \theta^+ I'_t \lambda + \theta^-(1 - I'_t)\lambda$.

2 Estimating k

In the preceding discussion k was viewed as a known positive integer or real value. For empirical purposes we recognize that k will be unknown in most instances. We consider two approaches to estimation, both based on the conditional least squares estimator. In the first, we estimate the other parameters $\psi_1 = (\theta^+, \theta^-, \lambda)'$ for any given k. The estimated value of k is next determined by AIC, SBIC or some other related criterion. In the second approach, we view k as a real positive but unknown parameter, so that the vector to be estimated is $\psi_2 = (\theta^+, \theta^-, \lambda, k)'$. We may approximate the indicator function $I(u > k)$ by say a distribution function centered at the unknown k. A simple choice is the logistic distribution function, i.e. $I \approx 1 - 1/[1 + \exp(-c(u-k))]$, where c is some large and given value. To reduce the number of unknowns we could consider setting $k = \lambda$ so that $\psi_3 = (\theta^+, \theta^-, \lambda)'$ and with I a function of λ.

The one-step-ahead prediction error

$$z_t = y_t - \lambda - \theta^- u_{t-1} - \delta I_{t-1} u_{t-1} \tag{9}$$

is real valued and related to u_t in the sense that $z_t + \lambda$ has the same conditional and unconditional first and second moments as u_t. Using the relationship to z_t we calculate u_t sequentially. This leap is also employed in Brännäs and Hall (2001).

Given this, the easiest general approach to estimation which is also robust towards both distributional misspecification and alternative forms of dependence among thinning operations (cf. Brännäs and Hall, 2001) is the conditional least squares estimator based on (9). Hence, the parameter vectors $\psi_i, i = 1, 2, 3$ minimize the least squares criterion function $S = \sum_t^T z_t^2$, where z_t is the prediction error. The conditional least squares estimator for this different moving average model differs from the conventional one with respect to the $E(u_t) \neq 0$ and k. The derivatives with respect to the parameters are:

$$\begin{aligned}
\frac{\partial z_t}{\partial \theta^+} &= \frac{\partial u_t}{\partial \theta^+} = -I_{t-1}u_{t-1} - \theta^- \frac{u_{t-1}}{\partial \theta^+} - \delta \left[\frac{\partial I_{t-1}}{\partial \theta^+} u_{t-1} + I_{t-1} \frac{\partial u_{t-1}}{\partial \theta^+} \right] \\
\frac{\partial z_t}{\partial \theta^-} &= \frac{\partial u_t}{\partial \theta^-} = -u_{t-1} + I_{t-1}u_{t-1} - \theta^- \frac{\partial u_{t-1}}{\partial \theta^-} - \delta \left[\frac{\partial I_{t-1}}{\partial \theta^-} u_{t-1} + I_{t-1} \frac{\partial u_{t-1}}{\partial \theta^-} \right] \\
\frac{\partial z_t}{\partial \lambda} &= \frac{\partial u_t}{\partial \lambda} - 1 = -1 - \theta^- \frac{\partial u_{t-1}}{\partial \lambda} - \delta \left[\frac{\partial I_{t-1}}{\partial \lambda} u_{t-1} + I_{t-1} \frac{\partial u_{t-1}}{\partial k} \right] \\
\frac{\partial z_t}{\partial k} &= \frac{\partial u_t}{\partial k} = -\theta^- \frac{\partial u_{t-1}}{\partial k} - \delta \left[\frac{\partial I_{t-1}}{\partial k} u_{t-1} + I_{t-1} \frac{\partial u_{t-1}}{\partial \lambda} \right]
\end{aligned}$$

which can all be calculated recursively from zero initial values.

For the covariance matrix of the parameter vector $\hat{\psi}_i'$ we advocate a sandwich estimator that is robust against both violations in assumptions about thinning operations as well as conditional heteroskedasticity. Its form is

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

where

$$F = \sum_{t=2}^{T} \frac{\partial z_t}{\partial \psi} \frac{\partial z_t}{\partial \psi'} \quad \text{and} \quad J = \sum_{t=2}^{T} z_t^2 \frac{\partial z_t}{\partial \psi} \frac{\partial z_t}{\partial \psi'}.$$

3 Discussion

The note has focused on the simplest of asymmetric integer-valued or count data moving average models. For continuous variables, the basic asMA model has been extended in particular for financial time series to capture asymmetries in both returns and volatilities as well simultaneity for multiple time series. Extending the current study to cover longer INMA lag structures and an INAR part appears tractable. If one wishes to include additional flexibility in conditional heteroskedasticity/variance the direct approach would be to relax the Poisson assumption about u_t and replace it with $\sigma_t u_t$ such that this has conditional variance σ_t^2 which modelwise could borrow from the GARCH literature.

Appendix

The density of $I_t u_t$ is left censored such that the outcome zero is observed when $u_t \leq k$ with probability $1-p = \Pr(u \leq k) = \sum_{i=0}^{k} p_i$. Hence, the expected value is $(1-p) \times 0 + p \sum_{i=k+1}^{\infty} i p_i / p = \lambda - e_1$. The density of $(1-I_t)u_t$ is right censored such that all $u_t \leq k$ outcomes are observed while for $u_t > k$ only zeroes are observed. With the censoring value being zero the probability for the zero outcome is increased by the probability for $u_t > k$, i.e. $p_0 + p$. There will be no impact on expected values e_m, since $\sum_{i=0}^{k} i^m p_i = \sum_{i=1}^{k} i^m p_i$ for any k.

For the thinning operation $\theta \circ (1-I)u$ the expected value is obtained from $E(\theta \circ (1-I)u) = E(\theta \circ (1-I)u|I=0)\Pr(I=0) + E(\theta \circ (1-I)u|I=1)\Pr(I=1)$. Here, $E(\theta \circ (1-I)u|I=0) = E(\sum_{i=1}^{u} v_i|I=0) = \theta E(u|u \leq k) = \theta(1-p)^{-1}\sum_{i=1}^{k} i p_i = \theta(1-p)^{-1}e_1$. Conditioning on $I=1$ instead gives conditional expectation equal to zero. For $E(\theta \circ Iu|I=1)$ we get $\theta p^{-1}(\lambda - e_1)$, and $E(\theta \circ Iu|I=0)=0$. It follows that $E(y_t) = \lambda + \theta^+(\lambda - e_1) + \theta^- e_1 = \lambda + \theta^+\lambda - (\theta^+ - \theta^-)e_1$

The variance is obtained using

$$\begin{aligned} E\left[V(y_t|Y_{t-1})\right] &= \lambda + \theta^-(1-\theta^-)\lambda + \left[\delta - \left({\theta^+}^2 - {\theta^-}^2\right)\right](\lambda - e_1) \\ V\left[E(y_t|Y_{t-1})\right] &= \lambda\left[{\theta^+}^2 + 2\delta\theta^+ e_1\right] - \delta^2\left[e_2 + e_1^2\right] - 2\theta^-\delta e_2, \end{aligned}$$

where $e_m = \sum_{i=0}^{k} i^m p_i$. Then $V(y_t) = E\left[V(y_t|Y_{t-1})\right] + V\left[E(y_t|Y_{t-1})\right]$.

The autocovariance function at lag 1 is

$$\begin{aligned} E(y_t y_{t-1}) &= E\left[u_t + \theta^+ I_{t-1} \circ u_{t-1} + \theta^- \circ u_{t-1} - \theta^- I_{t-1} \circ u_{t-1}\right] \\ &\quad \times \left[u_{t-1} + \theta^+ I_{t-2} \circ u_{t-2} + \theta^- \circ u_{t-2} - \theta^- I_{t-2} \circ u_{t-2}\right] \\ &= \lambda^2 + 2\theta^+\lambda - \delta(\lambda e_1 + e_2) + (\theta^+\lambda - \delta e_1)^2. \end{aligned}$$

Referenses

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

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 Ohlsson, H. (1999). Asymmetric Time Series and Temporal Aggregation. *Review of Economics and Statistics* **81**, 341-344.

De Gooijer, J.G. and Brännäs, K. (1995). Invertibility of Nonlinear Time Series Models. *Communications in Statistics: Theory and Methods* **24**, 2701-2714.

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

Wecker, W.E. (1981). Asymmetric Time Series. *Journal of the American Statistical Association* **76**, 16-21.