The Number of Traded Shares: A Time Series Modelling Approach

Kurt Brännäs
Department of Economics
Umeå School of Business and Economics, Umeå University
kurt.brannas@econ.umu.se

Umeå Economic Studies 860, 2013

Abstract

This short paper proposes a characterization for the number of traded shares or trading volume in terms of its data generating process. Share ownership plays a vital role. An empirical illustration based on the Nokia stock is included. Long memory in trading volume is linked to the long memory feature of ownership.

Key Words – Integer-valued, INAR, Ownership, Trading volume, Stock, Nokia
JEL – C22, C51, C58, G10, G32

1 Introduction

The financial asset pricing literature is a huge one, while the corresponding quantity or volume aspect of such trades has received much less attention. The paper focuses on the quantity side of trades by studying the impact of the number of old and new owners (investors) and separately by the number of trades (transactions) on the number of traded shares (trading volume) in individual stocks. Our focus is on time series characteristics rather than on any cross-sectional or high frequency aspects since to illustrate we only have access to aggregate monthly stock data.

There are some studies dealing with related questions but where other issues are stressed or data are of a different type than ours. Lo and Wang (2000) and Lo, Mamaysky and Wang (2004) empirically as well as theoretically studied pricing and volume together. Their preferred measure for volume is a relative one, the number of traded stocks relative to the number of outstanding shares, which they label turnover.

This measure is particularly convenient for their equilibrium model. In both their empirical and theoretical works they view the number of owners or the number of trades as given, while we stress the importance of these in view of the data generating process. Lee, Tsai and Lee (2013) recently emphasized the role of trading volume for asset pricing with a disequilibrium price adjustment. There is also a strand of literature on long memory aspects in trading volume and volatility which was recently summed up by Fleming and Kirby (2011) and Rossi and Santucci de Magistris (2013). They find for high frequency data that trading volume is fractionally integrated and that trading volume may be useful in forecasting daily volatility. Whether such long memory remains at the monthly frequency is an empirical question.

This paper departs from a description of the data generating process with its integer-valued data. The integer-valued time series models can through their parameters be interpreted in interesting ways and estimation is quite straightforward. The approach enables us to, e.g., study the origin of the long memory found in trading volume series. Integer-valued time series models were first introduced by McKenzie (1985) and Al-Osh and Alzaid (1987), for a survey see McKenzie (2003).

2 Modelling

2.1 Approach based on Ownership

We denote the number of share owners in a single stock at time t by y_t. If each share owner has $X_{ti}, i = 1, 2, \dots, y_t$, shares, the total number of shares at time t is

$$z_t = \sum_{i=1}^{y_t} X_{ti}. \tag{1}$$

Small company stocks are usually traded only in one market place, while stocks of large companies may be traded in several market places. There may also be share splits, share repurchases, etc. As the total number of shares in a company is fixed, at least, in the short term, the z_t can be viewed as fixed for small companies. For large companies, z_t may perhaps best be viewed as random. In small stocks it appears reasonable to expect a negative relationship between y_t and the $\{X_{ti}\}$ sequence, while for larger stocks such dependence is likely to be much weak.

The changes in the individual holdings of those owning shares at $t-1$ are, given one trade, given as:

$$x_{ti}^* = X_{ti} - X_{t-1,i} \begin{cases} = -X_{t-1,i}, & \text{owner } i \text{ is not an owner at } t \\ \in (-X_{t-1,i}, 0), & \text{owner } i \text{ has reduced the holdings} \\ = 0, & \text{no change} \\ > 0, & \text{owner } i \text{ has increased the holdings} \end{cases} \tag{2}$$

The number of traded shares is then $x_{ti} = |x^*_{ti}|$ for every owner i at time $t-1$. For a new owner $x^*_{ti} = e_{ti} = X_{ti}$ among a total of ϵ_t new owners.

Defining $1_i = 1$ for $x^*_{ti} \leq 0$ and $1_i = 0$ otherwise, we may write $-\sum_{i=1}^{y_{t-1}} 1_i x^*_{ti} = \sum_{i=1}^{y_{t-1}} (1-1_i) x^*_{ti} + \sum_{j=1}^{\epsilon_t} x^*_{tj}$, i.e. for a constant total number of shares, the number of sold shares must equal the number of bought shares by owners at $t-1$ or by the new owners. Hence, the entrants may acquire the difference between sold and bought shares by the previous owners. For longer time intervals and/or in the case of frequent trading, x^*_{ti} is best viewed as being a count variable with no upper limit.

The total number of traded shares w_t, in a short interval $(t-1, t]$ or when trading only takes place at discrete and equidistant points in time, arises from trading in the holdings of each owner $i = 1, \ldots, y_{t-1}$ at time $t-1$, but also from the entry of new owners or entrants in the period. The number of new owners is denoted ϵ_t. Given that the number of sold shares is equal to the number of bought shares we need to be observant and to avoid double counting. We may write the number of traded (here bought) shares as

$$w_t = \sum_{i=1}^{y_{t-1}} x_{ti} + \sum_{i=1}^{\epsilon_t} e_{ti}. \tag{3}$$

When trading is in continuous time and many trades can take place in a time interval $(t-1, t]$ we decompose the number of traded shares in the stock at time t as in (3). The x_{ti} is then interpreted as the aggregate number of traded (either bought or sold) shares in the time interval for owner i and analogously for e_{ti}. In our case we can observe both the number of traded shares w_t and the number of owners y_{t-1}.

To arrive at a model for the $\{w_t\}$ sequence we first let the unobservable second sum in (3) be a random term $\xi^e_t = \sum_{i=1}^{\epsilon_t} e_{ti}$ with mean $E(\xi^e_t) = \lambda > 0$. Second, we need to impose assumptions separately about the y_t and $\{x_{ti}\}$ sequences and about the independence/dependence between y_{t-1} and $\{x_{ti}\}$.

The strongest assumptions are those of independence between y_{t-1} and x_{ti}, within the $\{x_{ti}\}$ sequence, as well as time invariance. These assumptions give that $E(w_t) = E_y[E(w_t|y_{t-1})] = \mu_x E_y(y_{t-1}) + \lambda = \mu_x \mu_y + \lambda$. Conditioning on the information available at time $t-1$, i.e. $W_{t-1} = (w_1, \ldots, w_{t-1})$ gives

$$E(w_t|W_{t-1}) = \mu_x y_{t-1} + \lambda \tag{4}$$

for $E(x_{ti}|W_{t-1}) = \mu_x$, for all i. When, the lagged number of owners, i.e. y_{t-1}, can be observed, the model in (4) potentially with time dependent parameters can be estimated by, e.g., a conditional least squares (CLS) estimator.

Brännäs (2013) recently discussed the INAR(1) specification $y_t = \alpha \circ y_{t-1} + \eta_t$ for the number of owners in individual stocks and found for monthly data that for the residual $\hat{\eta}_t$ to be serially uncorrelated the α and $\lambda_\eta = E(\eta_t)$ parameters needed to be made time-dependent. The α parameter was empirically found to be close to one and then indicating quite long average holding times for some large Finnish and Swedish stocks.

To incorporate such an INAR(1) dependence structure we may write $w_t = E(w_t|W_{t-1}) + \xi_t$ with $E(\xi_t|W_{t-1}) = 0$ and use (4) to get, at least, the following alternative model representations

$$\begin{aligned} w_t &= \mu_x y_{t-1} + \lambda + \xi_t \\ &= \beta w_{t-1} + \lambda - (\beta w_{t-1} - \mu_x y_{t-1}) + \xi_t \\ &= \alpha w_{t-1} + (1-\alpha)\lambda + \mu_x y_{t-1} - \alpha\mu_x y_{t-2} + \xi_t - \alpha\xi_{t-1}. \end{aligned} \tag{5}$$

Hence, with an observable y_{t-1} the random error term is ξ_t and the parameters can be estimated separately by a CLS estimator based on the first expression. The second expression is a simple rewrite that incorporates a first order lag of the number of traded shares. The third expression is based on the second, with $\alpha = \beta$ and where the conditional INAR(1) representation is used for lagged ownership y_{t-1}. Empirical evidence suggests that α is close to one, so that a long memory property in trading volume can be attributed to ownership. Again, this model representation with its parameter restrictions can be estimated by a CLS estimator.

To include a time-varying model for the integer-valued x_{ti} sequences we need to assume some simplifying model as there are no available cross-sectional data. We may, e.g., consider a conditional representation $x_{ti} = \mu_{t-1}^x + \theta_i + \epsilon_{ti}$, where θ_i is viewed as a random effect and where ϵ_{ti} has zero mean. If $\theta_i = 0$, for every i, this is a replicated time series model. The conditional mean of x_{ti} is then $\mu_{t-1}^x + \mu_\theta$, where $\mu_\theta = E(\theta_i)$. Hence, the conditional representation for w_t can be written as

$$w_t = (\mu_{t-1}^x + \mu_\theta)y_{t-1} + \lambda + \xi_t. \tag{6}$$

If μ_t^x arises in accordance with, say, an INARMA model, the conditional mean of w_t will contain unobservable lags related to x_{ti}, therefore suggesting that such a specification is not empirically tractable. With $\mu_{t-1}^x = \mathbf{d}_{t-1}'\gamma$, where $\mathbf{d}_t$ is a vector of a constant and dummy variables that, e.g., reflect calendar features, we may write (6) as $w_t = y_{t-1}\mathbf{d}_{t-1}'\gamma + \lambda + \xi_t$. Such a representation as well as a modified versions of (5), i.e. $w_t = \alpha w_{t-1} + (1-\alpha)\lambda + \mathbf{d}_{t-1}'\gamma y_{t-1} - \alpha\mathbf{d}_{t-2}'\gamma y_{t-2} + \xi_t - \alpha\xi_{t-1}$ can be estimated.

The treatment above is largely in terms of conditional expectations. A reason for this is that higher order conditional and unconditional moments depend explicitly on assumptions on, e.g., the thinning operations in the INAR(1) (cf. Brännäs and Hellström, 2001). To account for dependence between y_t and the x_{ti} sequence as well as over time it appears necessary to add distributional assumptions. It then appears unlikely that closed form expressions would arise. Therefore, such more general but not necessarily

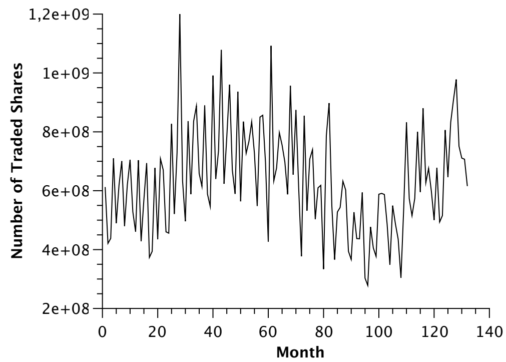

Figure 1: The monthly series of the number of traded Nokia shares (in Helsinki), January 2002 – December 2112, $T = 132$.

more reasonable treatments will additionally necessitate more sophisticated estimation approaches.

2.2 Approach based on Transactions

We may take a different view and depart from $w_t = \sum_{i=1}^{z_t} x_{ti}$, where we view z_t as indicating the number of transactions or trades in the time interval $(t-1, t]$, and where x_{ti} is viewed as the number of traded shares in each of these transactions. Studies of intraday number of transactions in five minute intervals indicate that there is long memory (e.g., Brännäs and Quoreshi, 2010; Quoreshi, 2012). Hence, the main features of this formulation are close to those of the number of owners characterization. For our purposes, however, a full length series for the number of transaction is not available.

3 Empirical Illustration

A times series of the monthly number of traded shares in Nokia (total volume, January 2002 – December 2112, $T = 132$, downloaded January 2013 from www.nasdaqomxnordic.com/shares) is used for illustrative as well as explorative purposes. In searching for model specifications related to those of Section 2 we look for specifications that are free of remaining serial correlation and where key parameters are significantly estimated. Estimation is throughout by the CLS estimator, and robust standard errors (with respect to heteroskedasticity and serial correlation up to lag three) are used throughout.

The time series of the number of traded shares in Nokia (w_t) is exhibited in Figure 1 and its sample autocorrelation function is in Figure 2. The number of traded shares

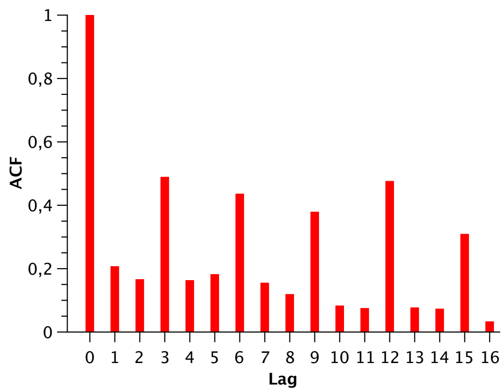

Figure 2: Autocorrelation function for the monthly series of the number of traded Nokia shares (in Helsinki), January 2002 – December 2112, $T = 132$.

is quite large with an average per day of about $30 \cdot 10^6$ shares (21 trading days per month). From Figure 2 it is quite evident that there is a strong, quarterly, seasonal effect which we interpret as due to news related to the reporting cycle (an annual report and quarterly reports) of Nokia.

To account for the seasonal pattern we employ a dummy variable $r_t = 1$ for months with a report and $r_t = 0$ otherwise. We also have access to a time series for the number of share owners (y_t) in Nokia for the same time period (mean 159408 registered owners in the Finnish depository). Unfortunately, there is no complete series for the number of transactions (z_t) in Nokia for the full period. However, for daily records in the final part of the period $(T = 1383$, mean 12310 daily transactions) we find an INAR(1) pattern $\hat{z}_t = 0.44\,\hat{z}_{t-1} + 12316$. Time aggregating this model to a monthly one suggests that no serial correlation is to be expected in a monthly number of transactions series. On the other hand, disaggregating to, say, five minute intervals suggests that α in that case is about 0.99, so there is support for the frequently found long memory property for intra-daily data. For sparsely sampled series, such as the current monthly series, any long memory component in trading volume can then be attributed to the long memory inherent in ownership series.

A model that departs from the final expression in (5) and accounts for the cyclical feature of the data can be specified and estimated by CLS. Note, that x_{ti} has a time invariant first order moment in this case. The estimated model is

$$\begin{aligned} \hat{w}_t &= 0.94\,\hat{w}_{t-1} - 43 \cdot 10^6 + 228 \cdot 10^6\, r_t + 568\, y_{t-1} - 0.94 \cdot 568\, y_{t-2} \\ &\quad -0.94\,\hat{\xi}_{t-1} + 0.16\,\hat{\xi}_{t-3}, \end{aligned} \tag{7}$$

where except for $\hat{\mu}_x = 568$ and the constant term in the time-varying $\hat{\lambda}_t = -43 \cdot 10^6 + 228 \cdot 10^6\, r_t$ the other estimates are significant. There is no significant serial correlation in the residual (Ljung-Box $\mathrm{LB}_{16} = 24.4, p = 0.08$) and $R^2 = 0.34$. The third order lag of ξ_t and the included r_t both account for the reporting cycle. Without these two variables there is substantial serial correlation. The estimated average number of traded shares in a time period for an individual owner is 568 and the mean holding time $1/(1-\alpha)$ is estimated to be 16.7 months.

Including a third order lag w_{t-3} instead of ξ_{t-3} gives a model that fits data slightly better and has no remaining serial correlation. However, this gives an unrealistically short mean holding time of 4.5 months. For this model we have a largest eigenvalue of 0.98 which still indicates quite a long memory. Note that in an INAR(1) model for monthly ownership data the lag one parameter estimate is 0.97. In either specification an increase in ownership increases the number of traded shares.

Attempts to estimate models with extended calendar effects such as $\mu_t^x = \mathbf{d}_t'\gamma$ (cf. Section 2.1) resulted in severe serial correlation in residuals and results are therefore not reported. A standard ARMA time series approach based on only the w_t series gives the CLS estimated model

$$\hat{w}_t = 0.93\, \hat{w}_{t-3} + 689 \cdot 10^6 + 0.14\, \hat{\xi}_{t-1} - 0.71\, \hat{\xi}_{t-3}, \tag{8}$$

where all estimates are significant. There is no remaining serial correlation ($\mathrm{LB}_{29} = 33.3, p = 0.27$) and the goodness-of-fit is $(R^2 = 0.36)$. If this model was to be interpreted as an INARMA, we note that the lag three INMA parameter estimate is violating the sign restriction. This violation remains for other related model specifications and also if r_t is included as an explanatory variable. In (7) there is no such sign violation.

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.

Brännäs, K. (2013). The Number of Shareholders – Time Series Modelling and Some Empirical Results. Umeå Economic Studies 855.

Brännäs, K. and Hellström, J. (2001). Generalized Integer-Valued Autoregression. *Econometric Reviews* **20**, 425-443.

Brännäs, K. and Quoreshi, A.M.M.S. (2010). Integer-Valued Moving Average Modelling of the Number of Transactions in Stocks. *Applied Financial Economics* **20**, 1429-1440.

Fleming, J. and Kirby, C. (2011). Long Memory in Volatility and Trading Volume. *Journal of Banking & Finance* **35**, 1714-1726.

Lee, C-F., Tsai, C-M. and Lee, A.C. (2013) Asset Pricing with Disequilibrium Price Adjustment: Theory and Empirical Evidence. *Quantitative Finance* **13**, 227-239.

Lo, A.W., Mamaysky, H. and Wang, J. (2004). Asset Prices and Trading Volume Under Fixed Transactions Costs. *Journal of Political Economy* **112**, 1054-1090.

Lo, A.W. and Wang, J. (2000). Trading Volume: Definitions, Data Analysis, and Implications of Portfolio Theory. *Review of Financial Studies* **13**, 257-300.

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

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

Quoreshi, A.M.M.S. (2012). A Long-Memory Integer-Valued Time Series Model, INARFIMA, for Financial Application. To appear in *Quantitative Finance*.

Rossi, E. and Santucci de Magistri, P. (2013). Long Memory and Tail Dependence in Trading Volume. *Journal of Empirical Finance* **22**, 94-112.