Order Flow, Transaction Clock, and Normality of Asset Returns: A Comment on Ané and Geman (2000)*

Anthony Murphy and Marwan Izzeldin
Nuffield College, Oxford and Lancaster University

November 2005

Abstract

We investigate the procedure used by Ané and Geman (2000) to recover the moments of information flow from high frequency data in a model which generalizes the subordinated / mixture of distributions process in Clark (1973). Using Monte Carlo experiments we show that the third and higher moments of the latent information flow cannot be accurately recovered using their univariate procedure. We explain why this happens. In our data, returns conditioned on the recentered number of trades are not Gaussian.

Introduction

Since the highly influential work of Clark (1973), researchers have devoted a great deal of time and effort to examining the relationship between the volatility of returns and measures of market activity such as volume and the number of transactions.[1] Clark was the first to propose using a stochastic clock as a time changer in order to recover the normality of returns. Clark wrote the price process for cotton futures as a subordinated process, where the economic interpretation of the subordinator was the cumulative volume of traded contracts.

Recently, in a widely cited and important paper, Ané and Geman (AG, 2000) revisited Clark's method of dealing with the non-normality of observed returns. They considered a general time change process. Using a novel non-parametric procedure, AG claim to have recovered the moments of the time change process which, apart from it's first moment, matched the moments of the observed

*We thank Professor Hélyette Geman and Clive Bowsher for their helpful comments. Corresponding author: Anthony Murphy, Nuffield College, New Road, Oxford OX1 1NF, UK. Email: anthony.murphy@nuffield.ox.ac.uk. Phone: +44 (0) 1865 278526.

[1] For example, see Besembinder and Seguin (1993), Chan and Fong (2000), Easley and O'Hara (1992), Easley et. al. (1997), Epps and Epps (1976), Gallant *et. al.* (1992), Harris (1986, 1987), Hasbrouk (1999), Jones *et. al.* (1994), Karpoff (1987, 1988), Lamaourex and Lastrapes (1990, 1994), Li (2004), Liesenfeld (1998, 2001), Richardson and Smith (1994) and Tauchen and Pitts (1983).

number of transactions but not those of volume in their data. Geman (2005, p.2712) summarized this result as showing, "that in order to recover a quasi perfect normality of returns, the transactions clock is better represented by the number of trades than the volume".

This result is not uncontroversial. On the one hand it is consistent with the findings of Hasbrouk (1999), Easley and O'Hara (1992), Jones, Kaul and Lipton (1994) and Lillo, Farmer and Mantegna (2003). On the other hand, Li (2004) and Izzeldin (2005), for example, cannot reproduce this finding. In this paper, we examine the procedure AG used to recover the moments of the time changer or information flow. We identify some problems with the procedure and show, using Monte Carlo experiments, that their procedure produces extremely inaccurate estimates of the higher moments of the time changer. We suggest that, contrary to the claim in AG, it is very difficult to model the stochastic time change non-parametrically.

The outline of this note is as follows. In Section I, we sketch the AG model of returns and critically examine their procedure for recovering the moments of the stochastic time change generating the non-normality of returns. Our Monte Carlo experiments are described in Section II. The results of these experiments are discussed in Section III. In Section IV we present some empirical evidence showing that the use of transactions and volume clocks do not produce near normal returns. We conclude with a brief summary in Section V.

I. Ané and Geman's Procedure for Recovering the Moments of the Stochastic Time Changer

Ané and Geman (AG, 2000) consider a general return process $r(t) = x(i(t))$ where $x(.)$ is a Brownian motion and $i(t)$ is some stochastic time change or information flow process, which may include a jump component. The stochastic time change process generalizes the subordinated processes considered by Clark (1973). The Brownian motion assumption is innocuous since, as AG note, any arbitrage-free return process can be expressed as a time-changed Brownian motion process (Monroe, 1978). At a point in time, a time changed Brownian motion process is just a normal mixture.

We consider a discrete time version of the AG process since the notation is somewhat simpler and our arguments remain valid in continuous time. In discrete time, AG assume that, conditional on the exogenous time changer / information flow i_t, returns r_t are normally distributed with mean $\mu_r i_t$ and variance $\sigma_r^2 i_t$. Thus r_t is distributed as a normal mixture:

$$r_t \mid i_t \sim N(\mu_r i_t, \sigma_r^2 i_t). \tag{1}$$

Let $m_1^i, ..., m_6^i$ denote the first six (central) moments of i_t. Then, it is straight forward to show that the first six unconditional (central) moments $m_1^r, ..., m_6^r$ of the normal mixture r_t are as follows:

$$
\begin{aligned}
m_1^r &= \mu_r m_1^i \\
m_2^r &= \sigma_r^2 m_1^i + \mu_r^2 m_2^i \\
m_3^r &= 3\mu_r \sigma_r^2 m_2^i + \mu_r^3 m_3^i \\
m_4^r &= \mu_r^4 m_4^i + 6\sigma_r^2 \mu_r^2 m_3^i + 6\sigma_r^2 \mu_r^2 m_1^i m_2^i + 3\sigma_r^4 (m_2^i + (m_1^i)^2) \\
m_5^r &= \mu_r^5 m_5^i + 10\sigma_r^2 \mu_r^3 m_4^i + (10\sigma_r^2 \mu_r^3 m_1^i + 15\mu_r \sigma_r^4) m_3^i + 30\mu_r \sigma_r^4 m_2^i \\
m_6^r &= \mu_r^6 m_6^i + 15\sigma_r^6 (m_3^i + 3m_2^i m_1^i + (m_1^i)^3) + 15\mu_r^4 \sigma_r^2 (m_5^i + m_4^i m_i^i) \\
&\quad +45\mu_r^2 \sigma_r^4 (m_4^i + 2m_3^i m_1^i + m_2^i (m_1^i)^2)
\end{aligned}
\tag{2}
$$

See Harris (1987) and Richardson and Smith (1994) for example. These moment conditions are the discrete time analogues of those in Appendix A of AG.

There are a number of identification problems. Firstly, the parameters μ_r and σ_r^2 are only identified up to scale since i_t is not observed. Suppose i_t is replaced by λi_t with $\lambda > 0$ so $m_1^i, \ldots, m_6^i$ become $\lambda m_1^i, \ldots, \lambda^6 m_6^i$ in the moment conditions. Then μ_r/λ and σ_r^2/λ satisfy the new moment conditions. Li (2004) also noted this problem. Thus μ_r, σ_r^2 and m_1^i are not separately identified. One solution to this scaling problem is to normalize the mean of the unobserved information flow process m_1^i to one. It is not clear how AG deal with this issue.

Secondly, with high frequency data, returns are zero on average so, a priori, setting μ_r to zero is a plausible restriction. However, when $\mu_r = 0$, the three uneven moment conditions are identically zero so identification becomes even more problematic. Finally, even when μ_r is non zero and m_1^i is set to one, the seven remaining parameters $(\mu_r, \sigma_r^2, m_2^i, \ldots, m_6^i)$ are not identified from six moment conditions. They can only be recovered if additional restrictions are imposed or additional moment conditions are added.

AG augment the six moment conditions with a number of approximate moment generating function (MGF) conditions:

$$
E[\exp(\beta_i r_t)] \approx \exp(\alpha_i m_1^i) \left[1 + \frac{\alpha_i^2}{2} m_2^i + \frac{\alpha_i^3}{6} m_3^i + \frac{\alpha_i^4}{24} m_4^i \right], \tag{3}
$$

where $\alpha_i = \beta_i \mu_r + \frac{1}{2}\beta_i^2 \sigma_r^2$ and different values of β_i are used. This approach is similar to the exact MGF approach set out in Quandt and Ramsey (1978) and Schmidt (1982). AG do not explicitly discuss the likely approximation errors involved in these restrictions or the choice of the β_i's. We follow Ané and Geman (1996), a closely related paper, and use the values -0.7,-0.5, 0.3 and 0.6 for the β_i's.

AG suggest that the results are not sensitive to the particular choice of the β_i's as long as very large and small values are not used. Our results confirm this. However, our Monte Carlo simulations suggest that the approximate MGF conditions are not very informative about the higher moments of i_t. One reason for this is that the higher order m_3^i and m_4^i terms in AG's approximate MGF conditions are extremely small. As a result, m_3^i and m_4^i are poorly identified, if at all. Another reason is that very large sample sizes are required to accurately estimate $E[\exp(\beta_i r_t)]$ using $\frac{1}{T}\sum_t \exp(\beta_i r_t)$. In addition, $\frac{1}{T}\sum_t \exp(\beta_i r_t)$ and $\frac{1}{T}\sum_t \exp(\beta_j r_t)$ are highly collinear when β_i and β_j have the same sign. For these reasons, the use of additional approximate MGF conditions does not help all that much to identify the higher moments of i_t, especially with high frequency data when μ_r is basically zero. This is not a small sample problem.

We also consider a bivariate data generation process or DGP similar to those considered by Clark (1973), Tauchen and Pitts (1983), Harris (1987) and Richardson and Smith (1994) inter alia. Conditional on the information flow i_t, we assume that returns r_t and observed "market activity" a_t (volume, log volume, the number of trades etc.) are independently and normally distributed as:

$$\begin{pmatrix} r_t \\ a_t \end{pmatrix} \mid i_t \sim N\left(\begin{pmatrix} \mu_r i_t \\ \mu_a i_t \end{pmatrix}, \begin{pmatrix} \sigma_r^2 i_t & 0 \\ 0 & \sigma_a^2 i_t \end{pmatrix}\right), \tag{4}$$

with the means and variances of both r_t and a_t linear in i_t. There is no compelling theoretical reason to assume that market activity, however defined or transformed, is normally distributed and linear in i_t.

As before, the unconditional bivariate moments $m_{jk}^{ra} = E(r_t - Er_t)^j(a_t - Ea_t)^k$ of r_t and a_t are easily calculated. The univariate and bivariate moments can be used to identify and estimate μ_r, μ_a, σ_r^2 and σ_a^2 as well as the moments of the information flow process $m_2^i, .., m_6^i$. The bivariate approach should, assuming a correctly specified data generation process (DGP), produce more precise estimates since it exploits more information and uses exact moments. It also provides more potential over-identifying restrictions.

II.**The Monte Carlo Experiments**

In our Monte Carlo experiments, we simulate returns from three DGP's and see how well the AG procedure works. Our three DGP's are the normal lognormal, the normal inverse Gaussian and the normal gamma. All three normal mixture distributions have been widely used in the empirical finance literature. The three DGP's simply assume different (exogenous) distributions for i_t, with $r_t \mid i_t \sim N(\mu_r i_t, \sigma_r^2 i_t)$ in every case. They are all easy to calibrate and simulate from.

The normal lognormal mixture was initially used by Clark (1973) to model the returns on cotton futures. In this model, the unobserved information flow i_t is assumed to be lognormally distributed. In the normal inverse Gaussian DGP, introduced by Barndorff-Nielsen (1995), the information flow is distributed as an inverse Gaussian random variable. In the normal gamma (or variance gamma) DGP, associated with Madan and Seneta (1990), the information flow has a gamma distribution. In all three cases, the distributions of i_t depend on two parameters. However since the first moment of i_t is normalised to unity, the two parameters are not independent.

We used the following settings in the Monte Carlo's. We set μ_r equal to 0 or 0.1 and σ_r^2 equal to 0.1. The zero mean setting is probably the appropriate one to use when considering high frequency data. The parameters of the information flow distributions were chosen so that $m_2^i = 0.5$ given the normalisation $m_1^i = 1$.

The other moments of i_t vary by distribution. In our bivariate simulations, we use $\mu_a = 3$ and $\sigma_a^2 = 1.3$ for our "market activity" variable. These settings were suggested by the results of using the bivariate moment conditions to recover the moments of i_t using 10 years of Nastraq return and volume data for the Dell stock at the 5 minute frequency.

In our experiments, we restrict our attention to recovering the first four central moments of i_t, as the higher moments of i_t are more difficult to estimate precisely. The simulations and moment estimation / recovery were carried out in the econometrics package TSP (Hall and Cummins 1997). AG used a method of moments like procedure to try and recover the moments of i_t. We used the generalized method of moments or GMM procedure (Hansen, 1982) which should be more efficient, a point also noted by Li (2004). In practice, the method of moments and GMM procedures produce very similar results, so we only present GMM results here.

We use sample sizes of 500, 1000, 2500, 5000 and 10000 observations in our experiments. Sample sizes of 5000 and 10000 observations are not uncommon in studies using, respectively, daily and high frequency data. Finally, our Monte Carlo results are all based on 1000 replications.

III. **The Monte Carlo Results**

In Tables I, II and III we look at the performance of the AG univariate procedure for recovering the moments of the time changer i_t. We use the 2nd, 4th and 6th moment of the return process r_t and two approximate moments based on the MGF. The results for the Normal Lognormal DGP are set out in Table I. The results show the third and fourth order moments of the time changer / information flow i_t cannot be accurately recovered using the AG procedure. This is true even in very large samples of 10,000 observations, so the poor performance of the AG procedure is not just a small sample problem. For example, with 10,000 observations, the average value of $\widehat{m}_3^i$ is close to the true value of 0.875 in the DGP, but the associated average standard error of 0.495 is still very large. As expected, the results for m_4^i are a good deal worse than those for m_3^i. Note that the GMM test statistic gives no indication of this poor performance.

We carried our further Monte Carlo experiments to see if our poor performance finding is robust. The results for the Normal Gamma and Normal Inverse Gaussian DGP's in Tables II and III are similar to those for the Normal Log Normal GDP in Table I. We also ran some Monte Carlo experiments using DGP's (i) with non zero means for returns, (ii) rescaling returns by multiplying them by 100, (iii) with different and/or additional moment conditions and (iv), as noted above, using least squares rather than GMM to estimate the moments of the time changer. In all cases, the AG procedure cannot recover the higher moments of the unobserved time changer / information flow. The approximate MGF moment conditions are just not informative about the moments of i_t.

In Table IV we set out some representative GMM results for the case where we use the moments of returns and some activity variable (such as the number of trades or volume) to recover the moments of the time changer / information flow. The results in Table IV show that the third and fourth order moments of i_t can be recovered in this bivariate setup. Of course, the bivariate setup involves more assumptions and large samples are required to estimate the higher moments of i_t precisely. This finding holds for the two other bivariate DGP's - Normal Gamma and Normal Inverse Gaussian - which we examined.

IV. **Some Further Evidence**

To support our Monte Carlo results, we tried to recover the moments of i_t using actual data for Dell and WorldCom stocks. We used 5 minute binned Nastraq data for 68 trading days from 8th March to the 8th June 2000, which yielded 5,304 observations per stock.

Table V reports the estimated moments of the time changer / information flow i_t using AG's univariate moments and two different sets of bivariate moments using the number of trades or volume as our activity measure. The AG procedure produces large and implausible estimates of m_3^i and m_4^i, in line with our Monte Carlo results, whereas the bivariate results appear more plausible and significant. The volume and trade results are quite similar to each other, which should come as no surprise since the number of trades and volume are highly correlated. The estimated moments of i_t do not match those of either volume or the number of trades, contrary to AG's claim that the transactions clock should be based on the number of trades. The similarity of the estimated moments of i_t for both Dell and WorldCom using trades and volume is interesting.

Figure 1 and the more formal results in Table VI confirm the fact that returns conditioned by recentered volume or the number of transactions are not normally distributed.[2] The conditioned returns are more Gaussian than the raw returns but the normal distribution still provides a poor approximation. The Jarque-Bera test statistics bear this out - the null of normality is always decisively rejected. Again the evidence does not support AG's claim regarding the transactions clock. It would be nice if all the important features of the latent information flow were captured by a combination of returns and the number of trades or volume. Unfortunately, this does not appear to be the case.

V. **Conclusions**

We show that the univariate procedure used by Ané and Geman (2000) to recover the moments of the latent time changer / information flow from high frequency data is not reliable. Our Monte Carlo results show that the third and higher moments of the time changer cannot be accurately recovered using AG's procedure because the approximate MGF conditions used are not informative. Our Monte Carlo results show that bivariate procedures work fairly well assuming the mean and variance of market "activity" is linear in the time changer. We also present some empirical evidence that returns conditioned on the recentered number of trades or volume are not Gaussian.

[2] The recentred variables are scaled to have a mean of unity.

References

Ané, T. & H. Geman (1996), 'Stochastic subordination', *Risk*, **9**(9), 145–150.

Ané, T. & H. Geman (2000), 'Order flow, transaction clock, and normality of asset returns', *The Journal of Finance*, **55**(5), 2259–2284.

Barndorff-Nielsen, O.E. (1995), 'Normal inverse Gaussian processes and the modelling of stock returns', research report 300, Dept. of Theoretical Statistics, Aarhus University.

Bessembinder, H. & P. J. Seguin (1993), 'Price volatility, trading volume, and market depth: Evidence from futures markets', *Journal of Financial and Quantitative Analysis,* **28**, 21–39.

Chan, K. & W.Fong (2000), 'Trade size, order imbalance, and the volatility-volume relation', *Journal of Financial Economics,* **57**(2), 247–273.

Clark, P.K. (1973), 'A subordinated stochastic process model with finite variance for speculative prices', *Econometrica*, **41**(1), 135–155.

Easley, D. & M. O'Hara (1992), 'Time and the process of security price adjustment', *The Journal of Finance*, **47**(2), 577–605.

Easley, D., N. Kiefer & M. O'Hara (1997), 'One day in the life of a very common stock', *Review of Financial Studies*, **10**, 805–835.

Epps, W. & M. Epps. (1976), 'The stochastic dependence of security price changes and transaction volumes: implications for the mixture of distribution hypothesis', *Econometrica,* **44**(2), 305–321.

Gallant, A.R., J. Rossi & G. Tauchen (1992), 'Stock prices and volumes', *Review of Financial Studies*, **5**, 199–242.

Geman, H. (2005), 'From measure change to time changes in asset pricing', *Journal of Banking and Finance*, **29**, 2701–2722.

Hall, B.H. & C. Cummins (1997), *TSP 4.5 User's Guide*, TSP International, Palo Alto, California.

Hansen, L.P. (1982), 'Large sample properties of generalized methods of moments estimators', *Econometrica*, **50**, 1029–1054.

Harris, L. (1986), 'Cross-security tests of the mixture of distributions hypothesis', *Journal of Financial and Quantitative Analysis*, **21**, 39–46.

Harris, L. (1987), 'Transaction data test of the mixture of distributions hypothesis', *Journal of Financial and Quantitative Analysis*, **22**, 127–41.

Hasbrouk, J. (1999), 'Trading fast and slow: Security market events in real time', working paper, New York Stock Exchange.

Izzeldin, M. (2005), 'Essays on the mixture of distribution hypothesis model: Investigating model validity in high frequency data', Ph.D. thesis, Cass Business School, City University, London.

Jones, M., G. Kaul & M. Lipson (1994), 'Transactions, volume and volatility', *Review of Financial Studies*, **7**(4), 631–651.

Karpoff, J.M. (1987), 'The relation between price changes and trading volume: A survey', *Journal of Financial and Quantitative Analysis*, **22**(1), 109-126.

Karpoff, J.M. (1988), 'Costly short sales and the correlation of returns with volume', *Journal of Financial Research*, **11**, 173–188.

Lamaourex, C.G. & D. Lastrapes (1990), 'Heteroskedasticity in stock return data: Volume versus GARCH effects', *Journal of Finance* **45**, 221–229.

Lamaourex, C.G. & D. Lastrapes (1994), 'Endogenous trading volume and momentum in stock-return volatility', *Journal of Business and Economic Statistics*, **2**, 253–260.

Li, Yan (2004), 'Estimation of information time in stock returns', working paper, Department of Economics, Yale University.

Liesenfeld, R. (1998), 'Dynamic bivariate mixture models: Modelling the behavior of prices and trading volume', *Journal of Business and Economic Statistics*, **16**, 101–109.

Liesenfeld, R. (2001), 'A generalized bivariate mixture model for stock price volatility and trading volume', *Journal of Econometrics*, **104**, 141–178.

Lillo, F., J.D. Farmer & R. Mantegna (2003), 'Master curve for price-impact function', *Nature*, **421**(6919), 129–130.

Madan, D.B. & E. Seneta. (1990), 'The VG model for share market returns', *Journal of Business*, **63**, 511–524.

Monroe, I. (1978), 'Processes that can be embedded in Brownian motion', *Annals of Probability*, **6**(1), 42–56.

Quandt, R.E. & J.B. Ramsey (1978), 'Estimating mixtures of normal distributions and switching regressions', *Journal of the American Statistical Association*, **73**, 730–738.

Richardson, M. & T. Smith (1994), 'A direct test of the mixture of distribution hypothesis: measuring the daily flow of information', *The Journal of Financial and Quantative Analysis*, **29**(1), 101–116.

Schmidt, P. (1982), 'An improved version of the Quandt-Ramsey MGF estimator for mixtures of normal distributions and switching regressions', *Econometrica*, **50(**2**)**, 501–516.

Tauchen, G. & M. Pitts (1983), 'The price varability-volume relationship on speculative markets', *Econometrica,* **51(**2**)**, 485–505.

Table I

Estimated Moments of the Time Changer - Normal Lognormal DGP for Returns

Zero mean for returns ($\mu_r = 0$) and one over-identifying restriction

Sample	Ave. GMM Parameter Estimates and Std. Errors						Size of GMM Test Statistic		
Size	μ_r	σ_r^2	m_1^i	m_2^i	m_3^i	m_4^i	10% Level	5% Level	1% Level
T = 500	0.000	0.099	1.000	0.457	0.584	-248.9	11%	5%	1%
	-	(0.008)	-	(0.218)	(0.847)	(5330.2)			
T = 1000	0.000	0.099	1.000	0.471	0.671	-135.2	9%	6%	1%
	-	(0.005)	-	(0.160)	(0.560)	(3741.8)			
T = 2500	0.000	0.099	1.000	0.488	0.805	-86.8	9%	5%	1%
	-	(0.003)	-	(0.122)	(0.602)	(2565.0)			
T = 5000	0.000	0.099	1.000	0.497	0.839	15.7	9%	6%	1%
	-	(0.002)	-	(0.094)	(0.534)	(1890.6)			
T = 10000	0.000	0.099	1.000	0.500	0.898	-52.1	11%	5%	1%
	-	(0.001)	-	(0.073)	(0.495)	(1387.8)			
True Parameters	0.000	0.100	1.000	0.500	0.875	3.890	-	-	-

Notes: The moments of the time changer are estimated using GMM and AG's univariate procedure. The following five moment conditions are used - the 2nd, 4th and 6th moment of r and two approximate MGF moments with β = -0.7 and 0.6. The results are based on 1000 converged replications. The table entries are the average parameter estimates and standard errors (in parentheses) in the 1000 Monte Carlo experiments. Details of the DGP are set out in Section II of the paper.

Table II

Estimated Moments of the Time Changer - Normal Gamma DGP for Returns

Zero mean for returns ($\mu_r = 0$) and one over-identifying restriction

Sample	Ave. GMM Parameter Estimates and Std. Errors						Size of GMM Test Statistic		
Size	μ_r	σ_r^2	m_1^i	m_2^i	m_3^i	m_4^i	10% Level	5% Level	1% Level
T = 500	0.000	0.098	1.000	0.470	0.366	28.459	9%	5%	1%
	-	(0.008)	-	(0.181)	(0.422)	(4978.098)			
T = 1000	0.000	0.099	1.000	0.493	0.459	-115.658	11%	5%	1%
	-	(0.005)	-	(0.146)	(0.447)	(3622.802)			
T = 2500	0.000	0.099	1.000	0.500	0.527	16.134	9%	5%	1%
	-	(0.003)	-	(0.104)	(0.421)	(2407.032)			
T = 5000	0.000	0.100	1.000	0.489	0.445	80.300	12%	6%	1%
	-	(0.002)	-	(0.073)	(0.285)	(1674.744)			
T = 10000	0.000	0.100	1.000	0.496	0.476	-16.435	11%	6%	1%
	-	(0.001)	-	(0.054)	(0.241)	(1214.866)			
True Parameters	0.000	0.100	1.000	0.500	0.500	1.500	-	-	-

Notes: See notes to Table I.

Table III

Estimated Moments of the Time Changer - Normal Inverse Gaussian DGP for Returns

Zero mean for returns ($\mu_r = 0$) and one over-identifying restriction

Sample	Ave. GMM Parameter Estimates and Std. Errors						Size of GMM Test Statistic		
Size	μ_r	σ_r^2	m_1^i	m_2^i	m_3^i	m_4^i	10% Level	5% Level	1% Level
T = 500	0.000	0.099	1.000	0.453	0.495	86.875	11%	6%	1%
	-	(0.008)	-	(0.195)	(0.529)	(5020.108)			
T = 1000	0.000	0.099	1.000	0.467	0.548	-101.931	9%	5%	1%
	-	(0.005)	-	(0.152)	(0.475)	(3688.182)			
T = 2500	0.000	0.099	1.000	0.495	0.718	-118.704	9%	5%	1%
	-	(0.003)	-	(0.117)	(0.510)	(2543.679)			
T = 5000	0.000	0.099	1.000	0.497	0.737	69.395	10%	5%	1%
	-	(0.002)	-	(0.089)	(0.452)	(1884.414)			
T = 10000	0.000	0.100	1.000	0.500	0.747	-42.212	11%	6%	1%
	-	(0.001)	-	(0.066)	(0.367)	(1327.396)			
True Parameters	0.000	0.100	1.000	0.500	1.500	2.625	-	-	-

Notes: See notes to Table I.

Table IV

Estimated Moments of the Time Changer - Bivariate Normal Lognormal DGP for Returns and "Activity"

Zero mean for returns and one over-identifying restriction

Sample	Ave. GMM Parameter Estimates and Std. Errors								Size of GMM Test Statistic		
Size	μ_r	σ_r^2	μ_a	σ_a^2	m_1^i	m_2^i	m_3^i	m_4^i	10% Level	5% Level	1% Level
T = 500	0.000	0.096	2.980	1.818	1.000	0.439	0.745	2.508	22%	15%	7%
	-	(0.007)	(0.107)	(0.955)	-	(0.106)	(0.326)	(1.806)			
T = 1000	0.000	0.098	2.988	1.697	1.000	0.462	0.758	2.963	20%	14%	7%
	-	(0.005)	(0.076)	(0.714)	-	(0.082)	(0.268)	(1.741)			
T = 2500	0.000	0.099	2.995	1.593	1.000	0.482	0.844	3.565	17%	11%	6%
	-	(0.003)	(0.048)	(0.487)	-	(0.058)	(0.207)	(1.565)			
T = 5000	0.000	0.099	2.998	1.524	1.000	0.491	0.858	3.755	16%	11%	5%
	-	(0.002)	(0.034)	(0.357)	-	(0.043)	(0.161)	(1.377)			
T = 10000	0.000	0.099	2.998	1.525	1.000	0.495	0.863	3.773	15%	10%	5%
	-	(0.001)	(0.024)	(0.256)	-	(0.031)	(0.120)	(1.088)			
True Parameters	0.000	0.100	3.000	1.500	1.000	0.500	0.875	3.890	-	-	-

Notes: GMM results based on 1000 replications using the following seven moment conditions - the second and fourth moments of returns r, the first four moments of "activity" a and the covariance between r^2 and a.

Table V

Estimated Moments of the Time Changer i_t for Dell and WorldCom

Using the Univariate AG Procedure and Two Bivariate Procedures

Moments	Dell			WorldCom		
	m_2^i	m_3^i	m_4^i	m_2^i	m_3^i	m_4^i
Univariate moments	1.200	3.275	0.144e+11	3.953	65.409	0.478e+12
	(0.169)	(0.976)	(0.341e+10)	(0.095)	(25.115)	(0.124e+11)
Bivariate moments with volume	0.498	1.940	10.304	0.618	1.317	7.366
	(0.044)	(0.312)	(2.423)	(0.133)	(0.398)	(0.027)
Bivariate moments with transactions	0.603	1.686	10.650	0.622	1.197	5.706
	(0.043)	(0.306)	(3.427)	(0.092)	(0.279)	(1.618)
Moments of recentered volume	0.806	2.361	14.905	0.534	1.191	6.657
Moments of recentered transactions	0.611	1.707	10.851	0.390	0.773	3.488

Notes: GMM results with standard errors in parentheses. The moment conditions used are the same as those in Tables I to III. The bivariate moments are the same as in Table IV. The binned 5 minute data used are described in Section V of the paper. Volume and transactions are scaled so that they have mean of one. The sample size is 5,236.

Table VI
The Moments of Returns and Returns Conditioned by the Recentred Number of Trades and Volume

	Dell			WorldCom		
	r	$\frac{r}{\sqrt{v}}$	$\frac{r}{\sqrt{t}}$	r	$\frac{r}{\sqrt{v}}$	$\frac{r}{\sqrt{t}}$
Mean	0.0008	-0.0072	-0.0073	-0.005	-0.0076	-0.0064
Variance	0.105	0.108	0.088	0.064	0.058	0.051
Skewness	0.054	-0.147	-0.114	-0.038	0.047	0.040
Excess Kurtosis	3.774	3.138	0.579	11.864	2.762	2.637
Jera-Barque (JB) Test	3111.0	301.6	84.8	30714.2	1666.6	1519.3

Notes: r = returns, v = recentered volume, t = recentred no of transactions.
The data are the same as those in Table V.

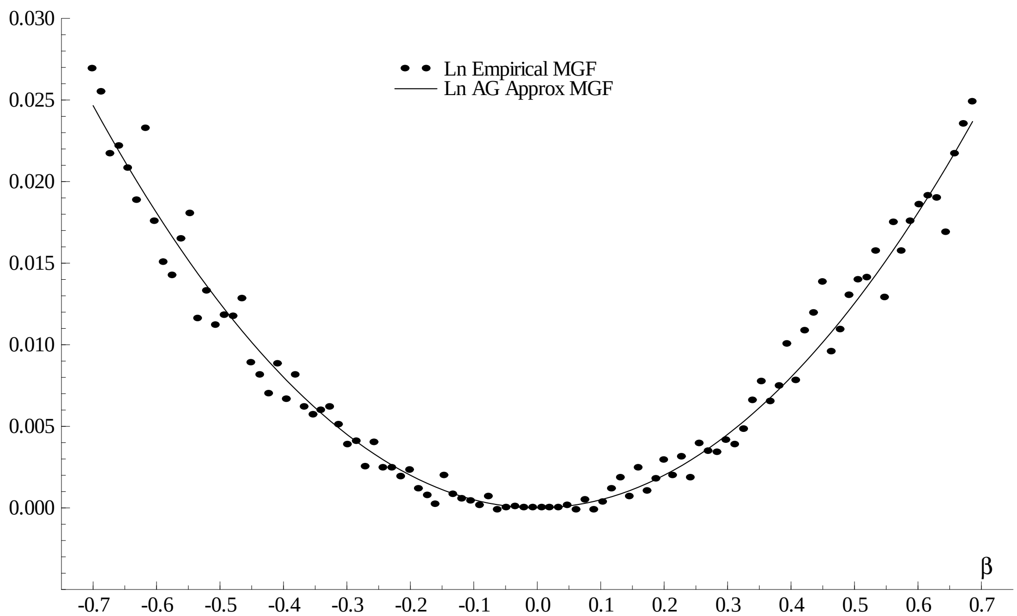

Figure 1. The AG approximate and empirical MGFs for a NIG return process with 10,000 observations. The natural logs of the empirical and approximate MGFs are plotted against β, which ranges from -0.7 to +0.7. In the DGP, $r_t \mid i_t \sim N(0, \frac{1}{10} i_t)$ and i_t is Inverse Gaussian with parameters $\mu = 1$ and $\lambda = 2$ AG's approximate MGF is $\exp(\alpha m_1^i)\left[1 + \frac{\alpha^2}{2} m_2^i + \frac{\alpha^3}{6} m_3^i + \frac{\alpha^4}{24} m_4^i\right]$, where $m_1^i ..., m_4^i$ are the moments of i_t and $\alpha = \frac{1}{2}\beta^2 \frac{1}{10}$ (since $\mu_r = 0$ and $\sigma_r^2 = \frac{1}{10}$ in this case). The empirical MGF is just $\frac{1}{T}\sum_t \exp(\beta r_t)$.

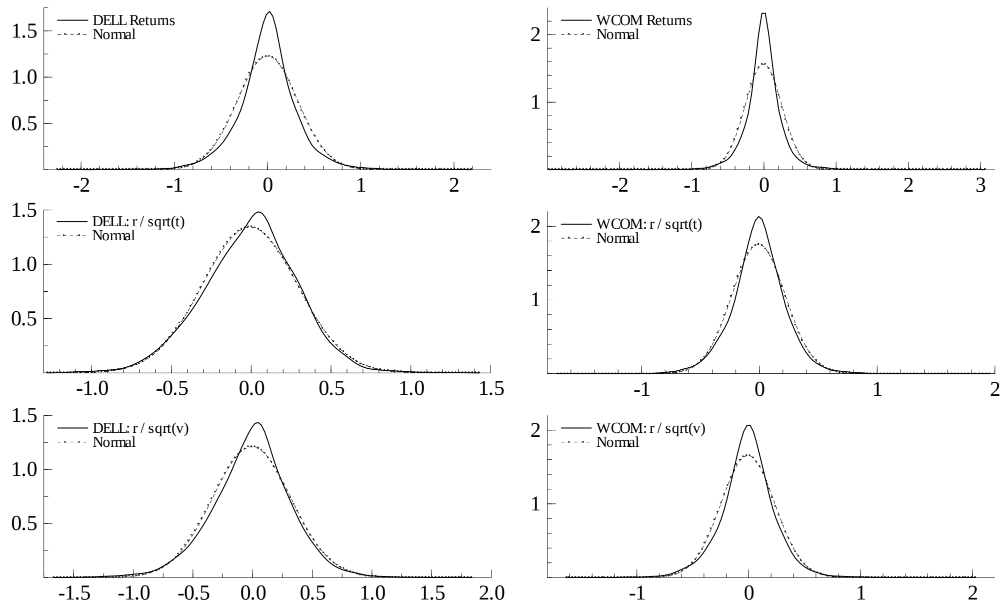

Figure 2. Estimated densities of returns and returns conditioned by the recentred numbers of trades and volume. The DELL returns are shown on the right and WCOM returns on the left. The densities of the raw returns, returns conditioned by trades and returns conditioned by volume are displayed in top, middle and bottom panels respectively. Normal distributions with the same mean and variances are also shown.