高頻度取引に対応した tobit 型 ACD モデル

株式会社 ライトストーン/専修大学経済学研究科

高　英模

はじめに

株取引における約定間隔に意味付けを与えた初期の研究に Diamond and Verrecchia (1987) や Easley and O'Hara (1992) がある．彼らは約定間隔の長さが Good News や Bad News の存在を示すものであることを Glosten and Milgrom (1985) のモデルを拡張して示した．その後, Engle and Russell (1998) は約定間隔を利用する Autoregressive Duration Model(ACD モデル) を提案し, Engle (2000) では ACD モデルと UHF-GARCH モデルを組合わせることで, 約定間隔が株価の瞬間ボラティリティに与える影響を実証分析した．

東京証券取引所では 2010 年 1 月から新世代の株式売買システム, アローヘッドによる取引が開始された．これにより我々は正確な約定時点の情報が入手できるようになった．本稿では ACD モデルとその推定上の問題点について解説し, さらにその対応方法について新たな提案を行う．

ACD モデルとは

株価収益率のボラティリティをモデル化した GARCH モデルを, 約定間隔に応用したものが ACD モデルである．Engle and Russell (1998) の提案した ACD モデルの基本的な構成を次に示す．

$$x_i = \Psi_i \epsilon_i \tag{1}$$

$$\Psi_i = \omega + \alpha x_{i-1} + \beta \Psi_{i-1} \tag{2}$$

これは ACD(1,1) モデルと呼ばれ, ϵ_i は期待値が 1, 分散 σ^2 の非負の確率分布に従う確率変数である．x_i は調整済み約定間隔, Ψ_i は x_i の条件付き期待値で期待約定間隔と呼ばれている．観測した約定間隔 τ_i には日中の時間経過とともに変化するイントラデイトレンド (図 1) が存在する．したがって, この確定的なトレンドを次のように除去してトレンド調整済みの約定間隔 x_i とする．

$$x_i = \frac{\tau_i}{\phi\left(t_{i-1}\right)} \tag{3}$$

$\phi\left(t_i\right)$ はイントラデイのトレンド関数である．

ACD モデル推定上の問題点

ACD モデルには推定上の問題点が 2 つ存在する．1 つはゼロ約定間隔の処理の問題であり, もう一つはモデルの推定手法である．

ゼロ約定間隔

約定データには約定価格や株数, 最良気配などの情報が約定時刻 t_i とともに記録されている．ACD モデルで利用するのは約定情報だけであり, 気配に関する情報は利用しない．約定間隔は単純に $\Delta t_i = t_i - t_{i-1}$ として求め, $\Delta t_i = 0$ となるデータをゼロ約定間隔 (zero duration) と呼ぶ．Engle (2000) で利用した 1990

年の IBM の株価データにはゼロ約定間隔が 13% 存在したが, 実証分析ではそれらのデータを単純に削除している．このゼロ約定間隔を無視しても本当に良いのだろうかという疑問が研究の動機である．約定間隔がゼロに等しいということは, 何らかの理由により取引が活発に行われている状態が実現していることを示している．ゼロ約定間隔を無視することは, 活発な取引の情報を敢えて捨てることになると考えられる．

Engle (2000) とその後の筆者の研究で用いた秒単位のデータにおけるゼロ約定間隔の個数を次に示す．

研究	データ	単位	ゼロ約定間隔
Engle (2000)	IBM (1990 年)	秒	13%
高 (2012)	日産自動車 (2010 年)	秒	32%
高 (2016)	JFE (2012 年)	秒	52%
	京セラ (2012 年)	秒	46%
	日産自動車 (2012 年)	秒	44%
	東京海上日動 (2012 年)	秒	51%

表 1. ゼロ約定間隔の割合

表 1 からも明らかなように, アローヘッド稼働後, 流動性の高い銘柄においては同じタイムスタンプの付いたデータが大量に発生しており, 高 (2016) では約半数の約定間隔データがゼロになっている．

パラメータ修正

高 (2016) による数値実験の結果, ゼロ約定間隔を削除した場合, シミュレーションの設定値と最も大きく乖離するのが ω であり, α と β の変化はさほど大きくないことが分かった．ここでは高 (2016) の提案した ω の修正方法を簡単に紹介する．約定間隔が 2 式に従う場合, その無条件期待値は次の式に従う．

$$\mathrm{E}\left(x\right)=\frac{\omega}{1-\alpha-\beta} \tag{4}$$

タイムスタンプが微小な単位まですべて正確に分かり, ゼロ約定間隔が存在しない場合, すべての約定情報が利用でき, 次のような関係が成り立つ．

$$\frac{\tilde{\omega}}{1-\tilde{\alpha}-\tilde{\beta}}=\frac{\sum^{full}x_i}{OBS_{full}}$$

一方, ゼロ約定間隔を削除した時は, 次のようになる．ここで $OBSz$ はゼロ約定間隔を除去した後のデータの個数を示す．

$$\frac{\hat{\omega}_z}{1-\hat{\alpha}_z-\hat{\beta}_z}=\frac{\sum^{z}x_i}{OBS_z}$$

数値実験の結果から大胆に $\tilde{\alpha}\simeq\hat{\alpha}_z, \tilde{\beta}\simeq\hat{\beta}_z$ と仮定すると,ω の修正済みパラメータは

$$\hat{\omega}_c=\hat{\omega}_z\frac{OBS_z}{OBS_{full}} \tag{5}$$

となることが分かる．つまり, ゼロ約定間隔を削除した時の推定値である $\hat{\omega}_c$ に, (ゼロ約定間隔を除いたデータの個数)/(全約定間隔データの個数) を掛けることで $\hat{\omega}$ についての近似が可能となる．

同時推定

Engle (2000) などでは, 3 式の $\phi\left(t_i\right)$ に線形スプライン関数 ($\phi\left(t_i\right)=a+bt_i$) などを利用してトレンドを除去した後, 独立して 2 式と 3 式を推定している．しかし, 2 つの式に時点 t_i の情報が含まれることから, 高 (2016) では次の尤度関数を用いて両者を同時推定した．

$$\Psi_i = \omega + \alpha \left(\frac{\tau_{i-1}}{\phi\left(t_{i-2}\right)} \right) + \beta \Psi_{i-1} \tag{6}$$

$$LL = -\sum \left\{ \frac{\tau_i}{\Psi_i \phi\left(t_{i-1}\right)} + \log \Psi_i + \log \phi\left(t_{i-1}\right) \right\} \tag{7}$$

x_i は 4 式から ω を修正すると小さくなることが明らかであるが, これは 5 式の関係を使って OBS_z/OBS_{full} を掛けて単純に小さくする．一方, 同時推定が完了した時点で得られる期待約定間隔 $\hat{\Psi}_i$ は 2 式を満たすものなので, 修正済みパラメータ $\hat{\omega}_c$ と修正した x_i を用いて逐次計算によって求める．

実証分析

Engle (2000) はイントラデイの瞬間ボラティリティを推定する UHG-GARCH モデルを提案した．そこでは $\hat{\Psi}_i$ や x_i を分散方程式で用いて, イントラデイボラティリティと約定間隔の関係を仮説検定している．瞬間ボラティリティの定義は次の通りである．

$$V_{i-1}\left(\frac{r_i}{\sqrt{x_i}} | x_i \right) = \sigma_i^2$$

r_i はティックデータにおける約定間の収益率である．例えば, 実際のモデル推定では平均方程式と分散方程式を次のようにして利用する．

$$r_i/\sqrt{x_i} = kx_i + \rho r_{i-1}/\sqrt{x_{i-1}} + e_i + \phi e_{i-1} \tag{8}$$

$$\sigma_i^2 = \omega + \alpha e_{i-1}^2 + \beta \sigma_i^2 + \gamma_1 \left(1/x_i\right) + \gamma_2 \left(x_i/\Psi_{i-1}\right) \tag{9}$$

高 (2016) ではパラメータ修正の有無によって ACD モデルの $\omega, x_i, \hat{\Psi}_i$ が変化することで, $k, \alpha, \beta, \gamma_1, \gamma_2$ の大きさやその有意性が影響される事を示した．

より高精度なデータ

ここまでの解説はすべて秒単位のデータを用いた分析に関するものである．現在ではより精度の高いデータ (ミリ秒単位) も入手可能となっている．そこで, ミリ秒単位のデータを利用した分析例[1]を紹介する．利用したデータは 2013 年 4 月 1 日から同年 6 月 29 日までの日経平均先物である．取引は日中立会と夜間立会に分かれているが, ここでは約定間隔の日中立会のデータのみを利用した．この間の約定間隔の個数とそこに占めるゼロ約定間隔の個数を次に示す．

	全約定間隔	約定間隔	ゼロ約定間隔
日中立会	644,953	513,451	131,502(20.4%)

表 2. 日経平均先物のゼロ約定間隔

ミリ秒単位のデータであるが, 20% 以上のゼロ約定間隔が発生している．この時の約定間隔の 5 分平均を図 1 として示す．

[1]研究協力: 乾　孝治教授, 明治大学総合数理学部

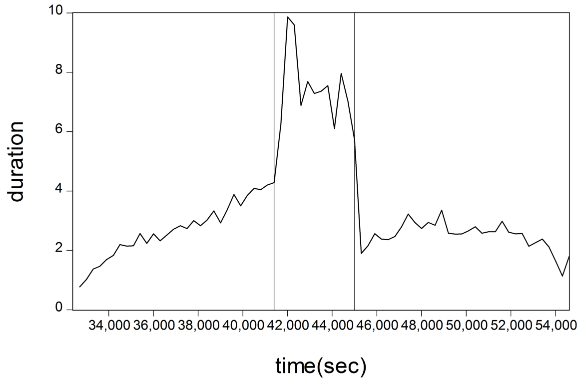

図 1. 約定間隔の 5 分平均 (日中立会)

取引開始直後が最も活発であり, 東証の昼休みにかけて徐々に約定間隔が長くなっている. 昼休みの間の取引間隔は約 2 倍程度に間延びし, 後場の開始とともに先物取引も活発化している様子が見てとれる. このデータに対し, 回帰スプラインとワイブル分布による ACD モデルの同時推定を行うと, トレンド調整後の約定間隔は図 2 のようになる.

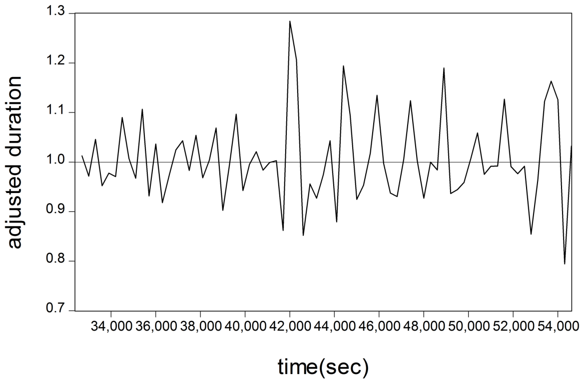

図 2. トレンド調整後の約定間隔

約 20% のゼロ約定間隔を削除し, 1 式における ϵ の密度関数として指数分布を用いた時の ACD モデルの推定結果は次の通りである.

	ω	α	β	LL	OBS.
Coef	0.044	0.261	0.728	-777014	513451
S.E	0.002	0.000	0.000		

表 3. ゼロ削除の推定結果

この推定に利用した尤度関数は次の通りである. 計算上の工夫のため, トレンド関数で利用する時点情報は 7 式における t_{i-1} から t_i に変更した.

$$LL = -\sum \left\{ \frac{\tau_i}{\Psi_i \phi\left(t_i\right)} + \log \Psi_i + \log \phi\left(t_i\right) \right\} \tag{10}$$

tobit タイプの ACD モデル

ミリ秒単位のデータでのゼロ約定間隔とは実際に 0.001 より小さな約定間隔である．そこで, パラメータ修正に代わる方法として tobit モデルの考え方を ACD モデルに応用し, 閾値となる約定間隔 $\Delta t'$ を設定し, 次のような尤度関数を考える．

$$LL = \sum \left\{ \log \left(1 - \exp \left(-\frac{\tau_i}{\Psi_i \phi(t_i)} \right) \right) \right\} I_1 - \sum \left\{ \frac{\tau_i}{\Psi_i \phi(t_i)} + \log \Psi_i + \log \phi(t_i) \right\} I_2$$

I_i は約定間隔が Δt 以下の場合に 1 をとり, それ以外ではゼロとなるダミー変数である．I_2 はその逆である．

	ω	α	β	LL	OBS.
Coef	0.022	0.260	0.735	-2227331	644953
S.E	0.000	0.000	0.000		

表 4.tobit タイプの ACD モデル

高 (2016) では簡便な修正方法で ω のみ修正したが, tobit タイプの ACD モデルを利用することで, ω だけでなく α と β を更新することができた．α はゼロ削除にくらべて僅かに小さく, β は逆にやや大きくなっており, これは高 (2016) が数値実験で示した結果と整合的である．

ここに示した tobit タイプの ACD モデルはミリ秒単位のデータにおけるゼロ約定間隔への対応方法である．約定間隔の閾値となる微小な Δt をゼロ約定間隔のセルに代入しても, 約定間隔の自己相関構造が殆ど影響されないところがポイントである．しかし, 秒単位のデータでは Δt の閾値が大きすぎて, 約定間隔の自己相関構造が変化してしまうので, この方法は適用できない.

まとめ

本稿の目的は ACD モデルの概略を解説し, 推定上の問題点を解決するための新たな提案を示すことである．ミリ秒単位のデータであれば, tobit モデルの考え方を応用することで, ゼロ約定間隔の問題を解決できることを示した．今後の研究は, tobit 型 ACD モデルの推定結果を用いて, UHF-GARCH モデルの推定を行い, 高頻度取引で利用可能なイントラデイボラティリティを正確に求めることである.

参考文献

高　英模 [2012]．“高速取引下の約定間隔と価格形成”．早稲田大学大学院ファイナンス研究科修士論文.

高　英模 [2016]．“ACD モデルにおけるパラメータ修正の効果”．横浜国立大学国際社会科学府 博士論文.

Bauwens,Luc and Pierre Giot.[2001].“Econometric modeling of stock market intraday activity”.Kluwer Academic Publishers.

Engle,Robert and Jeffrey R.Russell.[1998].“Autoregressive conditional duration:A new model for irregularly spaced transaction data”.Econometrica, (66).

Engle,Robert[2000]. “The econometrics of ultra-high-frequency data”.Econometrica, (68).

Tsay, Ruey.[2010].“Analysis of financial time series”.Wiley.