# 3. Other models for high-frequency trade volume

MFE Elective, Forecasting & Financial Time Series
Part I: High-Frequency Forecasting

Ryoko Ito[1]

Department of Economics and Nuffield College, Oxford University
Email: ryoko.ito@economics.ox.ac.uk
Website: www.itoryoko.com

Saïd Business School, Oxford, Trinity 2017



## Popular HF volume prediction method

Typical and popular ways of modelling HF trade volume were (are):

- Take average trade volume for each intra-day bin.
- Use of simple ARMA based models.
- Use Fourier approximation or a spline function to capture overall intra-day dynamics.
- Rolling average of nearest observations.
- Model announcement effect using historical data.

Some trading institutions still use one or a combination of these ad-hoc methods.

Largely due to the statistically highly complex features of HF trade volume.

# Popular HF volume prediction method

The primary objective of these (and any other) volume prediction methods is to accurately capture the pattern of intra-day periodicity.

- Irregular deviation from intra-day patterns are difficult to predict in some sense.

A new development was the **component-MEM** (CMEM) model of Brownlees et al. (2011).[1]

It is currently the state-of-the-art volume prediction model in the HF literature.

[1] Brownlees, C. T., Cippolini, F., and Gallo, G. M. (2011), "Intra-Daily Volume Modelling and Prediction for Algorithmic Trading," *Journal of Financial Econometrics*, 9, 489-518.



# Average trade volume per bin

Volume averaging by intra-day bin:

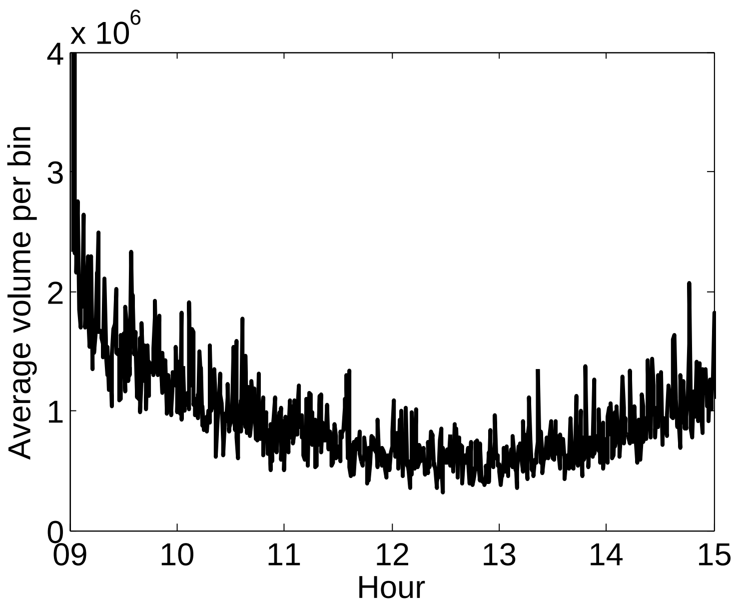


**Figure:** Average IBM trade volume per intra-day bin between Mon 28 Feb and Fri 31 Mar 2000. The sampling frequency is 30 seconds.

# Average trade volume per bin

Volume averaging by intra-day bin:

Almost U-shape but fluctuates.
The fluctuation is caused by daily irregular dynamics.
Large spikes can be observed around the moments the announcement effect dominates (e.g. FX volume).



# ARMA for trade volume

Classical ARMA type models:
They cannot capture data with one-sided distribution unless we take log(Trade Volume).

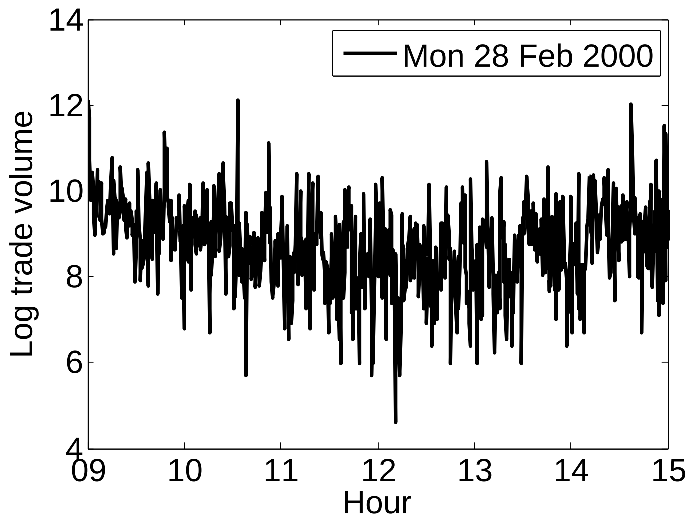


**Figure:** Log IBM trade volume on Mon 28 Feb 2000. The sampling frequency is 30 seconds.

# ARMA for trade volume

Need a component for intra-day U-shaped patterns.

Observe that Spline-DCS is like an ARMA-type component model for log trade volume.

$$\log(y_{t,\tau}) = \lambda_{t,\tau} + \log(\varepsilon_{t,\tau}).$$

Note the distribution of $\log(\varepsilon_{t,\tau})$ is the exponential GB2 distribution. (i.e. $\log(\varepsilon_{t,\tau})$ is GB2 after taking exponential transformation.)

$\Rightarrow$ Parameter restrictions may be needed to ensure $\mathbb{E}[\log(\varepsilon_{t,\tau})] = 0$ to be comparable with standard ARMA.



# Component-MEM

We adhere to the notations of Brownlees et al. (2011).
CMEM is:

$$\begin{aligned}
y_{t,\tau} &= \eta_t \phi_\tau \mu_{t,\tau} \varepsilon_{t,\tau}, \quad \varepsilon_{t,\tau} \text{ i.i.d.} \sim (1, \sigma^2), \\
\eta_t &= \alpha_0^{(\eta)} + \beta_1^{(\eta)} \eta_{t-1} + \alpha_1^{(\eta)} y_{t-1}^{(\eta)} \\
\mu_{t,\tau} &= \alpha_0^{(\mu)} + \beta_1^{(\mu)} \mu_{t,\tau-1} + \alpha_1^{(\mu)} y_{t,\tau-1}^{(\mu)} + \alpha_2^{(\mu)} y_{t,\tau-2}^{(\mu)} \\
\phi_{\tau+1} &= \exp\left\{ \sum_{k=1}^{\lfloor I/2 \rfloor} \left[ \delta_{1k} \cos\left(\frac{2\pi k \tau}{I}\right) + \delta_{2k} \sin\left(\frac{2\pi k \tau}{I}\right) \right] \right\}.
\end{aligned}$$

$\eta_t$ is the daily component.
$\mu_{t,\tau}$ is the intra-day non-periodic component.
$\phi_\tau$ is the intra-day periodic component. This approximates periodic patterns using the Fourier series.

# Component-MEM

$y_t^{(\eta)}$ is the *standardized* daily volume.
$y_{t,\tau}^{(\mu)}$ is the *standardized* intra-daily volume.

$$y_t^{(\eta)} = \eta_t I^{-1} \sum_{\tau=1}^{I} \varepsilon_{t,\tau}, \quad \text{and} \quad y_{t,\tau}^{(\mu)} = \mu_{t,\tau}\varepsilon_{t,\tau}.$$

These make the model analogous to the GARCH model in the return volatility literature. $\eta_t$ and $\mu_{t,\tau}$ are GARCH-like filters because

$$\mathbb{E}[y_{t,\tau}^{(\mu)} | \mathcal{F}_{t,\tau-1}] = \mu_{t,\tau}, \qquad \mathrm{Var}[y_{t,\tau}^{(\mu)} | \mathcal{F}_{t,\tau-1}] = \mu_{t,\tau}^2 \sigma^2,$$

and likewise for $y_t^{(\eta)}$ (with the normalization factor $1/I$ for the variance).
Then the conditional moment of volume in the $\tau$-th bin on the $t$-th observation day is $\mathbb{E}[y_{t,\tau} | \mathcal{F}_{t,\tau-1}] = \eta_t \phi_\tau \mu_{t,\tau}$.



# Component-MEM: remark on GMM

Important distinction from Spline-DCS:

CMEM is proposed to be estimated by the generalized method of moment (GMM).

GMM allows for a greater flexibility in the distribution of the error term.

- We don't assume anything about the distribution of $\varepsilon_{t,\tau}$.
- Only assume the first two moments of $\varepsilon_{t,\tau}$.
- Idea: try to match the first two moment conditions with the data.

Consistency and asymptotic normality of the resulting estimators can be still achieved.

## Component-MEM: remark on GMM

Let $\varphi$ denote the vector of all parameters of CMEM.

The GMM estimator, $\widehat{\varphi}_{IT}$, of $\varphi$ solves the MM equation

$$\frac{1}{IT}\sum_{t=1}^{T}\sum_{\tau=1}^{I} \boldsymbol{a}_{t,\tau} u_{t,\tau} = \boldsymbol{0},$$

where

$$\boldsymbol{a}_{t,\tau} = \eta_t^{-1}\nabla_\varphi \eta_t + \mu_{t,\tau}^{-1}\nabla_\varphi \mu_{t,\tau} + \phi_\tau^{-1}\nabla_\varphi \phi_\tau + e_{t,\tau}^{*-1}\nabla_\varphi e_{t,\tau}^{*},$$
$$u_{t,\tau} = y_{t,\tau}/(\eta_t \phi_\tau \mu_{t,\tau} e_{t,\tau}^{*}) - 1.$$



## Component-MEM: remark on GMM

Remark on GMM and the number of parameters:

Notice that this MM criterion includes...

$$\nabla_\varphi \eta_t,\ \nabla_\varphi \mu_{t,\tau},\ \text{and}\ \nabla_\varphi \phi_\tau.$$

They are the gradient vectors of each component with respect to the parameters.

$\Rightarrow$ During a numerical optimization routine, the number of recursive dynamic equations to evaluate **at each parameter value being tested** is:

(the number of components) $\times$ (1 + the number of parameters).

This can be large when the number of intra-day bins is high due to the Fourier component.

# Component-MEM: remark on GMM

The FX volume data is collected 24 hours a day, the total number of intra-day data is 144 at 10-minute frequency.

Then, the number of Fourier coefficients become

$$\lfloor 144/2 \rfloor \times 2 - 1 = 143$$

for the sine and cosine terms combined.

This means we have a large number of recursive equations to evaluate in the MM criterion.



# Idea behind CMEM

What does CMEM estimated by GMM capture?

Recall that trade volume series look like this:

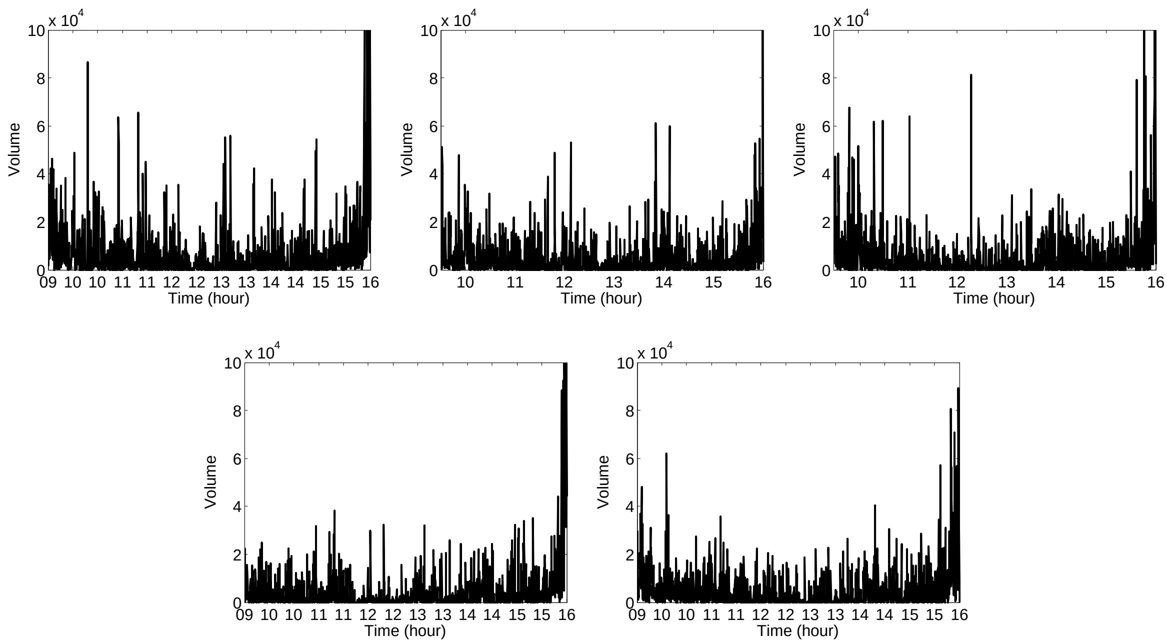


**Figure:** GE trade volume between Monday 1 June 2009 (top left) and Friday 5 June 2009 (bottom right). The sampling frequency is 15 seconds.

# Idea behind CMEM

CMEM estimated by GMM models the path of the conditional moment of the data.
(Compare with the quasi-maximum likelihood (QMLE) in GARCH with Gaussian likelihood.)

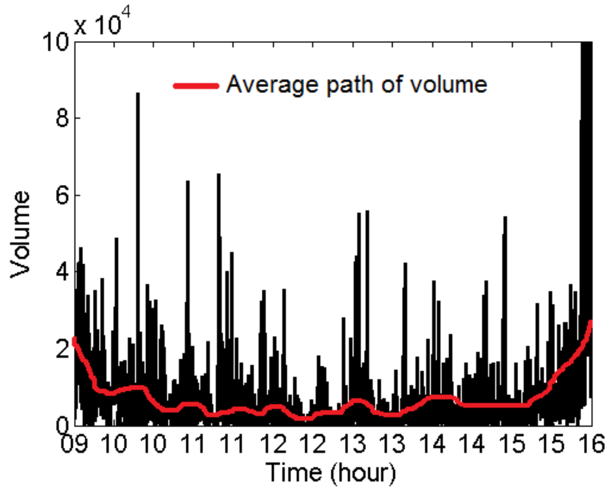


**Figure:** Picture illustration of modelling $\mathbb{E}[y_{t,\tau}|\mathcal{F}_{t,\tau-1}]$. GE trade volume on Monday 1 June 2009 (black line) and an imaginary (hand-drawn) mean path of the data (red line). The sampling frequency is 15 seconds.



# Idea behind CMEM

An advantage of GMM:

We may want to impose very little assumption about the overall distribution of the data when the data is highly non-Gaussian (e.g. asymmetric, heavy-tailed etc).

A disadvantage of GMM:

With a non-Gaussian data, the first two moments convey little information about the overall shape of the distribution.

$\Rightarrow$ If we want to know other quantities like risk statistics (e.g. value at risk, expected shortfall) or different quantiles or moments, different models and separate estimations will be required.

# Spline-DCS: remark on ML

Recall a key feature of Spline-DCS:

The distribution of $\varepsilon_{t,\tau}$ is fully specified. Estimate the distribution parameters to get the shape.

- It is a model for capturing the overall shape of the empirical distribution of the data.
- Estimation by the method of maximum likelihood (ML).



# Spline-DCS: remark on ML

ML is very easy to implement. Only need to specify the log-likelihood as the objective function to optimize.

Computing time typically taken to estimate Spline-DCS by ML is typically very short.

Model selection can be done by AIC and SIC. These are also easy to compute.

If we can find a distribution that captures the empirical distribution of the data well, this approach is useful. From density estimates and forecasts, we can:

- Derive the implied forecasts of moments and other statistical quantities.
- Obtain useful information about the degree of uncertainty and risk.

# Spline-DCS: remark on ML

A main objection to ML is that we often don't know the properties of the estimator if we get the distribution assumption wrong.

Exceptions to this situation include the quasi-ML (QML) in models like GARCH with the Gaussian likelihood.

But if a non-Gaussian likelihood is used to estimate models like GARCH and DCS by ML, the properties of the resulting QML estimators can be difficult to establish.

$\Rightarrow$ In the ML approach, we must check how well the distribution assumption for $\varepsilon_{t,\tau}$ is satisfied by inspecting the estimation residuals (denoted by $\widehat{\varepsilon}_{t,\tau}$, say).