2. Modelling high-frequency financial data

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

Features of trade volume

The empirical features of the data suggest that a good model should reflect:

- non-negative distribution
- heavy-tail
- overnight discontinuity
- intra-day seasonality
- non-periodic dynamics (i.e. autocorrelation structure)
- frequency mass of zero-valued observations

amongst other features.

Notations

Time index:

- Divide each trading day into $I \in \mathbb{N}_{>0}$ intra-day bins.
- $\cdot_{t,\tau}$ denotes an observation at the τ-th intra-day bin on the t-th trading day in the sample.
- $\tau = 0, \ldots, I$ and $t = 1, \ldots, T$.
- The market opens at $\tau = 0$ and closes at $\tau = I$.
- Set $y_{t,I} = y_{t+1,0}$ for all t.
  $\Rightarrow \tau = 1$ is the location of the first aggregated observation for each trading day.

Then, over a given T consecutive trading days, we have $I \times T$ observations.
We use the following set notations;

$$\begin{aligned} \Psi_{T,I} &= \{(t,\tau) \in \{1,2,\ldots,T\} \times \{1,2,\ldots,I\}\}, \\ \Psi_{T,I>0} &= \{(t,\tau) \in \{1,2,\ldots,T\} \times \{1,2,\ldots,I\} : y_{t,\tau} > 0\}. \end{aligned}$$

Applying DCS

Recall that the DCS(1,1) model is:

$$\begin{aligned} y_{t,\tau} &= s_{t,\tau} z_{t,\tau}, \quad z_{t,\tau} \sim \text{i.i.d. } F(\cdot) \\ s_{t,\tau} &= \text{link}(\lambda_{t,\tau}) \\ \lambda_{t,\tau} &= \delta + \phi \lambda_{t,\tau-1} + \kappa u_{t,\tau-1} \\ u_{t,\tau} &= \frac{\partial \log f_y(y_{t,\tau})}{\partial \lambda_{t,\tau}} \Bigg/ \mathbb{E}\left[- \frac{\partial^2 \log f_y(y_{t,\tau})}{\partial \lambda_{t,\tau}^2} \middle| \mathcal{F}_{t,\tau-1} \right] \end{aligned}$$

for each $(t,\tau) \in \Psi_{T,I}$. The observations $(y_{t,\tau})_{(t,\tau)\in\Psi_{T,I}}$ have the conditional distribution $F_y(\cdot)$ (c.d.f) and the density (p.d.f.) $f_y(\cdot)$.

(Note: I change the notation of the link parameter from $h_{t,\tau}$ to $\lambda_{t,\tau}$ for convenience.)

Applying DCS: GB2

For trade volume, our $y_{t,\tau}$ is non-negative.

- Choose a non-negative distribution for $z_{t,\tau} \sim$ i.i.d. $F(\cdot)$.
- The spikes in the empirical distribution are difficult to capture. $\Rightarrow$ Choose the aggregation interval wide enough so that a continuous distribution can be used.

A very useful non-negative distribution:
the generalized beta distribution of the second kind (GB2).

The p.d.f. is:

$$f(x;\nu,\xi,\zeta) = \frac{\nu x^{\nu\xi-1}(x^{\nu}+1)^{-\xi-\zeta}}{B(\xi,\zeta)}, \qquad x > 0, \text{ and } \nu,\xi,\zeta > 0$$

where $B(\cdot,\cdot)$ denotes the Beta function.

Applying DCS: GB2 and other distributions

GB2 nests (or is closely related to) some famous distributions.

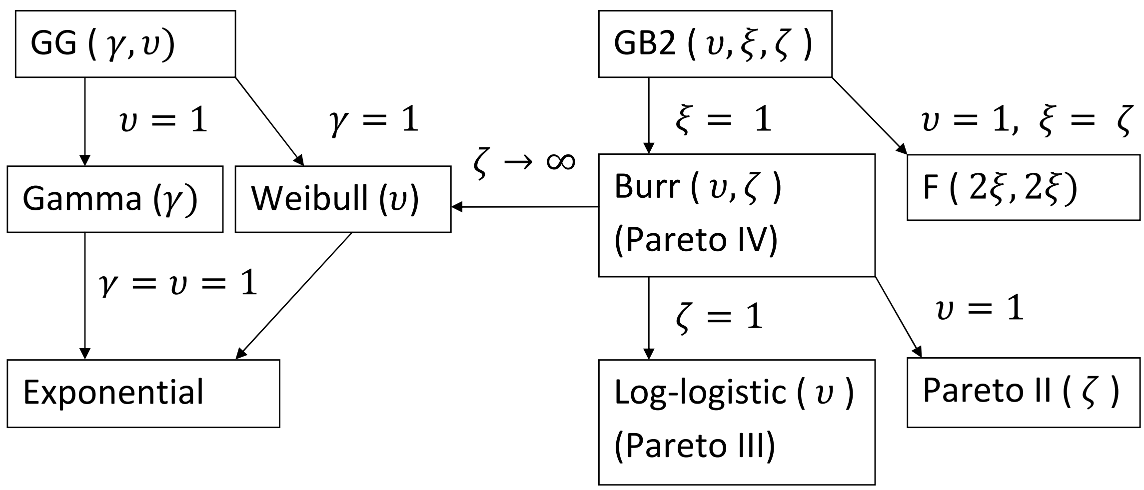

Figure: Nested diagram of some of the useful non-negative distributions. Scale factor is assumed to be one in all cases.

Applying DCS: the link function

The dynamic scale parameter $s_{t,\tau}$ has to be positive for all (t, τ).

Convenient to use the exponential link function:

$$s_{t,\tau} = \exp(\lambda_{t,\tau})$$

so that $s_{t,\tau} > 0$ for any $\lambda_{t,\tau} \in \mathbb{R}$.

Applying DCS: multiplicative error model revisited

Idea behind MEM: we want to estimate and forecast the sequence of standardization factors $(s_{t,\tau})_{(t,\tau)\in\Psi_{T,I}}$ such that the standardized observations,

$$z_{t,\tau} = y_{t,\tau}/s_{t,\tau},$$

are i.i.d.

Simulation example when $z_{t,\tau}$ is a two-sided random variable (as in volatility models):

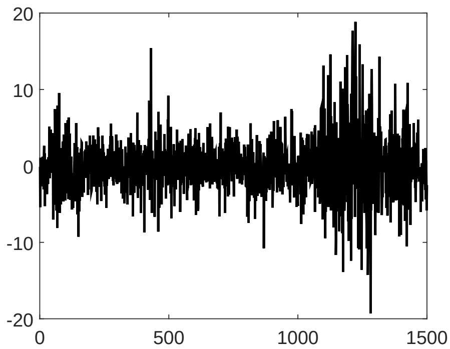

Figure: $y_{t,\tau}$

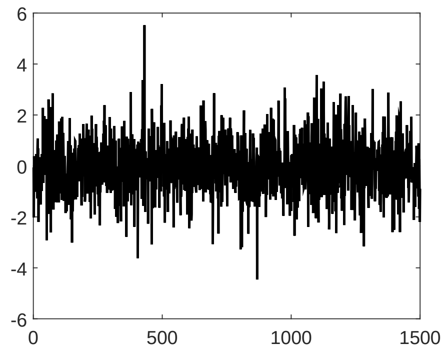

Figure: $z_{t,\tau} = y_{t,\tau}/s_{t,\tau}$

Applying DCS: multiplicative error model revisited

The idea is the same for trade volume.

We model $s_{t,\tau}$ that expands or shrinks the support of the distribution of $z_{t,\tau}$. When $\lambda_{t,\tau}$ is large, it is more likely to observe large $y_{t,\tau}$ etc.

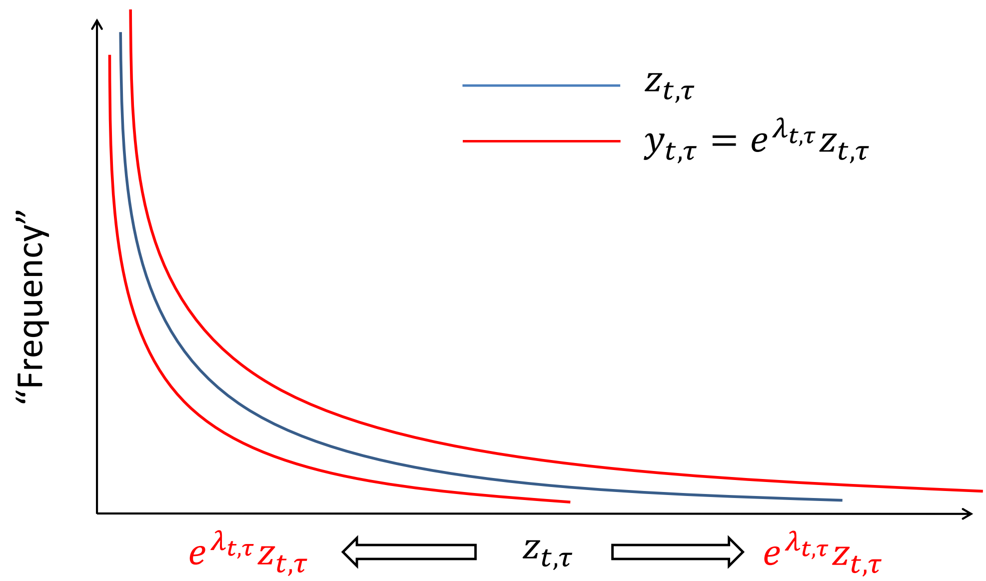

Figure: Picture illustration of DCS for volume

Applying DCS: zero-valued observations

Want to capture the frequency mass of zero-valued observations.

Define a c.d.f $F : \mathbb{R}_{\geq 0} \to [0, 1]$ of a standard random variable $X \sim F$ such that

- the origin has a discrete probability mass $p \in [0, 1]$.
- strictly positive support is captured by a conventional continuous distribution denoted by $F^* : \mathbb{R}_{>0} \to (0, 1]$ with the constant parameter vector θ^*.

Formally,

$$\mathbb{P}_F(X = 0) = p, \quad \mathbb{P}_F(X > 0) = 1 - p,$$
$$\mathbb{P}_F(X \leq x | X > 0) = F^*(x), \quad x > 0.$$

[c.f. Decomposition techniques of Amemiya (1973), Heckman (1974), McCulloch and Tsay (2001) Rydberg and Shephard (2003), Hautsch, Malec, and Schienle (2010).]

Applying DCS: zero-valued observations

- Our p is constant. This is ok if the number of zero-observations is small.
- When the sampling frequency is high, p becomes time-varying. (The moments of no trade are more likely during lunch than during busy morning or afternoon hours.)
  - Possible extension: time-varying $p_{t,\tau}$ using logit link [Rydberg and Shephard (2003), Hautsch, Malec, and Schienle (2010).]
- With this hybrid distribution, the DCS dynamics responds *only* to positive observations.
  - $u_{t,\tau}$ is the score of $F^*(\cdot)$, which is GB2.
  - Set $u_{t,\tau} = -\nu\xi$ when $y_{t,\tau} = 0$.

Applying DCS: ML estimation

All of the parameters of the model are estimated by ML.

The discussion thus far means that the joint log-likelihood function for the set of observations $(y_{t,\tau})_{(t,\tau)\in\Psi_{T,I}}$ is:

$$\log L = A\log(1-p)+(T\times I-A)\log(p)+\sum_{(t,\tau)\in\Psi_{T,I>0}}\log f_y^*(y_{t,\tau};\lambda_{t,\tau},\boldsymbol{\theta}^*),$$

where $A = |\Psi_{T,I>0}|$.

It is easy to check that the ML estimator (MLE) of p is $\widehat{p} = (T\times I - A)/(T\times I)$.

Applying DCS: the score of GB2

If a *non-standardized* (non-zero) observation $y_{t,\tau}$ follows the GB2 distribution given $\mathcal{F}_{t,\tau-1}$, its p.d.f $f_y^* : \mathbb{R}_{>0} \to \mathbb{R}$ with the scale parameter $s_{t,\tau} > 0$ is

$$f_y^*(y_{t,\tau}; s_{t,\tau}, \nu, \xi, \zeta) = f^*(y_{t,\tau}/s_{t,\tau}; \nu, \xi, \zeta)/s_{t,\tau}, \quad y_{t,\tau} > 0.$$

Recall that we set: $s_{t,\tau} = \exp(\lambda_{t,\tau})$. The log-likelihood function of a single (non-zero) observation $y_{t,\tau}$:

$$\log f_y^*(y_{t,\tau}; \lambda_{t,\tau}, \boldsymbol{\theta}^*) = \log(\nu) - \nu\xi\lambda_{t,\tau} + (\nu\xi - 1)\log(y_{t,\tau}) - \log B(\xi, \zeta) - (\xi + \zeta)\log[(y_{t,\tau}/\exp(\lambda_{t,\tau}))^\nu + 1].$$

(f_y^* is the p.d.f. of the centered non-standardized GB2 distribution.)

Applying DCS: the score of GB2

Then it is easy to check that

$$\frac{\partial \log f_y^*(y_{t,\tau})}{\partial \lambda_{t,\tau}} = \frac{\nu(\xi + \zeta)(y_{t,\tau}e^{-\lambda_{t,\tau}})^\nu}{(y_{t,\tau}e^{-\lambda_{t,\tau}})^\nu + 1} - \nu\xi = \nu(\xi + \zeta)b_{t,\tau} - \nu\xi$$

where we used the notation $b_{t,\tau} \equiv (y_{t,\tau}e^{-\lambda_{t,\tau}})^\nu/((y_{t,\tau}e^{-\lambda_{t,\tau}})^\nu + 1)$.

- $b_{t,\tau} \in (0, 1)$.
- By the property of the GB2 distribution, we know that $b_{t,\tau}$ i.i.d. and follows the beta distribution with parameters ξ and ζ.

Applying DCS: the score of GB2

The Fisher information quantity is a constant number, so it gets absorbed by κ.

So we have

$$u_{t,\tau} = \begin{cases} \nu(\xi+\zeta)b_{t,\tau} - \nu\xi & \text{if} \quad y_{t,\tau} > 0, \\ -\nu\xi & \text{if} \quad y_{t,\tau} = 0. \end{cases}$$

It is easy to check that $\mathbb{E}[u_{t,\tau}|y_{t,\tau} > 0] = 0$ and $-\nu\xi \leq u_{t,\tau} \leq \nu\zeta$. (The source of the robustness of the model.)

Applying DCS: capturing autocorrelation

The data showed intra-day periodic patterns, and some irregular movements around them.

The model assumes that z_t is i.i.d.

$\Rightarrow$ The periodic patterns and the autocorrelation structure in the data are entirely due to the dynamics in $s_{t,\tau}$.

$\Rightarrow$ Let $s_{t,\tau} = \exp(\lambda_{t,\tau})$ have a component structure to capture the autocorrelation structure of the data.

$$\lambda_{t,\tau} = \text{non-periodic}_{t,\tau} + \text{periodic}_{t,\tau}$$

Applying DCS: periodic component

Model the periodic structure of the data by a cubic spline function (see Harvey and Koopman (1993)).
Denote the periodic component by $\varphi_{t,\tau}$. The cubic spline function is:

$$\varphi_{t,\tau} = \sum_{j=1}^{k} \mathbb{1}_{\{\tau \in [\tau_{j-1}, \tau_j]\}} \, \underline{z}_j(\tau) \cdot \underline{\gamma}$$

- k: number of knots.
- $\tau_0 < \tau_1 < \cdots < \tau_k$: coordinates of the knots along the intra-day time-axis
- $\underline{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_k)^\top$: y-coordinates (height) of the knots
- $\underline{z}_j : [\tau_{j-1}, \tau_j]^{k+1} \to \mathbb{R}^{k+1}$: $k+1$-dimensional vector of functions. Conveys information about the (i) polynomial order, (ii) continuity, (iii) periodicity, and (iv) zero-sum conditions.
- Bowsher and Meeks (2008): "special type of dynamic factor model"

Applying DCS: cubic spline function

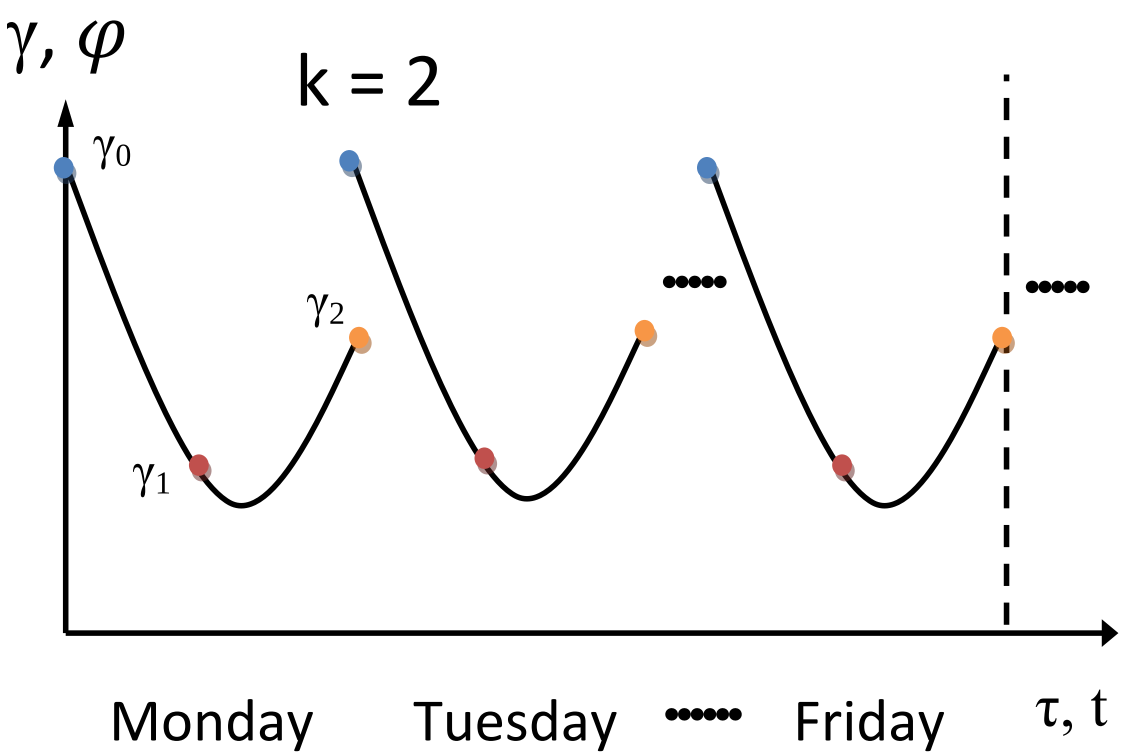

Figure: Picture illustration of static daily spline.

Applying DCS: cubic spline function

See the appendix document that formally defines this cubic spline function.

- The elements of $\underline{\gamma}$ are additional parameters to be estimated.
- τ_0 is at the start of each day. τ_k is at the end of each day.
- The location of $\tau_1, \ldots, \tau_{k-1}$ and the size of k are chosen by trial and error.

Note: by allowing for a discrepancy in the spline between the last knot and the first knot of any two consecutive days, this spline allows for the overnight effect.

(i.e. capturing the impact of overnight news flow, volume can start from a level different from the level the night before.)

Specifying knots

The location of knots, $\tau_1, \ldots, \tau_{k-1}$, and the size of k depend on things like:

- The empirical shape of periodicity.
- The number of intra-day observations.

Increasing k does not necessarily improve the fit of the model.

Using too many knots deteriorates the speed of computation.

Specifying knots

The location and the number of knots play a key role in achieving good estimation outcomes.

The following rules of thumb works well:

1. Place one knot approximately every 1 hour to 1.5 hours along the intra-day time axis.
2. Place relatively more knots around the hours in which the intensity of trading activity changes fast.
   - These hours correspond to the hours after NYSE opens or before it closes for the equity data.
3. Placed fewer knots around the hours in which trading intensity does not change much.
   - These hours typically correspond to lunch hours (in the NY local time) for the equity data.

Specifying knots

For the equity data, the following location of knots (in hours) along the intra-day time axis might work well:
`9.30, 11, 12.30, 14.30, 16`

The shape of the spline...

- up to 12.30pm captures the busy trading hours in the morning,
- between 12.30pm and 2.30pm captures the quiet lunch hours,
- after 2.30pm captures an acceleration in trading activities before the market closes.

Check if there is any improvement in the quality of the fit of the model to the data when you fiddle with the knots. (AIC and SIC are useful. Also check the residual autocorrelation etc.)

Specifying knots

The choice depends on the empirical cases. For FX, for instance, the intensity of trading activity

- changes fast during the hours trading in major markets around the world peaks.
- doesn't change much during the evening hours (in GMT) when both New York and Asia are closed.

It is useful to sketch how a piecewise function of cubic polynomials would fit the empirical shape of intra-day patterns when determining the location and the number of knots.

Applying DCS: non-periodic component

We can decompose any non-periodic dynamics in the data into several distinct components:

$$\text{non-periodic}_{t,\tau} = \mu_{t,\tau} + \eta_{t,\tau}.$$

- $\mu_{t,\tau}$: low-frequency component. Captures highly persistent nonstationary dynamics

$$\mu_{t,\tau} = \mu_{t,\tau-1} + \kappa_\mu u_{t,\tau-1}$$

- $\eta_{t,\tau}$: stationary component. A simple possibility is AR(P).

$$\eta_{t,\tau} = \phi_1 \eta_{t,\tau-1} + \phi_2 \eta_{t,\tau-2} \cdots + \phi_P \eta_{t,\tau-P} + \kappa_\eta u_{t,\tau-1}.$$

Adding AR processes

More generally, $\eta_{t,\tau}$ can be a mixture of AR components.

$$\eta_{t,\tau} = \sum_{j=1}^{J} \eta_{t,\tau}^{(j)}$$
$$\eta_{t,\tau}^{(j)} = \phi_1^{(j)} \eta_{t,\tau-1}^{(j)} + \phi_2^{(j)} \eta_{t,\tau-2}^{(j)} \cdots + \phi_{m^{(j)}}^{(j)} \eta_{t,\tau-m^{(j)}}^{(j)} + \kappa_\eta^{(j)} u_{t,\tau-1}$$

for some $J \in \mathbb{N}_{>0}$ and $m^{(j)} \in \mathbb{N}_{>0}$. By adding several AR dynamics, we can capture stationary dynamics that is highly persistent.

Adding AR processes

For instance, consider:

$$x_t = y_t + z_t, \quad y_t = \phi_1 y_{t-1} + \varepsilon_t^{(1)}, \quad z_t = \phi_2 z_{t-1} + \varepsilon_t^{(2)}$$
$$\varepsilon_t^{(1)} \sim \text{i.i.d.} N(0, \sigma_1^2), \quad \varepsilon_t^{(2)} \sim \text{i.i.d.} N(0, \sigma_2^2).$$

Then since y_t and z_t are independent, the autocorrelation of x_t is the sum of the autocorrelation of y_t and z_t.

Even when individual series have independent "short" memory, the added process x_t can capture dynamics similar to *long memory* while being a stationary process.

There is a literature on *long memory* and ARFIMA processes. [See, for instance, Baillie (1996), Harvey (1993), Taylor (2005), Bollerslev and Mikkelsen (1996), Granger and Joyeux (1980), Hosking (1981), Engle and Lee (1999), Harvey (2013).]

The stationary component $\eta_{t,\tau}$

$\eta_{t,\tau}$: the following structure works well in many empirical applications.

$$\eta_{t,\tau} = \eta_{t,\tau}^{(1)} + \eta_{t,\tau}^{(2)}$$
$$\eta_{t,\tau}^{(1)} = \phi_1^{(1)}\eta_{t,\tau-1}^{(1)} + \phi_2^{(1)}\eta_{t,\tau-2}^{(1)} + \kappa_\eta^{(1)} u_{t,\tau-1}$$
$$\eta_{t,\tau}^{(2)} = \phi_1^{(2)}\eta_{t,\tau-1}^{(2)} + \kappa_\eta^{(2)} u_{t,\tau-1}$$

Choose the model specification using AIC and SIC.

The stationary component $\eta_{t,\tau}$

In the above specification of $\eta_{t,\tau}$:

- $\eta_{t,\tau}^{(1)}$ is stationary if $-\phi_1^{(1)} + \phi_2^{(1)} < 1$, $\phi_2^{(1)} > -1$, and $0 < \phi_1^{(1)} + \phi_2^{(1)} < 1$
- $\eta_{t,\tau}^{(2)}$ is stationary if $0 < \phi_1^{(2)} < 1$.

See, for instance, Harvey (1993, p.19).

If there are more number of lags, you can check stationarity numerically.

Spline-DCS

Spline-DCS:

$$y_{t,\tau} = z_{t,\tau} \exp(\lambda_{t,\tau}), \quad z_{t,\tau} \sim \text{i.i.d. } F(\cdot)$$
$$\lambda_{t,\tau} = \delta + \mu_{t,\tau} + \eta_{t,\tau} + \varphi_{t,\tau}$$

- $\mu_{t,\tau}$: low-frequency component. Captures highly persistent nonstationary dynamics

$$\mu_{t,\tau} = \mu_{t,\tau-1} + \kappa_\mu u_{t,\tau-1}$$

- $\eta_{t,\tau}$: stationary (autoregressive) component. A mixture of AR components captures behavior similar to long-memory.

$$\eta_{t,\tau} = \sum_{j=1}^{J} \eta_{t,\tau}^{(j)}$$
$$\eta_{t,\tau}^{(j)} = \phi_1^{(j)} \eta_{t,\tau-1}^{(j)} + \phi_2^{(j)} \eta_{t,\tau-2}^{(j)} \cdots + \phi_m^{(j)} \eta_{t,\tau-m^{(j)}}^{(j)} + \kappa_\eta^{(j)} u_{t,\tau-1}$$

for some $J \in \mathbb{N}_{>0}$ and $m^{(j)} \in \mathbb{N}_{>0}$.

- $\varphi_{t,\tau}$: periodic component capturing diurnal patterns

Parameter restrictions

The role of each component is such that:

- $\mu_{t,\tau}$ should be less sensitive to changes in $u_{t,\tau-1}$ than $\eta_{t,\tau}^{(1)}$,
- $\eta_{t,\tau}^{(1)}$ should be, in turn, less sensitive than $\eta_{t,\tau}^{(2)}$.

Then we should have

- $|\kappa_\mu| < |\kappa_\eta^{(1)}| < |\kappa_\eta^{(2)}|$.

Parameter restrictions

Set $\eta^{(1)}_{1,1} = \eta^{(2)}_{1,1} = 0$ as we have $\mathbb{E}[\eta^{(1)}_{t,\tau}] = \mathbb{E}[\eta^{(2)}_{t,\tau}] = 0$.[1]

Since $\mathbb{E}[\mu_{t,\tau}] = \mu_{1,1}$, we assume $\mu_{1,1} = 0$ so that ω is identified.

The identification conditions of the parameters in $s_{t,\tau}$ are as laid out in the separate appendix document.

[1]With the asymmetry terms, this assumes that $r_{t,\tau}$ and $y_{t,\tau}$ are independent for any $(t,\tau) \in \Psi_{T,I}$, and $\mathbb{E}[r_{t,\tau}] = 0$.

Asymmetric effect in equity volume

Recall the leverage effect in return volatility (i.e. volatility increases when the price is falling).

For equity volume, we can test for asymmetric effects in volume relating to the direction of price change by including additional asymmetry terms:

$$\eta^{(j)}_{t,\tau} = \phi^{(j)}_1 \eta^{(j)}_{t,\tau-1} + \phi^{(j)}_2 \eta^{(j)}_{t,\tau-2} \cdots + \phi^{(j)}_m \eta^{(j)}_{t,\tau-m^{(j)}} + \kappa^{(j)}_\eta u_{t,\tau-1} + \kappa^{(j)}_{\eta,a} \text{sign}(-r_{t,\tau-1})(u_{t,\tau-1} + \nu\xi).$$

$\kappa^{(j)}_{\eta,a} > 0$ (or $\kappa^{(j)}_{\eta,a} < 0$) gives an increase (decrease) in the scale of volume when price falls.

($r_{t,\tau}$ is the log-difference in price.)

Asymmetric effect in equity volume

Test the significance of the coefficients, $\kappa_{\eta,\mathrm{a}}^{(j)}$.

$u_{t,\tau-1}$ is shifted up by the factor $\nu\xi$ so that the sign of $u_{t,\tau-1}$ does not confound the sign of the asymmetry term.

The sign function captures the asymmetric effect of price change in both the positive and negative directions. (i.e. $\kappa_{\eta,\mathrm{a}}^{(j)} > 0$ (or $\kappa_{\eta,\mathrm{a}}^{(j)} < 0$) gives a decrease (increase) in the scale of volume when price increases).

How to model leverage effects in the DCS models is discussed in Harvey (2013).

Asymmetric effect in equity volume

NOTE: although the presence of asymmetric effect in equity volume is tested in a number of studies, it is not clear what the sign of $\kappa_{\eta,\mathrm{a}}^{(j)} > 0$ should be in practice.

Should trade volume go up or down when price falls?

However, at least in equity, "a fall in price is a fall in price", it is straightforward to make sense of testing for this effect.

Asymmetric effect in FX volume

For the FX data, the ways in which we should assess the existence and the directional impact of such an effect is less straight forward than the equity data.

For instance, consider the euro-dollar currency exchange rate (EURUSD).

(Aside: in FX, volume is as measured by the traded units in the left hand currency, which is priced in the right hand currency.)

Asymmetric effect in FX volume

A depreciation of EUR against USD = An appreciation of USD against EUR.

- It costs less dollar to buy euro.
- It costs more euro to buy dollar.

So the "price" has increased for a seller of euro, the "price" has fallen for a buyer of euro.

$\Rightarrow$ the interpretation of a "fall in price" depends on the side of the trade.

Set $\kappa_{\eta,\mathrm{a}}^{(i)} = 0$ for $i = 1, 2$ for the FX data for simplicity.

Announcement effect in equity

We can add an event component, $e_{t,\tau}$, in $\lambda_{t,\tau}$ to capture the effect of *anticipated* macroeconomic events.

$$\begin{aligned} e_{t,\tau} &= \phi_e e_{t,\tau-1} + \boldsymbol{\kappa}_e^\top \boldsymbol{d}_{t,\tau} \\ \boldsymbol{d}_{t,\tau} &= (d_{t,\tau,1}, \ldots, d_{t,\tau,m})^\top \\ d_{t,\tau,i} &= \mathbb{1}_{\{\text{type } i \text{ event at time } (t,\tau)\}}, \quad i = 1, \ldots, m. \end{aligned}$$

Its dynamics are assumed to be deterministic.

Set $e_{1,1} = 0$. Then $e_{t,\tau}$ reverts to zero if $|\phi_e| < 1$.

Announcement effect in equity

For equity, events that affect trade volume are typically company specific.

They include:

- the company's quarterly or annual earnings announcement
- the company's dividend announcements,
- the competitors' earnings and dividend announcements (e.g. for IBM, they are Accenture, Hewlett-Packard, and Microsoft etc),
- important news in the industry (e.g. the technology industry for IBM).

Check relevant company's websites to get the timing of these announcements.

It can be difficult to get the exact timing of news releases in intra-day hours for some companies.

Announcement effect in FX

Major macroeconomic events that affect FX include:

- Monetary policy announcements
- Macroeconomic data releases
  - US non-farm payrolls,
  - Other labor market related data,
  - Gross domestic product (GDP),
  - Consumer prices,
  - Retail sales data,
  - Manufacturing data,
  - Home sales data,
  - Various indicators of house prices.

The release of US non-farm payroll data on the first Friday of each month is (one of) the most important event for currency pairs relating to USD.

Announcement effect in FX

There are usually several events per day that can impact a given currency pair.

E.g.) See the information provided in the Forex Economic Calendar by DailyFX (www.dailyfx.com).

Derivatives

If the standard asymptotic properties (e.g. consistency and asymptotic normality) of the MLE are known, we can test parameter significance by computing the analytical standard errors of the MLE.

This requires computing several recursive equations for the first derivatives of the log-likelihood with respect to the constant parameters of the model.

Derivatives

Example:

Let $\boldsymbol{\vartheta}$ denote the vector of all of the constant parameters of Spline-DCS.

The single log-likelihood is denoted by $\log f_y(y_{t,\tau};\boldsymbol{\vartheta})$.

Denote the i-th element of $\boldsymbol{\vartheta}$ by $\boldsymbol{\vartheta}_i$.

We compute the standard error of MLE for using the outer-product of the first-derivative of the joint log-likelihood as:

$$\mathrm{S.E.}(\widehat{\boldsymbol{\vartheta}}_{\mathrm{i}}) = \sqrt{\left(\sum_{(t,\tau)\in\Psi_{T,I}} \frac{\partial \log f_Y(y_{t,\tau};\widehat{\boldsymbol{\vartheta}})}{\partial \boldsymbol{\vartheta}} \frac{\partial \log f_Y(y_{t,\tau};\widehat{\boldsymbol{\vartheta}})}{\partial \boldsymbol{\vartheta}^{\top}}\right)^{-1}_{ii}},$$

where $\cdot_{ii}$ for $i = 1, \ldots, \dim(\boldsymbol{\vartheta})$ denotes the i-th diagonal element.

Derivatives

As mentioned above, computing this quantity requires computing several recursive equations.

To see this, note that if the error distribution, $F^*(\,\cdot\,;\theta^*)$, is GB2, $\boldsymbol{\vartheta}$ includes the distribution parameters $\theta^* = (\nu, \xi, \zeta)^\top$, as well as the Bernoulli parameter $p \in (0,1)$ and the parameters of $\lambda_{t,\tau}$.

Then we have the following derivatives with respect to the distribution parameters:

$$\frac{\partial \log f_Y(y_{t,\tau};\boldsymbol{\vartheta})}{\partial \nu} = \frac{1}{\nu} + \xi \log(y_{t,\tau} e^{-\lambda_{t,\tau}}) + u_{t,\tau}\frac{\partial \lambda_{t,\tau}}{\partial \nu},$$
$$\frac{\partial \log f_Y(y_{t,\tau};\boldsymbol{\vartheta})}{\partial \xi} = \log(b_{t,\tau}) - \frac{1}{B(\xi,\zeta)}\frac{\partial B(\xi,\zeta)}{\partial \xi} + u_{t,\tau}\frac{\partial \lambda_{t,\tau}}{\partial \xi},$$
$$\frac{\partial \log f_Y(y_{t,\tau};\boldsymbol{\vartheta})}{\partial \zeta} = -\frac{1}{B(\xi,\zeta)}\frac{\partial B(\xi,\zeta)}{\partial \zeta} + \log(1 - b_{t,\tau}) + u_{t,\tau}\frac{\partial \lambda_{t,\tau}}{\partial \zeta},$$
$$\frac{\partial \log f_Y(y_{t,\tau};\boldsymbol{\vartheta})}{\partial p} = \mathbb{1}_{\{y_{t,\tau}=0\}}/p - \mathbb{1}_{\{y_{t,\tau}>0\}}/(1-p).$$

Derivatives

Note that

- $\partial B(\xi,\zeta)/\partial\xi = B(\xi,\zeta)(\psi(\xi) - \psi(\xi+\zeta))$ with $\partial B(\xi,\zeta)/\partial\zeta = \partial B(\zeta,\xi)/\partial\zeta$ by the symmetry of the Beta function.
- $\psi(\cdot)$ is the digamma function.

With respect to other parameters of $\boldsymbol{\vartheta}$ (denoted by $\boldsymbol{\vartheta}_{(-)}$), we have

$$\frac{\partial \log f_Y(y_{t,\tau};\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}_{-\theta,i}} = u_{t,\tau}\frac{\partial \lambda_{t,\tau}}{\partial \boldsymbol{\vartheta}_{-\theta,i}},$$

where $\boldsymbol{\vartheta}_{(-),i}$ is the i-th element of $\boldsymbol{\vartheta}_{(-)}$.

Derivatives

Denoting the i-th element of $\boldsymbol{\vartheta}$ by $\boldsymbol{\vartheta}_i$, the derivatives of $\lambda_{t,\tau}$ are given by the following recursions:

$$\frac{\partial \lambda_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} = \mathbb{1}_{\{\boldsymbol{\vartheta}_i=\omega\}} + \frac{\partial \mu_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} + \frac{\partial \eta_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} + \frac{\partial s_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} + \frac{\partial e_{t,\tau}}{\partial \boldsymbol{\vartheta}_i}$$

$$\frac{\partial \mu_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} = \frac{\partial \mu_{t,\tau-1}}{\partial \boldsymbol{\vartheta}_i} + u_{t,\tau}\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\kappa_\mu\}} + \kappa_\mu \frac{\partial u_{t,\tau-1}}{\partial \boldsymbol{\vartheta}_i}$$

$$\frac{\partial \eta_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} = \frac{\partial \eta^{(1)}_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} + \frac{\partial \eta^{(2)}_{t,\tau}}{\partial \boldsymbol{\vartheta}_i},$$

Derivatives

$$\begin{aligned}\frac{\partial \eta^{(j)}_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} &= \eta^{(j)}_{t,\tau-1}\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\phi^{(j)}_1\}} + \eta^{(j)}_{t,\tau-2}\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\phi^{(1)}_2\}} + u_{t,\tau-1}\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\kappa^{(j)}_\eta\}} \\ &+ \phi^{(j)}_1 \frac{\partial \eta^{(j)}_{t,\tau-1}}{\partial \boldsymbol{\vartheta}_i} + \phi^{(j)}_2 \frac{\partial \eta^{(j)}_{t,\tau-2}}{\partial \boldsymbol{\vartheta}_i} + \kappa^{(j)}_\eta \frac{\partial u_{t,\tau-1}}{\partial \boldsymbol{\vartheta}_i} \\ &+ \operatorname{sign}(-r_{t,\tau})(u_{t,\tau-1} + \nu\xi)\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\kappa^{(j)}_{\eta,a}\}} \\ &+ \kappa^{(j)}_{\eta,a}\operatorname{sign}(-r_{t,\tau})\left(\frac{\partial u_{t,\tau-1}}{\partial \boldsymbol{\vartheta}_i} + \xi\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\nu\}} + \nu\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\xi\}}\right), \quad j = 1, 2,\end{aligned}$$

$$\frac{\partial e_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} = e_{t,\tau-1}\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\phi_e\}} + \phi_e \frac{\partial e_{t,\tau-1}}{\partial \boldsymbol{\vartheta}_i} + d_{t,\tau,m}\mathbb{1}_{\{\boldsymbol{\vartheta}_i=\kappa_{e,m}\}},$$

where $m = 1, \ldots, \dim(\boldsymbol{d}_{t,\tau})$ and $\kappa_{e,m}$ is m-th element of $\boldsymbol{\kappa}_e$.

Derivatives

As for the spline component, we have

$$\frac{\partial s_\tau}{\partial \boldsymbol{\vartheta}_i} = \sum_{j=1}^{k} \mathbb{1}_{\{\tau \in [\tau_{j-1}, \tau_j]\}} \mathbf{z}_{j,l}(\tau) \mathbb{1}_{\{\boldsymbol{\vartheta}_i = \gamma_l\}},$$

if it is a static daily spline, where $l = 0, \ldots, k-1$ and k is the number of knots.

If we assume that everything is constant at $\tau = 0$ on day $t = 1$, these recursions begin from zero.

Finally, for the score variable, we have

$$\begin{aligned}\frac{\partial u_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} &= ((\xi + \zeta) b_{t,\tau} - \xi) \mathbb{1}_{\{\boldsymbol{\vartheta}_i = \nu\}} + \nu (b_{t,\tau} - 1) \mathbb{1}_{\{\boldsymbol{\vartheta}_i = \xi\}} \\ &\quad + \nu b_{t,\tau} \mathbb{1}_{\{\boldsymbol{\vartheta}_i = \zeta\}} + \nu(\xi + \zeta) \frac{\partial b_{t,\tau}}{\partial \boldsymbol{\vartheta}_i}\end{aligned}$$

$$\frac{\partial b_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} = \begin{cases} -b_{t,\tau}(1 - b_{t,\tau}) \log(y_{t,\tau} e^{-\lambda_{t,\tau}})(y_{t,\tau} e^{-\lambda_{t,\tau}}) \frac{\partial \lambda_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} & \text{if } \boldsymbol{\vartheta}_i = \nu, \\ -\nu b_{t,\tau}(1 - b_{t,\tau}) \frac{\partial \lambda_{t,\tau}}{\partial \boldsymbol{\vartheta}_i} & \text{otherwise.} \end{cases}$$

Derivatives

Having said this....

Remark: we can notice from the above that programming all of these derivatives and recursive equations is not very easy.

- The specification of the recursive equations have to change every time the model specification changes.

Much easier to test model specification using AIC and SIC.
Then we only need to compute the optimized value of the log-likelihood, which is very easy.
This highlights the usefulness of these statistics.