News Impact on Stock Market

Directeur: Charles-Albert LEHALLE, Romain BURGOT

15 Septembre 2010
Contents

1 The Problem 3

2 The Model 6
   2.1 Motivation 6
   2.2 Latent Kalman Filter 7
   2.3 Parameter Estimation - EM 8
      2.3.1 Updating, Filtering, Smoothing 8
      2.3.2 EM 10
   2.4 Application on News Impact Exploration 14
      2.4.1 US-Employment Situation 16
      2.4.2 Earning Release 17
      2.4.3 Witching day 18

3 Extension: Non-homogeneous Kalman Filter 19
   3.1 Whole-day Impact 19
   3.2 ”Peak, Choc” Impact 23

4 Conclusion 25
1 The Problem

The problem we are facing is to understand the impact of financial news on stock market. From historical market data, observations of certain types of news’s impact seemed to be "periodic", which gave us the chance to predict. For example, on Witching days, the average daily volume of the stocks in CAC40 seems to be much larger than usual.

One difficulty in analysing news impact comes from the "coincidence" of different news on the same day, or even at same time. On Witching day I (WD-I) below, there are many other financial news which might also have large impact on the intra-day volume.
This requires us to attack the problem "inversely". We will focus on finding out similar patterns, when conditioned on certain type of news (e.g WD-1). From (below) we see another WD-1 on 2007, where a large peak around 12:00 (pattern) seems quite similar to that on 2009 (above).
Different types of news may have very different impact patterns. For news on Earning Release, the impact seems to be less localised than that of WD-1.
Perhaps we can "conclude" from the above figures that on Earning Release, the volume in the morning is much larger than usual, even in the afternoon if the stock "belongs to US". To get a more precise description of the above phenomenon, we are to modelise the problem.

## 2 The Model

We model the impact of news by studying **intra-day volume** based on its auto-correlation property.

### 2.1 Motivation

From the descriptive statistics on the intra-day volume below, we may observe some linear daily dynamics and inter-day seasonality. Let the observed intra-day volume be $V_{d,t}$, and define daily volume component be $x_d = \frac{1}{T} \sum_1^T \log V_{d,t}$, then $x_d$ and $x_{d+1}$ could be linearly related.
The intra-day component can be defined as \( r_{d,t} = \log V_{d,t} - x_d \), which is "centred" in sense of log additivity. The inter-day seasonality is characterised by \( \frac{1}{D} \sum_{d=1}^{D} r_{d,t} \), as shown below:

However, when there is news, the observed volume \( V_{d,t} \) tends to be largely impacted. From above, the effect of Witching day around 12:00 is quite significant among all the intra-day curves. We therefore regard them as latent variable to avoid using "outlier data" in our model. Whenever \( V_{d,t} \) is latent, the daily component \( x_d \) will become non-observable. A consistent way is to regard \( x_d \) as a latent variable all the time, since in reality we have no prior knowledge on which day there is "no news", i.e. when \( x_d \) be observable.

### 2.2 Latent Kalman Filter

The dynamics of daily and intra-day component of the volume can be regarded as a latent Kalman filter. The dynamics system comes from separating the daily and intra-day component of "usual" volume by \( Y_{d,t} \overset{\Delta}{=} V_{d,t} \) when there is no news impact,

\[
Y_{d,t} = W_d \cdot H_{d,t}. \tag{1}
\]

We may suppose the daily component \( \log W_d \in \mathbb{R} \) be a process of AR(1), and the intra-day component \( \log H_{d,} \in \mathbb{R}^T \) be an i.i.d multi-variate Gaussian process. Under some conventions, the system becomes a (linear Gaussian) Kalman Filter,

\[
\log Y_d = \mathbb{1} \cdot \log W_d + \log H_d \in \mathbb{R}^T. \tag{2}
\]

Let \( x_d = \log W_d, y_d = \log Y_d \), then for \( d = 0, \ldots, D - 1 \),

\[
x_{d+1} | x_d \sim \mathcal{N}(A x_d + B, C), \tag{3}
\]

\[
y_{d+1} | x_{d+1} \sim \mathcal{N}(\mathbb{1} x_{d+1} + E, F). \tag{4}
\]

To make the model identifiable, we may impose some conditions on \( E \), say

\[
\mathbb{1} \perp E, \text{ i.e. } \sum_{t=1}^{T} E_t = 0; \tag{5}
\]

As a starting point, we will initialise the system by \( x_0 \sim \mathcal{N}(\mu_0, \Sigma_0) \). Since only part of \( y_d \in \mathbb{R}^T \) is observable, let’s note the index of this part by \( \Lambda_d \subseteq \{1, \ldots, T\} \), and let \( y_{d}^{\Lambda} \) be this observable part.
The parameter space of the model will be (with constraint on $E$)
\[ \theta = (A, B, C, E, F, \mu_0, \Sigma_0). \]  

(6)

And the goal of parameter estimation is to maximize the likelihood,
\[ \hat{\theta} = \arg \max_{\theta} \log p(y_1^\Lambda, \ldots, y_D^\Lambda | \theta). \]

(7)

### 2.3 Parameter Estimation - EM

This part is to detail the parameter estimation for Latent Kalman Filter model, we will find in the end that the whole estimation by EM is **analytic** and **recursive**. The fundamental lemma we will use, either explicitly or implicitly, is

**Lemma 2.1.** Let $v \in \mathbb{R}^m$, $u \in \mathbb{R}^n$,
\[ v \sim \mathcal{N}(V, K), \quad u|v \sim \mathcal{N}(Lv + M, \Sigma), \]

(8)

\[ \Rightarrow \begin{pmatrix} v \\ u \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} V \\ LV + M \end{pmatrix}, \begin{bmatrix} K & (LK)^\top \\ LK & \Sigma + LKL^\top \end{bmatrix} \right). \]

(9)

Moreover,
\[ v|u \sim \mathcal{N}(V + K^\top L^\top (\Sigma + LKL^\top)^{-1}[u - (LV + M)], \]

(10)

\[ K - K^\top L^\top (\Sigma + LKL^\top)^{-1}L). \]

(11)

### 2.3.1 Updating, Filtering, Smoothing

Let’s start with the initial state $x_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$, and suppose $x_d \mid y_1^\Lambda, \ldots, y_d^\Lambda \sim \mathcal{N}(\mu_{d|d}, \Sigma_{d|d})$, by convention $\mu_{0|0} = \mu_0, \Sigma_{0|0} = \Sigma_0$, then
\[ x_{d+1} \mid y_1^\Lambda, \ldots, y_d^\Lambda \sim \mathcal{N}(\mu_{d+1|d|d}, \Sigma_{d+1|d|d}), \]

(12)

with $\mu_{d+1|d} = A\mu_{d|d} + B; \quad \Sigma_{d+1|d} = C + A\Sigma_{d|d}A^\top$;

(13)

because
\[ \begin{pmatrix} x_d \\ x_{d+1} \end{pmatrix} \mid y_1^\Lambda, \ldots, y_d^\Lambda \sim \mathcal{N}\left( \begin{pmatrix} \mu_{d|d} \\ A\mu_{d|d} + B \end{pmatrix}, \begin{bmatrix} \Sigma_{d|d} & (A\Sigma_{d|d})^\top \\ A\Sigma_{d|d}^\top & C + A\Sigma_{d|d}A^\top \end{bmatrix} \right); \]

(14)

One step ahead,
\[ x_{d+1} \mid y_1^\Lambda, \ldots, y_d^\Lambda \sim \mathcal{N}(\mu_{d+1|d+1}, \Sigma_{d+1|d+1}), \]

with $\mu_{d+1|d+1} = \mu_{d+1|d} + \Sigma_{d+1|d}(\mathbb{1}^\Lambda_{d+1})^\top (F_{d+1}^{A\Lambda} + \mathbb{1}^\Lambda_{d+1}\Sigma_{d+1|d}(\mathbb{1}^\Lambda_{d+1})^\top)^{-1}[y_{d+1}^\Lambda - (\mathbb{1}^\Lambda_{d+1}\mu_{d+1|d} + E_{d+1}^\Lambda)]$, 
\[ \Sigma_{d+1|d+1} = \Sigma_{d+1|d} - \Sigma_{d+1|d}(\mathbb{1}^\Lambda_{d+1})^\top (F_{d+1}^{A\Lambda} + \mathbb{1}^\Lambda_{d+1}\Sigma_{d+1|d}(\mathbb{1}^\Lambda_{d+1})^\top)^{-1}\mathbb{1}^\Lambda_{d+1}\Sigma_{d+1|d}, \]

because
\[ \begin{pmatrix} x_{d+1} \\ y_{d+1}^\Lambda \end{pmatrix} \mid y_1^\Lambda, \ldots, y_d^\Lambda \sim \mathcal{N}\left( \begin{pmatrix} \mu_{d+1|d} \\ \mathbb{1}^\Lambda_{d+1}\mu_{d+1|d} + E_{d+1}^\Lambda \end{pmatrix}, \begin{bmatrix} \Sigma_{d+1|d} & (\mathbb{1}^\Lambda_{d+1}\Sigma_{d+1|d})^\top \\ \mathbb{1}^\Lambda_{d+1}\Sigma_{d+1|d}^\top & F_{d+1}^{A\Lambda} + \mathbb{1}^\Lambda_{d+1}\Sigma_{d+1|d}(\mathbb{1}^\Lambda_{d+1})^\top \end{bmatrix} \right); \]
Write the above formulas by Kalman matrix $K_{d+1} = \Sigma_{d+1|d}^T (\mathbb{I}_{d+1} \Lambda + \Sigma_{d+1|d} (\mathbb{I}_{d+1} \Lambda)^T)^{-1}$, then for $d = 0, \ldots, D - 1$, the above filtering and updating process are summarised by

\begin{align}
\mu_{d+1|d} &= \Lambda \mu_{d|d} + B, \\
\Sigma_{d+1|d} &= C + A \Sigma_{d|d} A^T, \\
\mu_{d+1|d+1} &= \mu_{d+1|d} + K_{d+1} [y_{d+1} - (\mathbb{I}_{d+1} \mu_{d+1|d} + E_{d+1})], \\
\Sigma_{d+1|d+1} &= \Sigma_{d+1|d} - K_{d+1} \mathbb{I}_{d+1} \Sigma_{d+1|d}. 
\end{align}

The smoothing process is to calculate the distribution of $x_d \mid y_1, \ldots, y_D$, which will eventually be used in EM estimation. Start with $x_D \mid y_1^D, y_D^\Lambda \sim \mathcal{N}(\mu_{d|d}, \Sigma_{d|d})$, and suppose

\begin{equation}
x_{d+1} \mid y_{d+1}^A, \ldots, y_D^\Lambda \sim \mathcal{N}(\mu_{d+1|d+1}, \Sigma_{d+1|d+1}),
\end{equation}

then for each $d = D - 1, \ldots, 0$,

\begin{align}
x_d \mid x_{d+1}, y_{d+1}^A, \ldots, y_D^\Lambda &\sim x_d \mid x_{d+1}, y_{d+1}^A, \ldots, y_D^\Lambda \sim \mathcal{N}(\mu_{d|d+1|d}, \Sigma_{d|d+1|d}) \\
\text{with } \mu_{d|d+1|d} &= \mu_{d|d} + J_d [x_{d+1} - (A \mu_{d|d} + B)] = \mu_{d|d} + J_d (x_{d+1} - \mu_{d+1|d}), \\
\Sigma_{d|d+1|d} &= \Sigma_{d|d} - J_d A \Sigma_{d|d} = \Sigma_{d|d} - J_d \Sigma_{d+1|d} J_d^T,
\end{align}

Thus

\begin{equation}
\begin{bmatrix} x_{d+1} \\ x_d \end{bmatrix} \mid y_1^A, \ldots, y_D^\Lambda \sim \mathcal{N}\left(\begin{bmatrix} \mu_{d+1|d} \\ \mu_{d|d} \end{bmatrix}, \begin{bmatrix} \Sigma_{d+1|d} & (J_d \Sigma_{d+1|d})^T \\ J_d \Sigma_{d+1|d} & \Sigma_{d|d} \end{bmatrix}\right),
\end{equation}

\begin{align}
\text{with } \mu_{d|d} &= \mu_{d|d} + J_d [\mu_{d+1|d} - \mu_{d+1|d}], \\
\Sigma_{d|d} &= \Sigma_{d|d} + J_d (\Sigma_{d+1|d} - \Sigma_{d+1|d}) J_d^T.
\end{align}

Having obtained the posterior distribution of each $x_d$, we may further deduce that of the non-observable part $y_d^X$, i.e. $y_d^X \mid y_1^A, \ldots, y_D^\Lambda$. Since

\begin{equation}
y_d^X \mid x_d, y_1^A, \ldots, y_D^\Lambda \sim y_d^X \mid x_d, y_d^\Lambda,
\end{equation}

as well as $y_d \mid x_d \sim \mathcal{N}(\mathbb{1} x_d + E, F)$, we have

\begin{equation}
y_d^X \mid x_d, y_d^A \sim \mathcal{N}\left(\begin{bmatrix} x_d + E_d^X & F_d^X \Lambda \end{bmatrix}^{-1} (y_d^X - \mathbb{1} x_d - E_d^X), \right.
\end{equation}

\begin{equation}
F_d^X = F_d^X - F_d^X (F_d^X)^{-1} F_d^X.
\end{equation}

Set $L_d^X = \mathbb{1}_d^X - F_d^X (F_d^X)^{-1} \mathbb{1}_d^X$, we deduce by combining $x_d \mid y_1^A, \ldots, y_D^\Lambda \sim \mathcal{N}(\mu_{d|d}, \Sigma_{d|d})$,

\begin{equation}
y_d^X \mid y_1^A, \ldots, y_D^\Lambda \sim \mathcal{N}(E_{d|d}^X, F_{d|d}^X),
\end{equation}

\begin{align}
\text{with } E_{d|d}^X &= \mu_{d|d} + E_d^X + F_{d|d}^X (F_{d|d}^X)^{-1} (y_d^X - \mathbb{1}_d \mu_{d|d} - E_d^X), \\
F_{d|d}^X &= F_d^X - F_d^X (F_d^X)^{-1} F_d^X + L_d^X \Sigma_{d|d} (L_d^X)^T.
\end{align}
2.3.2 EM

Now we are prepared to do EM. Let \( x = (x_0, \ldots, x_D), \ y = (y_1, \ldots, y_D), \ y^A = (y^A_1, \ldots, y^A_D), \) then at each iteration the estimator \( \hat{\theta} \) will be updated by

\[
\arg\max_{\theta} \mathbb{E}[\log p(x, y|\theta) \mid y^A, \hat{\theta}].
\] (34)

To achieve this, let’s first ”decompose” the complete log likelihood by

\[
\log p(x, y|\theta) = \log p(x_0|\theta) + \sum_{d=0}^{D-1} \log p(x_{d+1}|x_d, \theta) + \sum_{d=1}^{D} \log p(y_d|x_d, \theta).
\] (35)

and then calculate each component’s expectation by ”centring” the non-observable variable \( \bar{x}_d = x_d - \hat{\mu}_{d|D}, \ \bar{y}^A_d = y^A_d - \hat{E}^A_{d|D} :\)

\[
\mathbb{E} [\log p(x_0|\theta) \mid y^A, \hat{\theta}] = \int \log p(x_0|\theta)p(x_0|y^A, \hat{\theta})dx_0
\] (36)

\[
= -\frac{1}{2} \mathbb{E} [(x_0 - \mu_0)^\top \Sigma_0^{-1}(x_0 - \mu_0) \mid y^A, \hat{\theta}] - \frac{1}{2} \log |\Sigma_0| - \frac{1}{2} \log(2\pi)
\] (37)

\[
\propto -\frac{1}{2} \mathbb{E} [\bar{x}_0 + \hat{\mu}_{0|D} - \mu_0)^\top \Sigma_0^{-1}(\bar{x}_0 + \hat{\mu}_{0|D} - \mu_0) \mid y^A, \hat{\theta}] - \frac{1}{2} \log |\Sigma_0|
\] (38)

\[
\propto -\frac{1}{2} \text{tr}(\Sigma_0^{-1}\hat{\Sigma}_{0|D}) - \frac{1}{2} (\hat{\mu}_{0|D} - \mu_0)^\top \Sigma_0^{-1}(\hat{\mu}_{0|D} - \mu_0) - \frac{1}{2} \log |\Sigma_0|.
\] (39)

\[
\mathbb{E} [\log p(x_{d+1}|x_d, \theta) \mid y^A, \hat{\theta}] = \int \log p(x_{d+1}|x_d, \theta)p(x_{d+1}, x_d|y^A, \hat{\theta})dx_{d+1}dx_d
\] (40)

\[
= -\frac{1}{2} \mathbb{E} [(x_{d+1} - Ax_d - B)^\top C^{-1}(x_{d+1} - Ax_d - B) \mid y, \hat{\theta}] - \frac{1}{2} \log |C| - \frac{1}{2} \log(2\pi)
\] (41)

\[
\propto -\frac{1}{2} \log |C|
\] (42)

\[
-\frac{1}{2} \mathbb{E} [(\bar{x}_{d+1} - \hat{A}\bar{x}_d + \hat{\mu}_{d+1|D} - A\hat{\mu}_{d|D} - B )^\top C^{-1} (\bar{x}_{d+1} - \hat{A}\bar{x}_d + \hat{\mu}_{d+1|D} - A\hat{\mu}_{d|D} - B) \mid y, \hat{\theta}]
\] (43)

\[
\propto -\frac{1}{2} \log |C| - \frac{1}{2} \text{tr}(C^{-1}\hat{\Sigma}_{d+1|D}) - \frac{1}{2} \text{tr}(C^{-1}\hat{A}\hat{\Sigma}_{d+1|D}A^\top)
\] (44)

\[
+ \frac{1}{2} \text{tr}(C^{-1}\hat{A}\hat{J}_d\hat{\Sigma}_{d+1|D}) + \frac{1}{2} \text{tr}(C^{-1}(A\hat{J}_d\hat{\Sigma}_{d+1|D})^\top) - \frac{1}{2} (\hat{\mu}_{d+1|D} - A\hat{\mu}_{d|D} - B)^\top C^{-1} (\hat{\mu}_{d+1|D} - A\hat{\mu}_{d|D} - B).
\] (45)

\[
\mathbb{E} [\log p(y_d|x_d, \theta) \mid y^A, \hat{\theta}] = \int \log p(y_d|x_d, \theta)p(x_d, y^A_d|y^A, \hat{\theta})dx_d dy^A_d
\] (46)

\[
= -\frac{1}{2} \mathbb{E} [(y_d - \mathbb{1}x_d - E)^\top F^{-1}(y_d - \mathbb{1}x_d - E) \mid y^A, \hat{\theta}] - \frac{1}{2} \log |F| - \frac{T}{2} \log(2\pi)
\] (47)

\[
\propto -\frac{1}{2} \text{tr}\left(F^{-1}\left[R_{1,d} - R_{2,d}E^\top - (R_{2,d}E^\top)^\top + EE^\top\right]\right) - \frac{1}{2} \log |F|,
\] (48)

with \( R_{1,d} = \mathbb{E} [(y_d - \mathbb{1}x_d)(y_d - \mathbb{1}x_d)^\top \mid y^A, \hat{\theta}], \ R_{2,d} = \mathbb{E} [(y_d - \mathbb{1}x_d) \mid y^A, \hat{\theta}]. \) (49)
By setting $R_1 = \sum_{d=1}^{D} R_{1,d}$, $R_2 = \sum_{d=1}^{D} R_{2,d}$, the above formulas are summarised by

\[
\begin{align*}
    l(\theta, \hat{\theta}) &\triangleq \mathbb{E}\left[ \log p(x, y \mid \theta) \mid y^A, \hat{\theta} \right] \\
    &\asymp - \frac{1}{2} \log |\Sigma_0| - \frac{D}{2} \log |F| - \frac{D}{2} \log |C| \\
    &- \frac{1}{2} \text{tr}\left( \Sigma_0^{-1} \left\{ (\hat{\mu}_0|D) - (\hat{\mu}_0) \right\} + \hat{\Sigma}_0|D \right) \right) \\
    &- \frac{1}{2} \text{tr}\left( F^{-1} \left[ R_1 - R_2 E^\top - (R_2 E^\top)^\top + D E E^\top \right] \right) \\
    &- \frac{1}{2} \text{tr}\left( C^{-1} \sum_{d=0}^{D-1} \left\{ (\hat{\mu}_{d+1}|D) - A \hat{\mu}_{d|D} - B) (\hat{\mu}_{d+1}|D) - A \hat{\mu}_{d|D} - B \right\}^\top \\
    &\quad + \hat{\Sigma}_{d+1|D} + A^\top \hat{\Sigma}_{d|D} A^\top - A \hat{J}_d \hat{\Sigma}_{d+1|D} - (A \hat{J}_d \hat{\Sigma}_{d+1|D})^\top \right) \right). \\
\end{align*}
\]

Some matrix calculus gives

\[
\begin{align*}
    \frac{\partial l}{\partial \mu_0} &\asymp \Sigma_0^{-1}(\hat{\mu}_0|D) - \mu_0, \\
    \frac{\partial l}{\partial \Sigma_0} &\asymp -\frac{1}{2} \Sigma_0^{-1} + \frac{1}{2} \Sigma_0^{-1} \left( \hat{\Sigma}_0|D + (\hat{\mu}_0|D) - \mu_0 \right)^\top \Sigma_0^{-1}. \\
\end{align*}
\]

Setting the derivatives to the zero, we get the unique solution

\[
\hat{\mu}_0 = \hat{\mu}_{0|D}, \quad \hat{\Sigma}_0 = \hat{\Sigma}_{0|D}. \\
\]

Now regard $\mu_0$ and $\Sigma_0$ as constant in $l(\theta, \hat{\theta})$ and rearrange, we get

\[
\begin{align*}
l(\theta, \hat{\theta}) &\asymp - \frac{D}{2} \log |F| - \frac{D}{2} \log |C| \\
    &- \frac{1}{2} \text{tr}\left( F^{-1} \left[ R_1 - R_2 E^\top - (R_2 E^\top)^\top + D E E^\top \right] \right) \\
    &- \frac{1}{2} \text{tr}\left( C^{-1} \sum_{d=0}^{D-1} \left\{ (\hat{\mu}_{d+1}|D) \hat{\Sigma}_{d+1|D} + \hat{\Sigma}_{d+1|D} \right\} + BB^\top + A \left[ \hat{\mu}_{d|D} \hat{\mu}_{d|D}^\top + \hat{\Sigma}_{d|D} \right] A^\top \\
    &\quad - A \left[ \hat{\mu}_{d|D} \hat{\mu}_{d+1|D}^\top + \hat{J}_d \hat{\Sigma}_{d+1|D} \right] - \left[ \hat{\mu}_{d|D} \hat{\mu}_{d+1|D}^\top + \hat{J}_d \hat{\Sigma}_{d+1|D} \right] A^\top \\
    &\quad + A \hat{\mu}_{d|D} B^\top + (A \hat{\mu}_{d|D} B^\top - \hat{\mu}_{d+1|D} B^\top - (\hat{\mu}_{d+1|D} B^\top)^\top \right) \right). \\
\end{align*}
\]

Use the symmetry of $C$, we get

\[
\frac{\partial l}{\partial B} = - \left( TC^{-1} B + C^{-1} A \sum_{d=0}^{D-1} \hat{\mu}_{d|D} - C^{-1} \sum_{d=0}^{D-1} \hat{\mu}_{d+1|D} \right),
\]
\[
\frac{\partial l}{\partial A} = - \left\{ C^{-1} A \sum_{d=0}^{D-1} \left[ \hat{\mu}_{d|D}(\hat{\mu}_{d|D})^\top + \hat{\Sigma}_{d|D} \right] - C^{-1} \left( \sum_{d=0}^{D-1} \left[ \hat{\mu}_{d+1|D}(\hat{\mu}_{d+1|D})^\top + \hat{J}_d \hat{\Sigma}_{d+1|D} \right] \right)^\top + C^{-1} B \left( \sum_{d=0}^{D-1} \hat{\mu}_{d|D} \right)^\top \right\}. \tag{65}
\]

Sufficient statistics help to simplify the formula:

\[
S_1 = \sum_{d=0}^{D-1} \hat{\mu}_{d|D}, \quad S_2 = \sum_{d=0}^{D-1} \hat{\mu}_{d+1|D}, \tag{67}
\]

\[
S_3 = \sum_{d=0}^{D-1} \left[ \hat{\mu}_{d|D}(\hat{\mu}_{d|D})^\top + \hat{\Sigma}_{d|D} \right], \tag{68}
\]

\[
S_4 = \sum_{d=0}^{D-1} \left[ \hat{\mu}_{d|D}(\hat{\mu}_{d+1|D})^\top + \hat{J}_d \hat{\Sigma}_{d+1|D} \right], \tag{69}
\]

\[
S_5 = \sum_{d=0}^{D-1} \left[ \hat{\mu}_{d+1|D}(\hat{\mu}_{d+1|D})^\top + \hat{\Sigma}_{d+1|D} \right]. \tag{70}
\]

\[
\left\{ \begin{array}{l}
\hat{B} = \frac{1}{D} [S_2 - \hat{A} S_1] \\
\hat{A} S_3 = S_4^\top - \hat{B} S_1^\top
\end{array} \right. \tag{71}
\]

Thus

\[
\hat{A} = DS_4^\top - S_2 S_1^\top, \quad \hat{B} = \frac{1}{D} [S_2 - \hat{A} S_1] \tag{72}
\]

\[
\hat{C} = S_5 + D \hat{B} \hat{B}^\top + \hat{A} S_3 \hat{A}^\top \tag{73}
\]

\[- \hat{A} S_4 - (\hat{A} S_4)^\top + \hat{A} S_1 \hat{B}^\top + (\hat{A} S_1 \hat{B}^\top)^\top - S_2 \hat{B}^\top - (S_2 \hat{B}^\top)^\top. \tag{74}
\]

To estimate \(E\) and \(F\), we are to solve

\[
\max_{E,F} \quad - \frac{1}{2} \text{tr} \left( F^{-1} \left[ R_1 - R_2 E^\top - (R_2 E^\top)^\top + D E E^\top \right] \right) - \frac{D}{2} \log |F| \tag{75}
\]

subject to \(E^\top E = 0\). \tag{76}

By Lagrange multiplier method, the problem becomes

\[
\min_{E,F} L(E, F, \lambda) = \frac{1}{2} \text{tr} \left( F^{-1} \left[ R_1 - R_2 E^\top - (R_2 E^\top)^\top + D E E^\top \right] \right) + \frac{D}{2} \log |F| + \lambda E^\top E. \tag{77}
\]

The local optima thus satisfies

\[
\frac{\partial L}{\partial E} = 0, \quad \frac{\partial L}{\partial F} = 0, \quad \frac{\partial L}{\partial \lambda} = 0. \tag{78}
\]
Again the matrix calculus gives
\[
\begin{align*}
\frac{\partial L}{\partial E} & \propto DF^{-1}E - F^{-1}R_2 + \lambda \mathbb{1} \quad (79) \\
\frac{\partial L}{\partial F} & \propto F^{-1}[R_1 - R_2E^\top - (R_2E^\top)^\top + DEE^\top]F^{-1} - DF^{-1} \quad (80) \\
\frac{\partial L}{\partial \lambda} & \propto \mathbb{1}^\top E. \quad (81)
\end{align*}
\]

Surprisingly, the above solution is analytic and unique. Setting the above derivatives to zero,
\[
\begin{align*}
\begin{cases}
\mathbb{1}^\top E = 0 \\
R_2 - DE = \lambda F \mathbb{1} \\
DF = [R_1 - R_2E^\top - (R_2E^\top)^\top + DEE^\top]
\end{cases}
\end{align*} \quad (82)
\]

Then
\[
R_2 - DE = \lambda F \mathbb{1} \quad (83)
\]
\[
= \frac{\lambda}{D}[R_1 - R_2E^\top - (R_2E^\top)^\top + DEE^\top] = \frac{\lambda}{D}[R_1 \mathbb{1} - ER_2^\top \mathbb{1}] \quad (84)
\]

Multiplying both side by \(\mathbb{1}^\top\), we get
\[
\mathbb{1}^\top R_2 = \frac{\lambda}{D} \mathbb{1}^\top R_1 \mathbb{1} \implies \hat{\lambda} = \frac{D \mathbb{1}^\top R_2}{\mathbb{1}^\top R_1 \mathbb{1}}. \quad (85)
\]

\(\hat{E}, \hat{F}\) follows immediately from \(\hat{\lambda}\),
\[
\hat{E} = \frac{R_2 - \frac{\hat{\lambda}}{D} R_1 \mathbb{1}}{D - \frac{\hat{\lambda}}{D} R_2^\top \mathbb{1}}, \quad \hat{F} = \frac{1}{D}[R_1 - R_2\hat{E}^\top - (R_2\hat{E}^\top)^\top + D\hat{E}\hat{E}^\top]. \quad (86)
\]

The last "stone" is to calculate \(R_1\) et \(R_2\). We separate each \(R_{1,d}\) by \(R_{1,A}^{\Lambda,\Lambda}\), \(R_{1,d}^{\Lambda,\Lambda}\), \(R_{1,d}^{\Lambda,\Lambda}\), \(R_{1,d}^{\Lambda,\Lambda}\), as well as \(R_{2,d}\) by \(R_{2,d}^{\Lambda,\Lambda}\) and \(R_{2,d}^{\Lambda,\Lambda}\),
\[
R_{1,d}^{\Lambda,\Lambda} = \mathbb{E}[(y_d^{\Lambda} - \mathbb{1}_d x_d) | y^{\Lambda}, \hat{\theta}] = y_d^{\Lambda} - \mathbb{1}_d \bar{\mu}_{d|D}, \quad (87)
\]
\[
R_{2,d}^{\Lambda,\Lambda} = \mathbb{E}[(y_d^{\Lambda} - \mathbb{1}_d x_d) | y^{\Lambda}, \hat{\theta}] = \hat{E}_{d|D}^{\Lambda,\Lambda} - \mathbb{1}_d \bar{\mu}_{d|D}. \quad (88)
\]

By \(R_{1,d} = \mathbb{E}[(y_d - \mathbb{1}_d x_d)(y_d - \mathbb{1}_d x_d)^\top | y^{\Lambda}, \hat{\theta}]\), we have
\[
R_{1,d}^{\Lambda,\Lambda} = \mathbb{E}[(y_d^{\Lambda} - \mathbb{1}_d x_d)(y_d^{\Lambda} - \mathbb{1}_d x_d)^\top | y^{\Lambda}, \hat{\theta}]
= R_{2,d}^{\Lambda,\Lambda}[(R_{2,d}^{\Lambda,\Lambda})^\top + \mathbb{1}_d \hat{\Sigma}_{d|D}(\mathbb{1}_d^{\Lambda})^\top],
\]
\[
R_{1,d}^{\Lambda,\Lambda} = \mathbb{E}[(y_d^{\Lambda} - \mathbb{1}_d x_d) | y^{\Lambda}, \hat{\theta}] = (R_{1,d}^{\Lambda,\Lambda})^\top
= R_{2,d}^{\Lambda,\Lambda}[(R_{2,d}^{\Lambda,\Lambda})^\top - \mathbb{1}_d (L_d^{\Lambda} \hat{\Sigma}_{d|D}) + \mathbb{1}_d \hat{\Sigma}_{d|D}^{\Lambda} (\mathbb{1}_d^{\Lambda})^\top],
\]
\[
R_{1,d}^{\Lambda,\Lambda} = \mathbb{E}[(y_d^{\Lambda} - \mathbb{1}_d x_d) | y^{\Lambda}, \hat{\theta}]
= (R_{1,d}^{\Lambda,\Lambda})^\top + \mathbb{1}_d \hat{\Sigma}_{d|D}(\mathbb{1}_d^{\Lambda})^\top + R_{2,d}^{\Lambda,\Lambda}[(R_{2,d}^{\Lambda,\Lambda})^\top - L_d^{\Lambda} \hat{\Sigma}_{d|D}(\mathbb{1}_d^{\Lambda})^\top - (L_d^{\Lambda} \hat{\Sigma}_{d|D}(\mathbb{1}_d^{\Lambda})^\top)^\top].
\]
2.4 Application on News Impact Exploration

Let $Q_{d,t}$ be indicator of news, taking values in $\{0, \ldots, C\}$. When there is no news (impact), we will say $Q_{d,t} = 0$, and in this case $y_d = v_d \triangleq \log V_d$ (as defined before). The observable part can be written as $\Lambda_d = \{t : Q_{d,t} = 0\}$, and the non-observable part $\Lambda^c_d$ can be further partitioned into $\Lambda_{d,c} = \{t : Q_{d,t} = c\}$.

However, for such types of news as Witching day or Earning Release, it’s impact may last whole day long. It turns out therefore non trivial to define $Q_{d,t}$ precisely without studying it’s whole day impact. So to fully explore the news impact, we will assign one specific value $c \in \{0, \ldots, C\}$ to $Q_{d,t}$ for each time $t$ at day $d$. In turn, either $\Lambda_d$ or $\Lambda^c_d$ will be the empty set.

We use one year data for stocks in CAC40, from ’01-Mar-2009’ to ’01-Mar-2010’, with news of type Witching day, Earning Release and US-Employment Situation.

The EM Estimation is run for each stock separately till the log-likelihood converges. Below is one example illustrating the EM Estimation Procedure.

The daily dynamic parameters $A, B, C$ didn’t converge to the pre-defined precision, and was stopped after 2000 iterations. $A$ equals 0.772 at the last iteration, showing strong daily auto-correlation between two consecutive days. Larger $A$ implies smaller $B$ because of the nature of linear regression. The intra-day seasonality in described by $E$ and $F$, where the covariance matrix $F$ shows the average intra-day estimation error. The largest errors (variance) seem to happen at 2:30PM: a clear long ”vertical” error on the right bottom figure came from 07-Aug-2009:
This may be caused by some news event, however, we have only limited information, thus this part of error (residue) goes eventually into $F$. To make $F$ "small", more information is needed, especially around 2:30PM.

**Instant News Impact**

The Latent Kalman Filter model helps us define the instant impact of news based on (31):

$$u_{d,t} = \frac{v_{d,t} - \mu_d|D - E_t}{\sqrt{F_{t,t} + \Sigma_d|D}}.$$ \hfill (89)

When there is no news, i.e. $Q_{d,t} = 0$, $u_{d,t} | y^A \sim \mathcal{N}(0, 1)$. Similarly, we may do some statistics on $u_{d,t}$ when there is news. As we have only limited news for each stock in CAC40, we will group them together to form a larger population. Below are basic statistics based on $u_{d,t}$ for each news type:
Almost all stocks in CAC40 are impacted, especially starting from 2:30PM.
Almost all stocks in CAC40 are impacted whole day long, especially in the morning.
2.4.3 Witching day

Not all stocks in CAC40 are impacted at 12:00PM, further classification of stocks seems to be necessary. Impact at closing auction and at 4:00PM is also quite large, so Witching day’s impact might also last whole day, alike Earning Release.
3 Extension: Non-homogeneous Kalman Filter

Having studied the instant impact of news each the impact day, we can continue modelling two typical news impacts: Whole-day Impact and "Peak, Choc" Impact in order to relate the intra-day volume \( v_d \) with \( y_d \).

3.1 Whole-day Impact

In this part, we will model such news impact as Witching day and Earning releases whose impact might last whole-day long.

Let \( Q_d \) be the indicator of news, taking values in \( \{0, 1, \ldots, C\} \). The model we develop applies to multiple types of news, however, any news of different type should not take place on the same day, i.e. \( Q_d \in \{0, 1, \ldots, C\} \) (rather than \( Q_d \subseteq \{0, 1, \ldots, C\} \)). Let \( i(Q_d) = (i_i(Q_d))_i \), then

\[
v_d = y_d + i(Q_d).
\]  

Suppose \( i(c) = (I_i(c))_i \sim N(E_c, F_c) \) for \( c \in \{1, \ldots, C\} \), and \( i(0) = 0 \), then we get the following Non-homogeneous Kalman Filter,

\[
v_d|x_d \sim \begin{cases} N(1x_d + E, F) & \text{if } Q_d = 0 \\ N(1x_d + E + E_c, F + F_c) & \text{if } Q_d = c \in \{1, \ldots, C\}. \end{cases}
\]  

To simplify the notion, let's introduce \( E_0 = 0 \) and \( F_0 = 0 \). The filtering, updating and smoothing can be deduced similarly as in Latent Kalman Filter. Suppose \( x_d | v_1, \ldots, v_d \sim N(\mu_{d|d}, \Sigma_{d|d}) \), then

\[
\begin{align*}
x_{d+1} | v_1, \ldots, v_d & \sim N(\mu_{d+1|d}, \Sigma_{d+1|d}), \\
\text{with} & \quad \mu_{d+1|d} = A\mu_{d|d} + B, \quad \Sigma_{d+1|d} = C + A\Sigma_{d|d}A^\top;
\end{align*}
\]  

and

\[
\begin{align*}
\begin{pmatrix} x_{d+1} \\ v_{d+1} \end{pmatrix} | v_1, \ldots, v_d & \sim N\left( \begin{pmatrix} \mu_{d+1|d} \\ \xi_{d+1|d} \end{pmatrix}, \begin{pmatrix} \Sigma_{d+1|d} & (I\Sigma_{d+1|d})^\top \\ (I\Sigma_{d+1|d}) & (I\Sigma_{d+1|d})^\top + \Xi_{d+1|d} + I\Sigma_{d+1|d}I^\top \end{pmatrix} \right) \\
\text{with} & \quad \xi_{d+1|d} = I\mu_{d+1|d} + E + E_{Q_d}, \quad \Xi_{d+1|d} = F + F_{Q_d}.
\end{align*}
\]  

Thus the filtering and updating process can be summarised by

\[
\begin{align*}
\mu_{d+1|d} & = A\mu_{d|d} + B, \quad \Sigma_{d+1|d} = C + A\Sigma_{d|d}A^\top, \\
\mu_{d+1|d+1} & = \mu_{d+1|d} + K_{d+1}[v_{d+1} - \xi_{d+1|d}], \\
\Sigma_{d+1|d+1} & = \Sigma_{d+1|d} - K_{d+1}\Sigma_{d+1|d}K_{d+1}^\top,
\end{align*}
\]  

where \( K_{d+1} = (I\Sigma_{d+1|d})^\top(\Xi_{d+1|d} + I\Sigma_{d+1|d}I^\top)^{-1}. \)
The smoothing process is also straightforward, essentially the same as in Latent Kalman Filter. Suppose
\[ x_{d+1} \mid v_1, \ldots, v_D \sim N(\mu_{d+1|D}, \Sigma_{d+1|D}), \] (100)
then by \( J_d = \Sigma_{d|d}^{\top} A^{\top} \Sigma_{d+1|d}^{-1} \), we get
\[ x_d \mid v_1, \ldots, v_D \sim N(\mu_{d|D}, \Sigma_{d|D}), \] (101)
with \( \mu_{d|D} = \mu_{d|d} + J_d[\mu_{d+1|D} - \mu_{d+1|d}] \),
\( \Sigma_{d|D} = \Sigma_{d|d} + J_d(\Sigma_{d+1|D} - \Sigma_{d+1|d}) J_d^\top. \) (102)

Recall \( y_d \mid x_d \sim N(1 x_d + E, F) \), then we have
\[ y_d \mid v_1, \ldots, v_D \sim N(\nu_{d|D}, \Omega_{d|D}), \] (104)
with \( \nu_{d|D} = 1 \mu_{d|D} + E, \quad \Omega_{d|D} = F + 1 \Sigma_{d|D} 1\top; \) (105)

To do EM, let’s "decompose" the complete log likelihood:
\[ \log p(x, y, v|\theta) = \log p(x_0|\theta) + \sum_{d=0}^{D-1} \log p(x_{d+1}|x_d, \theta) + \sum_{d=1}^{D} \left[ \log p(y_d|x_d, \theta) + \log p(v_d|y_d, \theta) \right]. \] (106)

Compared to Latent Kalman Filter, the only difference is the term \( \log p(y_d|x_d, \theta) + \log p(v_d|y_d, \theta) \), which depends on \( Q_d \). When \( Q_d = 0 \),
\[ \log p(y_d|x_d, \theta) + \log p(v_d|y_d, \theta) \] (107)
\[ = -\frac{1}{2}(v_d - 1 x_d - E)^\top F^{-1}(v_d - 1 x_d - E) - \frac{1}{2} \log |F| - T \frac{1}{2} \log(2\pi). \] (108)
When \( Q_d = c \in \{1, \ldots, C\} \),
\[ \log p(y_d|x_d, \theta) + \log p(v_d|y_d, \theta) \] (109)
\[ = -\frac{1}{2}(y_d - 1 x_d - E)^\top F^{-1}(y_d - 1 x_d - E) - \frac{1}{2} \log |F| - T \frac{1}{2} \log(2\pi) \] (110)
\[ -\frac{1}{2}(v_d - y_d - E_c)^\top F_c^{-1}(v_d - y_d - E_c) - \frac{1}{2} \log |F_c| - T \frac{1}{2} \log(2\pi). \] (111)

Let \( D_c = \{ d : Q_d = c \} \), we get for \( c = 0 \),
\[ \sum_{d \in D_0} \mathbb{E} \left[ \log p(y_d|x_d, \theta) + \log p(v_d|y_d, \theta) \mid v, \tilde{\theta} \right] \] (112)
\[ \propto -\frac{1}{2} \sum_{d \in D_0} \text{tr} \left( F^{-1}[H_d - R_d E^\top - (R_d E^\top)^\top + EE^\top] \right) - \frac{|D_0|}{2} \log |F|, \] (113)
with \( H_d = \mathbb{E}[(v_d - 1 x_d)(v_d - 1 x_d)^\top \mid v, \tilde{\theta}], \quad R_d = \mathbb{E}[(v_d - 1 x_d) \mid v, \tilde{\theta}]; \) (114)
and for \( c \in \{1, \ldots, C\} \),
\[
\sum_{d \in D_c} \mathbb{E} \left[ \log p(y_d | x_d, \theta) + \log p(v_d | y_d, \theta) \mid v, \hat{\theta} \right] = \frac{1}{2} \sum_{d \in D_c} \text{tr} \left( F^{-1} \left[ H_d - R_d E^\top - (R_d E^\top)^\top + EE^\top \right] \right) - \frac{|D_c|}{2} \log |F| \tag{115}
\]
\[
\propto - \frac{1}{2} \sum_{d \in D_c} \text{tr} \left( F_{c}^{-1} \left[ M_d - L_d E^\top - (L_d E^\top)^\top + E_c E_c^\top \right] \right) - \frac{|D_c|}{2} \log |F_c| \tag{116}
\]
with \( H_d = \mathbb{E} \left[ (y_d - 1 x_d)(y_d - 1 x_d)^\top \mid v, \hat{\theta} \right] \), \( R_d = \mathbb{E} \left[ (y_d - 1 x_d)^\top \mid v, \hat{\theta} \right] \), \( M_d = \mathbb{E} \left[ (v_d - y_d)(v_d - y_d)^\top \mid v, \hat{\theta} \right] \), \( L_d = \mathbb{E} \left[ (v_d - y_d) \mid v, \hat{\theta} \right] \). \tag{118}

To estimate the parameter \( E, F, E_c \) and \( F_c \), the log likelihood can be written as
\[
l(\theta, \hat{\theta}) \propto - \frac{1}{2} \sum_{d=1}^{D} \text{tr} \left( F^{-1} \left[ H_d - R_d E^\top - (R_d E^\top)^\top + EE^\top \right] \right) - \frac{|D|}{2} \log |F| \tag{120}
\]
\[
\sum_{c=1}^{C} \left\{ - \frac{1}{2} \sum_{d \in D_c} \text{tr} \left( F_{c}^{-1} \left[ M_d - L_d E^\top - (L_d E^\top)^\top + E_c E_c^\top \right] \right) - \frac{|D_c|}{2} \log |F_c| \right\}. \tag{121}
\]

The estimation of \( E \) and \( F \) is essentially the same as that in Latent Kalman Filter. The maximum likelihood estimator for \( E_c \) is also straightforward,
\[
\hat{E}_c = \frac{1}{|D_c|} \sum_{c \in D_c} L_d \tag{122}
\]

However, to estimate \( E_c \) and \( F_c \) for each news type \( c \) is no longer so evident since in reality we have only a few news from each type, i.e. \( |D_c| \) is small. The problem becomes worse when \( |D_c| < T \), because the maximum likelihood estimator of \( F_c \) becomes degenerate. This trade-off leads to the following assumption
\[
F_c = \sigma_c^2 I_d. \tag{123}
\]

Then
\[
\hat{\sigma}_{c}^2 = \frac{1}{T |D_c|} \sum_{d \in D_c} \text{tr} \left[ M_d - L_d \hat{E}_c^\top - (L_d \hat{E}_c^\top)^\top + \hat{E}_c \hat{E}_c^\top \right] \tag{124}
\]
\[
= \frac{1}{T \left( 1/|D_c| \sum_{d \in D_c} \text{tr} M_d - \hat{E}_c \hat{E}_c^\top \right)} \tag{125}
\]

The last ”stone” is also the calculation of \( H_d, R_d, M_d, L_d \):

For \( d \in D_0 \),
\[
R_d = v_d - 1 \hat{\mu}_{d|D}, \quad H_d = R_d R_d^\top + 1 \hat{\Sigma}_{d|D} 1^\top; \tag{126}
\]

For \( d \notin D_0 \),
\[
R_d = \hat{v}_{d|D} - 1 \hat{\mu}_{d|D} = \hat{E}, \quad H_d = R_d R_d^\top + \hat{F}, \tag{127}
L_d = v_d - \hat{v}_{d|D}, \quad M_d = L_d L_d^\top + \hat{\Omega}_{d|D}. \tag{128}
\]
Example  Impact of Witching day and Earning Release, stock id 290428, from 01/01/2009 to 01/01/2010, 5 minute per interval.
3.2 "Peak, Choc" Impact

In reality, news impact may be instantaneous. We are to model one kind of such impact resembling "Peak" or "Choc". These can be observed on US-Employment Situation and US-Market Opening.

Let $Q_{d,t}$ be the indicator of news, taking values in $\{0, 1, \ldots, C\}$. Unlike Whole-day Impact, here we allow different types of news to take place on the same day, however, not at the same time, i.e. $Q_{d,t} \in \{0, 1, \ldots, C\}$.

Extending from Latent Kalman Filter, let’s define $\Lambda_{d,c} = \{t : Q_{d,t} = c\}$, and let $v^\Lambda_{d,c}$ be this part of $v_d$. "Peak, Choc" Impact leads to the following Non-homogeneous Kalman Filter

$$ y_d|x_d \sim \mathcal{N}(\mathbb{1}x_d + E, F), $$

$$ v^\Lambda_{d,c}|y_d \sim \mathcal{N}(y^\Lambda_{d,c} + \gamma_c\mathbb{1}_d, \sigma_c^2\mathbb{I}_d); $$

$$ \implies v_d|x_d \sim \mathcal{N}(\mathbb{1}x_d + E + \sum_{c=1}^C \gamma_c\mathbb{1}_{\Lambda_{d,c}}, F + \sum_{c=1}^C \sigma_c^2\mathbb{I}_{\Lambda_{d,c}}). $$

The filtering, updating and smoothing process are exactly the same as that in Whole-day Impact, as long as we replace the formula (95) by

$$ \xi_{d+1|d} = \mathbb{1}\mu_{d+1|d} + E + \sum_{c=1}^C \gamma_c\mathbb{1}_{\Lambda_{d,c}}, \quad \Xi_{d+1|d} = F + \sum_{c=1}^C \sigma_c^2\mathbb{I}_{\Lambda_{d,c}}. $$

To do EM, let’s again decompose the complete log likelihood:

$$ \log p(x, y, v|\theta) = \log p(x_0|\theta) + \sum_{d=0}^{D-1} \log p(x_{d+1}|x_d, \theta) + \sum_{d=1}^D \left[ \log p(y_d|x_d, \theta) + \log p(v_d|y_d, \theta) \right]. \tag{133} $$

Let $u^\Lambda_{d,0} = v^\Lambda_{d,0}, u^\Lambda_{d,c} = y^\Lambda_{d,c}$, the term $\log p(y_d|x_d, \theta) + \log p(v_d|y_d, \theta)$ now becomes

$$ \log p(y_d|x_d, \theta) + \log p(v_d|y_d, \theta) \tag{134} $$

$$ = -\frac{1}{2}(u_d - \mathbb{1}x_d - E)^T F^{-1}(u_d - \mathbb{1}x_d - E) - \frac{1}{2} \log |F| - \frac{T}{2} \log(2\pi) \tag{135} $$

$$ + \sum_{c=1}^C \left\{ -\frac{1}{2\sigma_c^2}(v^\Lambda_{d,c} - y^\Lambda_{d,c} - \gamma_c\mathbb{1}_{\Lambda_{d,c}})^T (v^\Lambda_{d,c} - y^\Lambda_{d,c} - \gamma_c\mathbb{1}_{\Lambda_{d,c}}) - \frac{|\Lambda_{d,c}|}{2} \log \sigma_c^2 - \frac{|\Lambda_{d,c}|}{2} \log(2\pi) \right\}. $$

Then

$$ l(\theta, \hat{\theta}) \propto -\frac{1}{2} \sum_{d=1}^D \text{tr} \left( F^{-1} [H_d - R_d E^T - (R_d E^T)^T + E E^T] \right) - \frac{|D|}{2} \log |F| \tag{136} $$

$$ + \sum_{c=1}^C \sum_{d=1}^D \left\{ -\frac{1}{2\sigma_c^2} \text{tr} \left[ M^\Lambda_{d,c} - \gamma_c L^\Lambda_{d,c} (\mathbb{1}_{\Lambda_{d,c}})^T - \gamma_c (\mathbb{1}_{\Lambda_{d,c}})^T L^\Lambda_{d,c} + \gamma_c^2 \mathbb{I}_{\Lambda_{d,c}} \right] - \frac{|\Lambda_{d,c}|}{2} \log \sigma_c^2 \right\}, $$

where $M^\Lambda_{d,c}$ and $L^\Lambda_{d,c}$ are matrices defined in the context.
where
\[
H_d = \mathbb{E}[(u_d - \mathbbm{1} x_d)(u_d - \mathbbm{1} x_d)^\top | \nu, \hat{\theta}], \quad R_d = \mathbb{E}[u_d - \mathbbm{1} x_d | \nu, \hat{\theta}],
\]
\[
M_d^{\Lambda_{\nu}\Lambda_{\nu}} = \mathbb{E}[(v_d^{\Lambda_{\nu}} - \mathbbm{1} y_d^{\Lambda_{\nu}})(v_d^{\Lambda_{\nu}} - \mathbbm{1} y_d^{\Lambda_{\nu}})^\top | \nu, \hat{\theta}], \quad L_d^{\Lambda_{\nu}} = \mathbb{E}[v_d^{\Lambda_{\nu}} - y_d^{\Lambda_{\nu}} | \nu, \hat{\theta}].
\] (138)

The calculations of $H_d$, $R_d$, $M_d^{\Lambda_{\nu}\Lambda_{\nu}}$, $L_d^{\Lambda_{\nu}}$ are no difficult than previous ones.

\[
L_d^{\Lambda_{\nu}} = v_d^{\Lambda_{\nu}} - \hat{\nu}_d^{\Lambda_{\nu}} \hat{\mu}_d, \quad M_d^{\Lambda_{\nu}\Lambda_{\nu}} = L_d^{\Lambda_{\nu}}(L_d^{\Lambda_{\nu}})^\top + \hat{\Omega}_{d|D}^{\Lambda_{\nu}\Lambda_{\nu}}; \quad R_d^{\Lambda_{\nu}} = v_d^{\Lambda_{\nu}} - \mathbbm{1} \hat{\nu}_d^{\Lambda_{\nu}} \hat{\mu}_d.
\] (139)

Similarly,

\[
H_d^{\Lambda_{\nu}0,\Lambda_{\nu}0} = \mathbb{E}[(v_d^{\Lambda_{\nu}} - \mathbbm{1} \mathbbm{1}^{\Lambda_{\nu}0} x_d)(v_d^{\Lambda_{\nu}} - \mathbbm{1} \mathbbm{1}^{\Lambda_{\nu}0} x_d)^\top | \nu, \hat{\theta}] = R_d^{\Lambda_{\nu}0} (R_d^{\Lambda_{\nu}0})^\top + \mathbbm{1} \hat{\Omega}_{d|D}^{\Lambda_{\nu}0\Lambda_{\nu}0},
\]

\[
H_d^{\Lambda_{\nu}0,\Lambda_{\nu}5} = \mathbb{E}[(v_d^{\Lambda_{\nu}} - \mathbbm{1} \mathbbm{1}^{\Lambda_{\nu}5} x_d)(y_d^{\Lambda_{\nu}} - \mathbbm{1} \mathbbm{1}^{\Lambda_{\nu}5} x_d)^\top | \nu, \hat{\theta}] = R_d^{\Lambda_{\nu}0} (R_d^{\Lambda_{\nu}5})^\top + \hat{\Omega}_{d|D}^{\Lambda_{\nu}0\Lambda_{\nu}5},
\]

Thus

\[
\hat{\gamma}_{\nu} = \sum_{d=1}^D (\mathbbm{1} \mathbbm{1}^{\Lambda_{\nu}})^\top L_d^{\Lambda_{\nu}} = \sum_{d=1}^D (\mathbbm{1} \mathbbm{1}^{\Lambda_{\nu}})^\top L_d^{\Lambda_{\nu}}
\]

\[
\frac{\sum_{d=1}^D (\mathbbm{1} \mathbbm{1}^{\Lambda_{\nu}})^\top L_d^{\Lambda_{\nu}}}{\sum_{d=1}^D |\Lambda_{d,\nu}|},
\] (141)

and

\[
\hat{\sigma}_{\nu}^2 = \frac{\sum_{d=1}^D \text{tr}[M_d^{\Lambda_{\nu}\Lambda_{\nu}} - \hat{\gamma}_{\nu} L_d^{\Lambda_{\nu}} (\mathbbm{1} \mathbbm{1}^{\Lambda_{\nu}})^\top - \hat{\gamma}_{\nu} (\mathbbm{1} \mathbbm{1}^{\Lambda_{\nu}})^\top L_d^{\Lambda_{\nu}} + \hat{\gamma}_{\nu}^2 (\mathbbm{1} \mathbbm{1}^{\Lambda_{\nu}})^\top (\mathbbm{1} \mathbbm{1}^{\Lambda_{\nu}})^\top]}{\sum_{d=1}^D |\Lambda_{d,\nu}|} = \frac{\sum_{d=1}^D \text{tr}(M_d^{\Lambda_{\nu}\Lambda_{\nu}})}{\sum_{d=1}^D |\Lambda_{d,\nu}|} - \hat{\gamma}_{\nu}^2.
\]
4 Conclusion

Dans ce mémoire, on a utilisé un modèle très classique - Filtre de Kalman - pour analyser l’impact de news. L’idée originale provient de la discussion avec mes deux directeurs : Charles et Romain, qui ont grands expériences en modélisation du marché. Charles m’a proposé de modéliser intra-day volume par un Gaussien multi-varié, tandis que Romain m’a conseillé de caractériser le volume journalié par AR(1). Puisque News rend les données très "arbitraires", et il y a aussi des intervalles de volume qui nous donnent log 0, je propose finalement qu’on met tous ces variables latentes.

On avait supposé, au lieu de la formule (5),

$$\sum_{t=1}^{T} \exp E_t = 1,$$

(142)

ce qui rend le problème de Lagrangien en (76) non-linaire. Mais ce sera intéressant si on peut trouver une solution à ce restreint.

De plus, la combinaison de Latent Kalman Filter avec les deux extensions me parait aussi faisable: il suffit des hypothèses pour identifier les différents news impacts.

On peut aussi introduire une variable de contrôle dans le Filtre de Kalman pour avoir notre stratégie de trading.

Il y a toujours plein de choses à faire, surtout dans la recherche.

References