

Utilisation de modèles linéaires  
dynamiques pour l'optimisation de  
portefeuilles, une approche bayésienne

Mohammed Haddou      Jean-François Angers  
Robert Cléroux

**CRM-2761**

octobre 2001



## Résumé

Ce travail porte sur l'optimisation de portefeuilles. Le rapport se divise en trois parties principales. Dans la première partie, on fait une étude de prévision, c'est-à-dire qu'on prédit à chaque étape la prochaine valeur des rendements. Un modèle linéaire dynamique est utilisé pour cette fin et les densités prédictives des rendements sont obtenues, ce qui permet d'avoir un estimé de la prochaine valeur des rendements et de la variance de ces derniers. Nous avons aussi développé des intervalles de confiance conjoints pour les valeurs prédites. La deuxième partie est consacrée au choix du portefeuille qui permet un grand rendement avec le moins de risque possible, le risque étant mesuré en termes de variance. Le portefeuille optimal est obtenu en résolvant un problème de programmation non linéaire. Enfin, dans la dernière partie on effectue une étude de sensibilité pour mesurer l'effet de certains hyperparamètres sur la solution. Une application numérique où cinq devises fortes sont considérées illustre la méthode utilisée.



# 1 Introduction

On s'intéresse à l'optimisation de portefeuilles. Une attention particulière est donnée au cas où les titres sous étude sont des devises fortes.

Considérons un investisseur désirant répartir une somme d'argent  $A_0$  entre des titres de rendements aléatoires  $R_1, \dots, R_n$ . On appellera alors portefeuille tout vecteur  $w$  de  $\mathbb{R}^n$  noté  $w = (w_1, \dots, w_n)'$  avec l'interprétation suivante :  $w_i =$  fraction de la richesse investie dans l'actif  $i$ ;  $i = 1, \dots, n$  avec  $\sum_{i=1}^n w_i = 1$ .

En fin de période, la richesse investie dans le portefeuille  $w$  donne :

$$A = A_0 \times \sum_{i=1}^n w_i R_i.$$

Nous pouvons aussi tenir compte de la possibilité de contraintes quantitatives sur le portefeuille, comme par exemple l'interdiction de positions à découvert sur certains titres, c'est-à-dire  $w_i \geq 0$  pour certains  $i$ .

## Remarques.

1. Pour notre étude nous supposons que tous les actifs sont avec risque.
2. On peut aussi dans la définition de portefeuille supposer des  $w_i$  qui ne somment pas à un.

Dans le cas des devises fortes, si  $Z_t(i)$  désigne le taux de change de la devise  $i$  au temps  $t$ , le rendement de cette dernière au temps  $t$  est donné par :

$$R_t(i) = \frac{Z_t(i)}{Z_{t-1}(i)} - 1.$$

Soit  $R_t = (R_t(1), \dots, R_t(n))$  où  $R_t(i)$  désigne le rendement de la devise  $i$  au temps  $t$ . Il s'en suit donc que  $w_t' R_t = \sum_{i=1}^n w_t(i) R_t(i)$  n'est autre que le rendement réalisé au temps  $t$ .

Ce rapport se divise en trois sections principales. Dans la première section on étudie la prévision de la prochaine valeur des rendements. Pour cela on utilise un modèle linéaire dynamique (filtre de Kalman), l'approche est donc bayésienne. Ceci est fait en déterminant les densités prédictives des rendements à chaque temps  $t$ . Il s'avère que ces dernières sont des densités de Student multidimensionnelles dont les paramètres sont faciles à mettre à jour. Notre prédicteur pour la prochaine valeur sera donc :  $\hat{Y} = E(Y_t | D_{t-1})$ , de variance  $\hat{V} = Var(Y_t | D_{t-1})$ . Par la suite, on discute de la façon de choisir le portefeuille  $w_t$  qui permet un grand rendement avec le moins de risque possible ; le risque étant mesuré ici en termes des variances. Le portefeuille  $w_t$  est optimisé à chaque temps  $t-1$  de façon à ce que les rendements obtenus soient disponibles au temps  $t$ . Le portefeuille optimale est obtenu en résolvant un problème de programmation non linéaire. Enfin, nous faisons une étude de sensibilité pour mesurer l'effet de certains hyperparamètres sur la solution.

## 2 Prévision

Nous nous intéressons ici à la série  $\{Y_t, t = 1, 2, \dots\}$  où  $t$  représente le temps,  $Y_t \in \mathbb{R}^p$ , avec  $p > 1$ . Plus précisément, au temps  $t-1$  (le présent) nous désirons prédire  $Y_t$  (l'observation future). Notons par  $D_t$  l'ensemble de toute l'information (passée) disponible au temps  $t$ . Soit  $\{\mu_t, t = 1, 2, \dots\}$  un ensemble de paramètres ;  $\mu_t \in \mathbb{R}^p$ . Le paramètre  $\mu_{t-1}$  est défini tel que toute l'information passée nécessaire à la prévision d'observations futures soit contenue dans l'information apportée par  $\mu_{t-1}$ . Cette dernière information sera exprimée en termes de distributions conditionnelles ;  $\mu_{t-1} | D_{t-1}$ . En termes statistiques, sachant  $D_{t-1}$  (le présent),  $\mu_{t-1} | D_{t-1}$  est "suffisante" pour prédire le future. Ce qui implique que, sachant  $D_t$ , toute l'information concernant le future est contenu dans la distribution *a posteriori*  $\mu_t | D_t$ . Si, en particulier, cette distribution est, par exemple,  $N(m_t, \Sigma_t)$  ; normale de moyenne  $m_t$  et de matrice de variance-covariance  $\Sigma_t$ , alors sachant  $D_t$ , le couple  $(m_t, \Sigma_t)$  contient toute l'information sur le future.

Nous aurons aussi besoin à spécifier les suites de distributions suivantes :

$$\mu_t | \mu_{t-1}, D_{t-1}, \quad Y_t | \mu_t, D_{t-1} \quad \text{et} \quad \mu_{t-1} | D_{t-1}.$$

## 2.1 LE MODÈLE :

Notre modèle sera donc le suivant :

**Observation :**  $Y_t = \theta_t + v_t$ ;  $v_t \sim N_p(0, \Sigma)$  iid pour  $t = 1, 2, \dots$

**Paramètre :**  $\theta_t = \theta_{t-1} + w_t$ ;  $w_t \sim N_p(0, \frac{1}{n_0}\Sigma)$  iid pour  $t = 1, 2, \dots$

**Information initiale :**  $\theta_0 \mid \Sigma, D_0 \sim NW^{-1}(\alpha_0, m_0, n_0, S_0)$ ,  
c'est-à-dire  $\theta_0 \mid \Sigma, D_0 \sim N_p(m_0, \frac{1}{n_0}\Sigma)$ ,  $\Sigma \mid D_0 \sim W^{-1}(\alpha_0, S_0)$ .

où  $\alpha_0, S_0, n_0$ , et  $m_0$  sont supposés connus;  $S_0$  est une matrice  $(p \times p)$ , symétrique, définie positive,  $m_0$  est un vecteur de  $\mathbb{R}^p$  et  $\alpha_0, n_0$  sont des scalaires, avec bien entendu  $n_0 > 0$ .

### Remarques.

1.  $N_p$  désigne la loi normale  $p$ -dimensionnelle.
2.  $W^{-1}$  désigne la loi de Wishart inversée dont la densité est donnée par  $p(\Sigma) \propto |\Sigma|^{-\alpha_0/2+p} \times \exp\{-\frac{1}{2}tr\Sigma^{-1}(S_0)\}$  (voir Box & Tiao, p. 460.)
3. Les suites  $\{v_t\}$  et  $\{w_t\}$  sont supposées indépendantes et mutuellement indépendantes.
4. Les variables aléatoires  $\theta_t \mid D_{t-1}$  et  $v_t \mid D_{t-1}$  sont supposées indépendantes pour tout  $t$ . De même pour  $(\theta_{t-1} \mid D_{t-1})$  et  $(w_t \mid D_{t-1})$ .
5. Les équations du modèle supposent un conditionnement par  $D_{t-1}$ ; information passée incluant les données avant le temps  $t$ . On pourra, pour simplifier, noter  $D_t$  comme suite;  $D_t = \{D_0, Y_1, \dots, Y_t\} = \{D_{t-1}, Y_t\}$ .
6. Le modèle s'écrit aussi de la manière suivante :

$$\begin{aligned} Y_t \mid \theta_t, \Sigma, D_{t-1} &\sim N_p(\theta_t, \Sigma), \\ \theta_t \mid \theta_{t-1}, \Sigma, D_{t-1} &\sim N_p(\theta_{t-1}, \frac{1}{n_0}\Sigma), \\ \theta_0, \mid \Sigma, D_0 &\sim NW^{-1}(\alpha_0, m_0, n_0, S_0), \end{aligned}$$

### 2.1.1 Distribution *a posteriori* de $\theta_t$ sachant $\Sigma$ :

Dans cette section, nous présentons les équations de récurrence qui permettent de passer du temps  $t-1$  au temps  $t$  :

Pour  $m_{t-1}$  et  $n_{t-1}$  donnés, on a :

$$\theta_{t-1} \mid \Sigma, D_{t-1} \sim N_p(m_{t-1}, \frac{1}{n_{t-1}}\Sigma), \quad (1)$$

$$\theta_t \mid \Sigma, D_{t-1} \sim N_p(m_{t-1}, a_t\Sigma), \quad (2)$$

$$Y_t \mid \Sigma, D_{t-1} \sim N_p(m_{t-1}, b_t\Sigma), \quad (3)$$

$$\theta_t \mid \Sigma, D_t \sim N_p\left(m_t, \frac{1}{n_t}\Sigma\right), \quad (4)$$

$$\text{où } a_t = \frac{1}{n_{t-1}} + \frac{1}{n_0}, \quad b_t = a_t + 1, \quad n_t = 1 + \frac{1}{\frac{1}{n_0} + \frac{1}{n_{t-1}}},$$

$$m_t = \left(1 - \frac{1}{n_t}\right)m_{t-1} + \left(\frac{1}{n_t}\right)Y_t = m_{t-1} + \frac{1}{n_t}e_t, \quad e_t = Y_t - m_{t-1} \quad (5)$$

### 2.1.2 Distribution *a posteriori* pour $\Sigma$ :

Les distributions des variables aléatoires  $(\Sigma \mid D_t)$ ,  $t = 1, 2, \dots$ , sont obtenues par récurrence de la manière suivante : si pour  $\alpha_{t-1}$  et  $S_{t-1}$  donnés, la densité *a priori* au temps  $t - 1$  de  $\Sigma$  est :

$$\Sigma \mid D_{t-1} \sim W^{-1}(\alpha_{t-1}, S_{t-1}), \quad (6)$$

alors la densité *a posteriori* au temps  $t$  est donnée par :

$$\Sigma \mid D_t \sim W^{-1}(\alpha_t, S_t), \quad (7)$$

où

$$\alpha_t = \alpha_{t-1} + 1 = \alpha_0 + t, \quad S_t = S_{t-1} + b_t^{-1} e_t e_t'.$$

### 2.1.3 Marginalisation par rapport à $\Sigma$ :

La densité *a posteriori* de  $\theta_{t-1}$  au temps  $t - 1$  est donnée par :

$$\theta_{t-1} \mid D_{t-1} \sim T_p\left(\alpha_{t-1}, m_{t-1}, \frac{1}{\alpha_{t-1} n_{t-1}} S_{t-1}\right), \quad (8)$$

où  $T_p$  désigne la loi de Student  $p$ -dimensionnelle.

Pour résumer, les différentes densités au temps  $t$  sont :

**densité *a priori* au temps  $t$**

$$\theta_t \mid D_{t-1} \sim T_p(\alpha_{t-1}, m_{t-1}, \alpha_{t-1}^{-1} a_t S_{t-1}),$$

**densité prédictive**

$$Y_t \mid D_{t-1} \sim T_p(\alpha_{t-1}, m_{t-1}, \alpha_{t-1}^{-1} b_t S_{t-1}),$$

**densité *a posteriori* au temps  $t$**

$$\theta_t \mid D_t \sim T_p\left(\alpha_{t-1}, m_t, \frac{1}{\alpha_t n_t} S_t\right).$$

### 2.1.4 PREUVE :

On procède par récurrence dans les trois parties et on commence par les équations de la section 2.1.1.

(2) : (1) étant vraie (hypothèse de récurrence) on a que :

$$\theta_{t-1} \mid \Sigma, D_{t-1} \sim N_p\left(m_{t-1}, \frac{1}{n_{t-1}} \Sigma\right).$$

Comme  $\theta_t = \theta_{t-1} + w_t$ , avec  $(w_t \mid \Sigma, D_{t-1}) \sim N_p(0, \frac{1}{n_0} \Sigma)$  et, conditionnellement à  $\Sigma$  et  $D_{t-1}$ ,  $\theta_{t-1}$  et  $w_t$  sont indépendants, on a que :

$$\theta_t \mid \Sigma, D_{t-1} \sim N_p\left(m_{t-1} + 0, \frac{1}{n_{t-1}} \Sigma + \frac{1}{n_0} \Sigma\right) \equiv N(m_{t-1}, a_t \Sigma),$$

où

$$a_t = \frac{1}{n_{t-1}} + \frac{1}{n_0}.$$

(3) : On a  $\theta_t \mid \Sigma, D_{t-1} \sim N(m_{t-1}, a_t \Sigma)$  d'après l'équation (2) et  $v_t \mid \Sigma, D_{t-1} \sim N(0, \Sigma)$ , par hypothèse. Comme  $Y_t = \theta_t + v_t$ , il s'en suit que :

$$Y_t \mid \Sigma, D_{t-1} \sim N(m_{t-1} + 0, a_t \Sigma + \Sigma) \equiv N(m_{t-1}, b_t \Sigma),$$

où  $b_t = a_t + 1$ .

(4) : La formule de Bayes permet d'écrire :

$$p(\theta_t | \Sigma, D_t) \propto p(\theta_t | \Sigma, D_{t-1}) \times p(Y_t | \Sigma, D_{t-1}) \propto \exp(-Q/2),$$

où  $Q$  est donnée par :

$$\begin{aligned} Q &= (\theta_t - m_{t-1})'(a_t \Sigma)^{-1}(\theta_t - m_{t-1}) + (\theta_t - Y_t)' \Sigma^{-1}(\theta_t - Y_t) \\ &= \{\theta_t - A[(a_t \Sigma)^{-1}m_{t-1} + \Sigma^{-1}Y_t]\}' A^{-1}\{\theta_t - A[(a_t \Sigma^{-1})m_{t-1} + \Sigma^{-1}Y_t]\} + \text{Reste}, \end{aligned}$$

le reste ne dépend pas de  $\theta_t$  et

$$A = [(a_t \Sigma)^{-1} + \Sigma^{-1}]^{-1} = \frac{a_t}{a_t + 1} \Sigma.$$

Donc

$$\theta_t | \Sigma, D_t \sim N_p \left( m_t, \frac{1}{n_t} \Sigma \right),$$

où

$$\begin{aligned} \frac{1}{n_t} &= \frac{a_t}{a_t + 1}, \\ m_t &= \frac{a_t}{a_t + 1} \Sigma \times \Sigma^{-1} \left[ \frac{m_{t-1}}{a_t} + Y_t \right] = \left(1 - \frac{1}{n_t}\right) m_{t-1} + \frac{1}{n_t} Y_t = m_{t-1} + \frac{1}{n_t} e_t, \\ e_t &= Y_t - m_{t-1}. \end{aligned}$$

Montrons maintenant le résultat de la section 2.1.2.

Supposons que  $\Sigma | D_{t-1} \sim W^{-1}(\alpha_{t-1}, S_{t-1})$ , et montrons que  $\Sigma | D_t \sim W^{-1}(\alpha_t, S_t)$ , où

$$\alpha_t = \alpha_{t-1} + 1 \quad \text{et} \quad S_t = S_{t-1} + b_t^{-1} e_t e_t'.$$

La formule de Bayes permet d'écrire :

$$p(\Sigma | D_t) \propto p(\Sigma | D_{t-1}) \times p(Y_t | \Sigma, D_{t-1}).$$

Or, par hypothèse,  $p(\Sigma | D_{t-1}) \propto |\Sigma|^{-(\frac{\alpha_{t-1}}{2} + p)} \times \exp\{-\frac{1}{2} \text{tr}(\Sigma^{-1} S_{t-1})\}$  et d'après (3)  $p(Y_t | \Sigma, D_{t-1}) \propto |\Sigma|^{-1/2} \times \exp\{-\frac{1}{2} \text{tr}(b_t \Sigma)^{-1} (Y_t - m_{t-1})(Y_t - m_{t-1})'\}$

Donc,

$$\begin{aligned} p(\Sigma | D_t) &\propto |\Sigma|^{-(\alpha_{t-1} + 1)/2 + p} \exp\left\{-\frac{1}{2} \text{tr}(\Sigma^{-1} S_{t-1} + b_t^{-1} \Sigma^{-1} e_t e_t)\right\} \\ &\propto |\Sigma|^{-\alpha_t/2 + p} \times \exp\left\{-\frac{1}{2} \text{tr} \Sigma^{-1} (S_{t-1} + b_t^{-1} e_t e_t)\right\}, \end{aligned}$$

c'est-à-dire

$$\Sigma | D_t \sim W^{-1}(\alpha_t, S_t),$$

avec  $\alpha_t = \alpha_{t-1} + 1$  et  $S_t = S_{t-1} + b_t^{-1} \times e_t e_t'$ .

Pour la dernière section nous aurons besoin du résultat général suivant.

**Proposition.** Si  $\theta, \Sigma$  sont deux vecteurs aléatoires de  $\mathbb{R}^p$  tels que  $(\theta | \Sigma) \sim N_p(m, k\Sigma)$  et  $\Sigma \sim W^{-1}(\alpha, S)$  alors  $\theta \sim T_p(\alpha, m, \alpha^{-1}kS)$ .

*Démonstration.* En effet,  $p(\theta) = \int_{\Sigma} p(\theta, \Sigma) d\Sigma$ .



Ainsi

$$\begin{aligned}
p(\theta, \Sigma) &= p(\theta, \Sigma)p(\Sigma) \\
&\propto |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2} \text{tr}(k\Sigma)^{-1}(\theta - m)(\theta - m)'\right\} |\Sigma|^{-\left(\frac{\alpha}{2}+p\right)} \exp\left\{-\frac{1}{2} \text{tr}(\Sigma^{-1}S)\right\} \\
&\propto |\Sigma|^{-\frac{\alpha+1}{2}+p} \exp\left\{-\frac{1}{2} \text{tr}\Sigma^{-1}[k^{-1}(\theta - m)(\theta - m)' + S]\right\},
\end{aligned}$$

et  $p(\theta, \Sigma)$ , comme fonction de  $\Sigma$ , apparaît comme une  $W^{-1}(\alpha+1, k^{-1}(\theta - m)(\theta - m)' + S)$ . Par conséquent,

$$\begin{aligned}
p(\theta) &= \int_{\Sigma} p(\theta, \Sigma) d\Sigma \\
&\propto |k^{-1}(\theta - m)(\theta - m)' + S|^{-(\alpha+1+p-1)/2} \\
&\propto |S|^{-(\alpha+p)/2} |I_p + (kS)^{-1}(\theta - m)(\theta - m)'|^{-(\alpha+p)/2} \\
&\propto |I_p + (kS)^{-1}(\theta - m)(\theta - m)'|^{-(\alpha+p)/2}.
\end{aligned}$$

Comme  $|I_p - AB| = |I_k - BA| \forall A : (p \times k)$  et  $\forall B : (k \times p)$ , on a alors que :  $p(\theta) \propto |1 + (\theta - m)'(kS)^{-1}(\theta - m)|^{-(\alpha+p)/2}$ , ce qui implique que  $\theta \sim T_p(\alpha, m, k\alpha^{-1}S)$ .

Ceci étant établi, les résultats de la troisième partie découlent donc directement des deux premières.

### 3 Optimisation

Rappelons que notre but est d'obtenir un rendement élevé avec le moins de risque possible. Plusieurs approches sont possibles . Nous donnerons en premier un exemple d'approche standard puis nous définirons la fonction à maximiser avec laquelle on a travaillé.

Une approche standard consisterait à chercher le portefeuille optimale  $w_t$  (le prochain) pour un certain rendement  $r$  fixé à l'avance. En d'autres termes on se restreint aux portefeuilles  $w_t$  pour lesquels  $w_t' \hat{Y}_t = r$ . On choisira alors celui qui minimise la variance, à savoir  $w_t' \hat{V}_t w_t$ .

On veut donc résoudre le problème de programmation non linéaire suivant :

$$\begin{aligned}
\text{Prog 1 :} \quad & \min && w_t' \hat{V}_t w_t, \\
& \text{contrainte} && w_t' \hat{Y}_t = r.
\end{aligned}$$

On pourra aussi rajouter la contrainte  $\sum_{i=1}^p w_i = 1$  et résoudre le programme :

$$\begin{aligned}
\text{Prog 2 :} \quad & \min && w' \hat{V}_t w_t \\
& \text{contraintes} && w' \hat{Y}_t = r, \\
& && \sum_{i=1}^p w_i = w' \times 1 = 1,
\end{aligned}$$

où 1 désigne le vecteur de  $\mathbb{R}^p$  de composantes toutes égales à 1.

#### 3.1 Solution du programme 1 :

Soit  $\phi(w) = w'Vw$  (on omettra les indices pour alléger le texte). Le lagrangien est donné par :

$$L(w, \lambda) = \phi(w) + \lambda(w'Y - r) = w'Vw + \lambda(w'Y - r).$$

Comme  $w'Vw$  est une fonction convexe de  $w$  car  $V > 0$  (définie-positive), il s'en suit que  $L$  l'est aussi car le deuxième terme est linéaire en  $w$ . On est donc en présence d'un problème de programmation non linéaire *convexe*, il admet donc une solution unique qui est obtenu en résolvant le système suivant :

$$\begin{cases} \nabla_w L = 0, \\ \nabla_\lambda L = 0, \end{cases}$$

où  $\nabla$  désigne le gradient. Ça revient donc à résoudre le système :

$$2Vw + \lambda Y = 0, \quad (9)$$

$$w'Y - r = 0. \quad (10)$$

De (9) on tire  $w = -\frac{1}{2}\lambda V^{-1}Y$ , et en remplaçant cette valeur dans (10) on obtient :

$$Y' \left( -\frac{1}{2}\lambda V^{-1}Y \right) = r \implies \lambda^* = -2 \frac{r}{Y'V^{-1}Y}.$$

La solution est alors :  $w^* = \frac{r}{Y'V^{-1}Y} \times V^{-1}Y$ .

### 3.2 Solution du programme 2 :

Comme précédemment, on commence par définir le lagrangien, qui est donné par :

$$L(w, \lambda, \mu) = \phi(w) + \lambda(w'Y - r) + \mu(w'1 - 1) = w'Vw + \lambda(w'Y - r) + \mu(w'1 - 1).$$

$L$  est *convexe* car  $V > 0$  (définie-positive); les autres termes étant linéaires en  $w$  n'affectent pas la convexité. Donc la solution existe et est unique, et elle doit satisfaire le système suivant :

$$\begin{cases} \nabla_w L = 2Vw + \lambda Y + \mu 1 = 0, & (11) \\ \nabla_\lambda L = w'Y - r = 0, & (12) \\ \nabla_\mu L = w'1 - 1 = 0. & (13) \end{cases}$$

De (14) on tire  $w = -\frac{1}{2}V^{-1}(\lambda Y + \mu 1)$  et en substituant cette valeur dans (15) et (16) on aura les deux équations suivantes :

$$\begin{aligned} (Y'V^{-1}Y) \times \lambda + (Y'V^{-1}1) \times \mu &= -2r, \\ (1'V^{-1}Y) \times \lambda + (1'V^{-1}1) \times \mu &= -2, \\ \iff \begin{cases} (Y'V^{-1}Y) \times \lambda + (Y'V^{-1}1) \times \mu &= -2r, \\ \Delta \times \lambda &= -2[(1'V^{-1}1)r - (1'V^{-1}Y)], \end{cases} \end{aligned}$$

où  $\Delta = (1'V^{-1}1)(Y'V^{-1}Y) - (Y'V^{-1}1)^2$ . Ce qui donne  $\lambda^* = -21'V^{-1}(1r - Y)/\Delta$

Pour  $\mu$  on a :

$$\begin{aligned} (1'V^{-1}1)\mu &= -2 - (1'V^{-1}Y)\lambda \\ &= -2 \left[ 1 - \frac{(1'V^{-1}1)(Y'V^{-1}Y)r - (Y'V^{-1}1)^2}{\Delta} \right] \\ &= -\frac{2}{\Delta} [(1'V^{-1}1)(Y'V^{-1}Y) - (1'V^{-1}Y)r] \\ &\implies \mu^* = -\frac{2}{\Delta} \times Y'V^{-1}(Y - r1). \end{aligned}$$

Finalement  $w_t^* = \hat{V}_t^{-1}(\alpha_t \hat{Y}_t - \beta_t 1)$  avec,

$$\begin{aligned} \alpha_t &= 1'V^{-1}(r1 - Y), \\ \beta_t &= Y'V^{-1}(r1 - Y). \end{aligned}$$

**Remarque.**

Les programmes 1 et 2 ont une propriété très intéressante. En effet ils sont équivalents au programme suivant :

$$\begin{aligned} \max \quad & w'_t \hat{Y}_t, \\ \text{contrainte} \quad & w'_t \hat{V}_t w = \sigma^2. \end{aligned}$$

En d'autres termes, on cherche à maximiser le rendement avec un risque fixé en avance. En termes techniques, ce programme est le dual du programme 1. Il est clair que dans ce cas  $r$  et  $\sigma$  sont liés.

**3.3 Fonction choisie pour notre étude**

Considérons maintenant la fonction définie par :  $\phi(w) = \frac{w'Y}{(w'Vw)^{1/2}}$  où  $w$  et  $Y$  sont deux vecteurs de  $\mathbb{R}^p$ ,  $w$  étant bien entendu  $\neq 0$  et  $V$  est une matrice  $(p \times p)$  symétrique définie-positive. Cherchons donc les portefeuilles  $w$  qui maximisent cette fonction. Considérons donc le programme non linéaire suivant :

$$\begin{aligned} \text{Prog 3 :} \quad \max \quad & \phi(w) = \frac{w'Y}{(w'Vw)^{1/2}}, \\ \text{contrainte} \quad & \sum_{i=1}^p w_i = w' \times 1 = 1. \end{aligned}$$

Le lagrangien est donné par :

$$L(w, \lambda, \mu) = \phi(w) + \mu(w'1 - 1) = \phi(w) = \frac{w'Y}{(w'Vw)^{1/2}} + \mu(w'1 - 1).$$

On doit donc résoudre le système suivant :

$$\begin{cases} \nabla_w L = (w'Vw)^{-1/2}Y - \frac{w'Y}{(w'Vw)^{3/2}}Vw + \mu 1 = 0, \\ \nabla_\mu L = w'1 - 1 = 0. \end{cases}$$

En multipliant la première equation par  $w'$  et en tenant compte du fait que  $w'1 = 1$  on obtient  $\mu = 0$  et par suite on a :

$$\nabla_w L = 0 \iff Y - \frac{w'Y}{(w'Vw)}Vw = 0 \iff w = \left( \frac{w'Y}{w'Vw} \right)^{-1} V^{-1}Y, \quad (14)$$

c'est-à-dire  $w = (\text{scalaire}) \times V^{-1}Y$ . Pour obtenir la constante il suffit de multiplier les deux membres de la dernière equation par  $1'$ , on aura alors :  $1 = k \cdot 1'V^{-1}Y$ . Et par suite :  $k = \frac{1}{1'V^{-1}Y}$ .

**Remarques.**

1. Le lagrangien se réduit à la fonction  $\phi$ .
2. Si  $w$  est solution de  $\frac{\partial \phi(w)}{\partial w} = 0$ , alors  $k \times w$  l'est aussi pour tout  $k \neq 0$ .
3. Si l'on note  $w_0 = V^{-1}Y$ , alors  $\phi(kw_0) = \phi(w_0) = (Y'V^{-1}Y)^{1/2}$ , pour tout  $k > 0$ .

**3.3.1 Calcul de  $\frac{\partial^2 \phi(w)}{\partial w^2}$  :**

On sait que

$$\frac{\partial \phi(w)}{\partial w} = (w'Vw)^{-1/2}Y - \frac{w'Y}{(w'Vw)^{3/2}}Vw = g_1(w) - g_2(w).$$

Ainsi

$$\frac{\partial^2 \phi(w)}{\partial w^2} = \frac{\partial g_1(w)}{\partial w} - \frac{\partial g_2(w)}{\partial w},$$

où

$$\begin{aligned}\frac{\partial g_1(w)}{\partial w} &= \left[-\frac{1}{2}(w'Vw)^{-3/2}2Vw\right][Y]' + (w'Vw)^{-1/2}[0_{p \times p}]', \\ \frac{\partial g_2(w)}{\partial w} &= \left[\frac{\partial}{\partial w}\left(\frac{w'Y}{(w'Vw)^{3/2}}\right)\right] \left[\frac{\partial Vw}{\partial w}\right]'.\end{aligned}$$

Or

$$\begin{aligned}\frac{\partial}{\partial w}\left(\frac{w'Y}{(w'Vw)^{3/2}}\right) &= -\frac{3}{2}(w'Vw)^{-5/2}2Vww'Y + (w'Vw)^{-3/2}Y \\ &= (w'Vw)^{-3/2}\left[Y - 3\frac{w'Y}{w'Vw}Vw\right].\end{aligned}$$

Ce qui donne,

$$\frac{\partial g_2(w)}{\partial w} = (w'Vw)^{-3/2}\left[Yw'V - 3\frac{w'Y}{w'Vw}Vww'V + (w'Y)V\right].$$

D'où on a :

$$\frac{\partial^2 \phi(w)}{\partial w^2} = -(w'Vw)^{-3/2}\left[VwY' + Yw'V - 3\frac{w'Y}{w'Vw}Vww'V + (w'Y)V\right].$$

La dérivée seconde au point  $w = w^* = k \times V^{-1}Y$  vaut alors :

$$\frac{\partial^2 \phi(w)}{\partial w^2}\Big|_{w=w^*} = -\frac{1}{k^2} \frac{1}{(Y'V^{-1}Y)^{3/2}} [(Y'V^{-1}Y)V - YY'].$$

Soit  $x \in \mathbb{R}^p$ ,  $x \neq 0$ . Notons par  $H$  le hessien, c'est-à-dire la dérivée seconde au point  $w^*$ , on aura alors :

$$\begin{aligned}x'Hx &= -\frac{1}{k^2} \frac{1}{(Y'V^{-1}Y)^{3/2}} [(Y'V^{-1}Y)x'Vx - x'YY'x] \\ &= -\frac{1}{k^2} \frac{1}{(Y'V^{-1}Y)^{3/2}} [(Y'V^{-1}Y)x'Vx - (Y'x)^2]\end{aligned}$$

Or d'après l'inégalité de Cauchy-Schwarz on a que :

$$(Y'x)^2 \leq (Y'V^{-1}Y)(x'Vx)$$

car  $V$  est une matrice définie-positive et on a égalité quand  $Y = cVx$  c'est-à-dire ,  $x = cte \cdot V^{-1}Y$ . Donc finalement,  $x'Hx \leq 0 \forall x \neq 0$ .

Le hessien est donc une matrice semi-définie négative, les  $w$  qui annulent la dérivée sont alors des maximaux locaux, or il n'y en a qu'un seul, à savoir  $w^* = \frac{1}{1'V^{-1}Y} \cdot V^{-1}Y$ . Ceci implique que  $w^* = \operatorname{argmax}(\phi(w))$  et est tel que  $1'w = 1$ , c'est-à-dire  $\sum_{i=1}^n w_i = 1$ . Rappelons aussi que le maximum de cette fonction vaut  $\sqrt{Y'V^{-1}Y}$ .

### 3.4 Intervalles de confiance pour $\hat{Y}_t$ :

Pour construire un intervalle de confiance pour  $\hat{Y}_t$  on aura besoin des résultats suivants :

#### Lemme (Inégalité de Cauchy-Schwarz).

Soit  $a, b$  deux vecteurs de  $\mathbb{R}^p$  et  $\Sigma$  une matrice  $(p \times p)$  définie-positive, qu'on notera  $\Sigma > 0$ . Alors on a  $(a'b)^2 \leq (a'\Sigma a)(b'\Sigma^{-1}b)$  avec égalité lorsque  $a = cte \cdot \Sigma^{-1}b$ .

**Lemme.**

Soit  $f(a) = \frac{(a'b)^2}{a'\Sigma a}$ .

Du Lemme 1 on a :  $f(a) = \frac{(a'b)^2}{a'\Sigma a} \leq b'\Sigma^{-1}b \ \forall a \neq 0$ .

Donc  $b'\Sigma^{-1}b$  est un majorant de  $f(a)$ , et il est atteint pour  $a = cte\Sigma^{-1}b$ , c'est-à-dire :

$$\sup_{a \neq 0} \left\{ \frac{(a'b)^2}{a'\Sigma a} \right\} = f(cte.\Sigma^{-1}b) = b'\Sigma^{-1}b.$$

En d'autres termes :

$$b'\Sigma^{-1}b = \sup_{a \neq 0} \left\{ \frac{(a'b)^2}{a'\Sigma a} \right\} \quad \forall b \in \mathbb{R}^p \text{ et } \forall \Sigma > 0.$$

**Lemme.**

Si  $\theta \sim T_p(\hat{\theta}, \Sigma, \nu)$  alors la forme quadratique  $Q(\theta) = \frac{(\theta - \hat{\theta})'\Sigma^{-1}(\theta - \hat{\theta})}{p} \sim F(p, \nu)$ , où  $F(p, \nu)$  désigne la densité de Fisher de paramètres  $p$  et  $\nu$  (voir Box & Tiao (1973), p.117.)

Nous désirons ici trouver des intervalles de confiance simultanés pour les composantes de  $\theta \sim T_p(\hat{\theta}, \Sigma, \nu)$ .

Pour un niveau  $\alpha$  donné, nous avons que :

$$\begin{aligned} 1 - \alpha &= \Pr[Q(\theta) \leq F_\alpha(p, \nu)] \\ &= \Pr \left[ \frac{(\theta - \hat{\theta})'\Sigma^{-1}(\theta - \hat{\theta})}{p} \leq F_\alpha(p, \nu) \right] \\ &= \Pr \left[ (\theta - \hat{\theta})'\Sigma^{-1}(\theta - \hat{\theta}) \leq pF_\alpha(p, \nu) \right] \\ &= \Pr \left[ b'\Sigma^{-1}b \leq pF_\alpha(p, \nu) \right] \\ &= \Pr \left[ \sup_{a \neq 0} \left\{ \frac{(a'b)^2}{a'\Sigma a} \leq pF_\alpha(p, \nu) \right\} \right] \\ &= \Pr \left[ \frac{(a'b)^2}{a'\Sigma a} \leq pF_\alpha(p, \nu), \forall a \neq 0 \right] \\ &= \Pr \left[ \frac{|a'\theta - a'\hat{\theta}|}{\sqrt{a'\Sigma a}} \leq \sqrt{pF_\alpha(p, \nu)}, \forall a \neq 0 \right]. \end{aligned}$$

Donc on vient de construire un intervalle de confiance pour toute combinaison linéaire  $a'\theta$ , à savoir  $a'\hat{\theta} \pm [pF_p(p, \nu)]^{1/2}\sqrt{a'\Sigma a}$ .

En particulier, si on prend comme valeur de  $a$  les vecteurs colonnes de la matrice identité de  $\mathbb{R}^p$  on aura  $p$  intervalles de confiance simultanés pour les  $\theta_i$ , à savoir  $\hat{\theta}_i \pm [pF_p(p, \nu)]^{1/2}\hat{\sigma}_{ii}$ , où  $\hat{\sigma}_{ii}$  est l'élément  $(i, i)$  de la matrice  $\Sigma$ . Si  $E_i$  désigne l'événement :  $\theta_i \in [\hat{\theta}_i - [pF_p(p, \nu)]^{1/2}\hat{\sigma}_{ii}, \hat{\theta}_i + [pF_p(p, \nu)]^{1/2}\hat{\sigma}_{ii}]$ , alors  $P[\bigcap_{i=1}^p E_i] \geq 1 - \alpha$ .

Dans notre cas,  $\hat{Y}_t \sim T_p(m_{t-1}, \frac{b_t}{\alpha_{t-1}}S_{t-1}, \alpha_{t-1})$ , par conséquent

$$(\hat{\theta}, \Sigma, \nu) = \left( m_{t-1}, \frac{b_t}{\alpha_{t-1}}S_{t-1}, \alpha_{t-1} \right),$$

et un intervalle de confiance pour  $\hat{Y}_t(i)$  est alors donné par

$$m_{t-1}(i) \pm [pF_p(p, \alpha_{t-1})]^{1/2} \left[ \frac{b_t}{\alpha_{t-1}} s_{t-1}^2(i, i) \right]^{1/2} \text{ où } i = 1, 2, \dots, p.$$

Les figures suivantes montrent les bornes des intervalles de confiance au niveau  $\alpha = .05$  pour  $\hat{Y}_t$  (le prochain rendement),  $t = 1, \dots, 250$ , ainsi que les observations (points) et la courbe représentant les  $\hat{Y}_t$ . Les devises étudiées sont respectivement : le dollar canadien ; \$ CA, le franc français ; FF, le Mark allemand ; DEM, le Yen japonais ; ¥ et la livre sterling ; £.

Toutes ces devises ont été calculées relativement au dollar US. Ces courbes ont été obtenues pour  $p = .2$ .

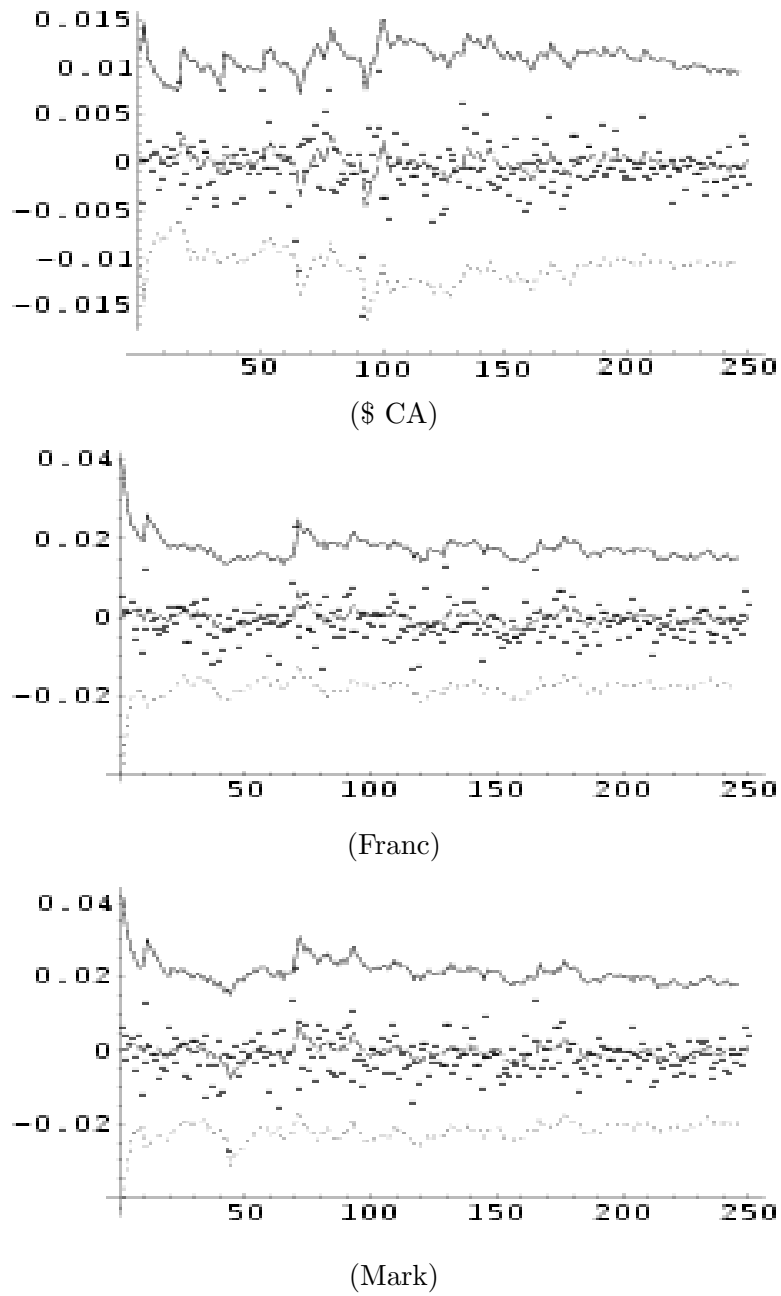


FIG. 1 – Intervalles de confiance pour le rendement du dollars Canadien, du Franc français et du Mark allemand par rapport au dollars américain.

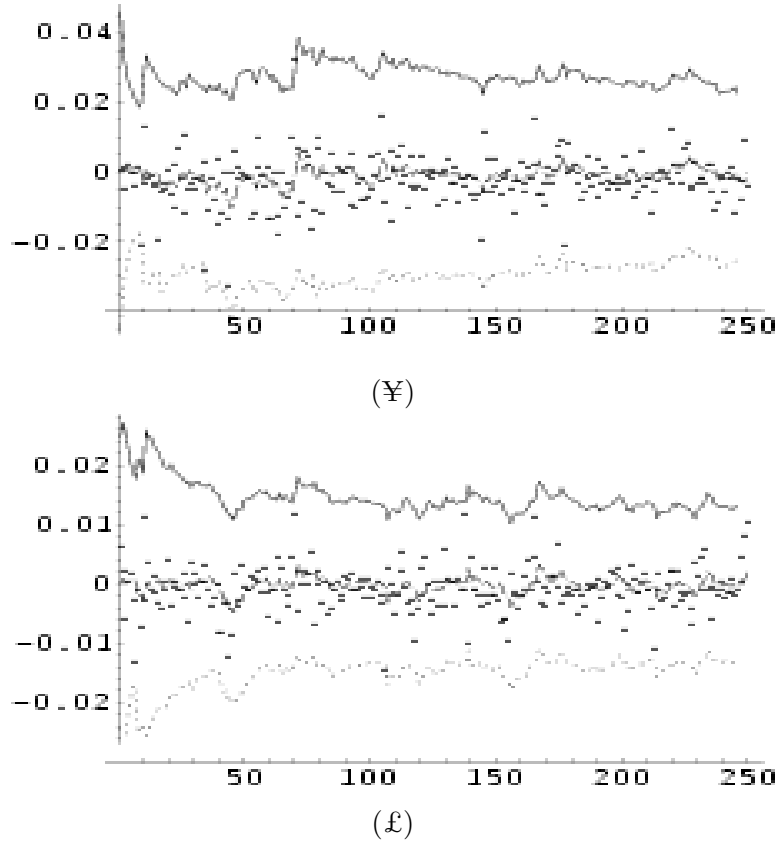


FIG. 2 – Intervalles de confiance pour le rendement du Yen et de la livre sterling par rapport au dollars américain.

## 4 Étude de sensibilité

Dans cette partie nous nous intéressons à l'effet de  $m_0$ ,  $S_0$ ,  $\alpha_0$  et  $n_0$  (hyperparamètres) sur la solution et par conséquent au bon choix de  $\hat{Y}_t$  et  $\hat{V}_t$ . Rappelons ici que  $\hat{Y}_t = m_{t-1}$ , donc ce dernier ne dépend que de  $m_0$ ,  $n_0$  et des observations ;  $Y_t$ . En effet  $m_t = (1 - \frac{1}{n_t})m_{t-1} + \frac{1}{n_t}Y_t$ , ce qui donne

$$m_t = \left(1 - \frac{1}{n_t}\right) \left(1 - \frac{1}{n_{t-1}}\right) \cdots \left(1 - \frac{1}{n_1}\right) m_0 + S,$$

où  $S$  ne dépend que de  $n_0$  et de  $Y_t$ . Il est clair alors que l'effet de  $m_0$  s'affaiblit avec le temps ; le coefficient de  $m_t$  devenant négligeable.

D'un autre côté

$$S_t = S_{t-1} + b_t^{-1} e_t e'_t = S_0 + \sum_{j=1}^t b_j^{-1} e_j e'_j,$$

donc l'effet de  $S_0$  devient négligeable après un certain temps. Et pour  $t$  assez grand on aura

$$\hat{V}_t = \frac{b_t S_{t-1}}{\alpha_{t-1} - 2} = \frac{b_t S_{t-1}}{\alpha_0 + t - 3} \approx \frac{b_t S_{t-1}}{t - 3}.$$

Il devient alors clair de ce qui précède que le seul hyperparamètre qui influe sur la solution ( $\hat{Y}_t$  et  $\hat{V}_t$ ) est  $n_0$ . Le paramètre  $n_0$  intervient dans les expressions de  $n_t$  et  $b_t$  qui à leur tour rentrent dans les expressions de  $\hat{Y}_t$  et  $\hat{V}_t$ . Comme  $b_t = 1 + \frac{1}{n_{t-1}} + \frac{1}{n_0}$ , il s'en suit donc que le bon choix de  $\hat{Y}_t$  et  $\hat{V}_t$  dépendra du comportement de la suite  $\{n_t\}$  ou de façon équivalente de la suite  $\{\frac{1}{n_t}\}$ .

#### 4.1 Étude de la suite $\frac{1}{n_t}$ :

Rappelons les expressions suivantes :

$$\frac{1}{n_t} = \frac{a_t}{a_t + 1}; \quad a_t = \frac{1}{n_{t-1}} + \frac{1}{n_0}; \quad n_t = 1 + \frac{1}{\frac{1}{n_{t-1}} + \frac{1}{n_0}}.$$

Il est évident que  $0 < \frac{1}{n_t} < 1$  car  $n_0$  est supposé  $> 0$ . Donc la suite  $\{\frac{1}{n_t}\}$  est bornée. Montrons qu'elle est aussi monotone.

On a :

$$\frac{1}{n_t} - \frac{1}{n_{t-1}} = \frac{a_t}{a_t + 1} - \frac{a_{t-1}}{a_{t-1} + 1} = \frac{1}{(a_t + 1)(a_{t-1} + 1)} \left( \frac{1}{n_{t-1}} - \frac{1}{n_{t-2}} \right).$$

Comme  $\frac{1}{(a_t+1)(a_{t-1}+1)} > 0$ , la suite est donc monotone. Elle est strictement croissante si  $\frac{1}{n_1} > \frac{1}{n_0}$ , c'est-à-dire  $n_0 > 2$ , strictement décroissante si  $n_0 < 2$  et constante égale à 2 si  $n_0 = 2$ .

Bornée et monotone, elle est donc convergente. Notons par  $p$  (poids) sa limite.

#### 4.2 Calcul de $p$ :

Comme  $n_t = 1 + \frac{1}{\frac{1}{n_{t-1}} + \frac{1}{n_0}}$ ,  $p$  vérifie donc la relation :

$$\frac{1}{p} = 1 + \frac{1}{p + \frac{1}{n_0}}$$

ce qui donne après simplifications :  $p^2 + \frac{1}{n_0}p - \frac{1}{n_0} = 0$ . La solution est donnée par :

$$p = \frac{2}{1 + \sqrt{4n_0 + 1}} \iff n_0 = \frac{1}{p} \left( \frac{1}{p} - 1 \right) = \frac{1-p}{p^2}.$$

La convergence de la suite  $\{\frac{1}{n_t}\}$  se trouve être très rapide. En effet pour  $n_0 \neq 2$  on a que :

$$\begin{aligned} n_t - \frac{1}{p} &= 1 + \frac{1}{\frac{1}{n_{t-1}} + \frac{1}{n_0}} - 1 - \frac{1}{p + \frac{1}{n_0}} \\ &= \frac{pn_0^2}{(n_0 + n_{t-1})(n_0p + 1)} \times \left[ n_{t-1} - \frac{1}{p} \right] \\ \left| n_t - \frac{1}{p} \right| &= \frac{pn_0^2}{(n_0 + n_{t-1})(n_0p + 1)} \times \left| n_{t-1} - \frac{1}{p} \right|. \end{aligned}$$

Si  $n_0 > 2$  alors  $\{n_t\}$  est décroissante et on a :  $p < n_{t-1} < n_0$ . Sinon,  $\{n_t\}$  est croissante et on a :  $n_0 < n_{t-1} < p$ .

On aura donc

$$\left| n_t - \frac{1}{p} \right| \leq C \left| n_{t-1} - \frac{1}{p} \right|,$$

où  $C = C(n_0, p) = \frac{n_0}{\min[2n_0, n_0+p]} \times \frac{pn_0}{pn_0+1} < 1$ .

Ce qui donne :

$$\left| n_t - \frac{1}{p} \right| < C^{t-1} \left| n_1 - \frac{1}{p} \right|.$$

Remarquons que dans tous les cas  $C$  est le produit de deux quantités positives strictement inférieures à 1. Notre prédicteur  $\hat{Y}_t$  tend alors rapidement vers  $\hat{Y}_t = (1-p)m_{t-1} + pY_t$ . Le choix de  $\hat{Y}_t$  dépend donc du bon choix de  $p$ , qui dépend à son tour de  $n_0$  car  $p = \frac{2}{\sqrt{4n_0+1}+1}$ .

Un bon prédicteur doit être, d'une part, sensible au changement dans la série, c'est-à-dire dans les observations et d'autre part tenir compte de l'état du système ; ne pas trop s'éloigner du  $m_{t-1}$ . Il y a donc ici un choix difficile à faire. Tout dépendra du bon choix de  $p$ .



### 4.3 Critères utilisés pour le choix de p :

Utilisation de la densité prédictive, à savoir

$$P(Y_1, Y_2, \dots, Y_n | D_0) = \prod_{t=1}^n P(Y_t | D_{t-1}).$$

Cette dernière est regardée comme fonction de p et elle est en fait une vraisemblance pour ce paramètre ; les modèles différent par leur  $n_0$  et donc par le p. Il s'agit donc de choisir le ou les p qui maximisent P. Les échantillons sur lesquelles on a travaillé suggèrent un poids p entre 0,2 et 0,4 en général (voir figure 3.)

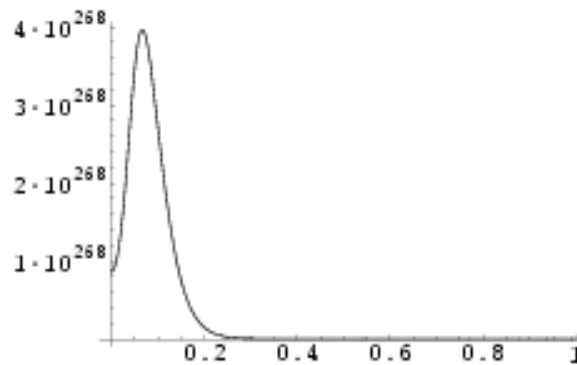


FIG. 3 – Densité prédictive fonction de p

## 5 Les données

Pour ce projet, on a travaillé avec plusieurs jeux de données. Les données utilisées pour illustration ont été prises d'un fichier contenant 2567 valeurs quotidiennes de taux de changes de douze devises fortes dont les valeurs sont données relativement au dollar US. Elles ont été prises sur une période de dix ans, du 9 octobre 1986 au 8 septembre 1996. On pourra trouver ces données sur le site web suivant :

[www.stat.duke.edu/datasets/mw/ts\\_data/all\\_exrates.html](http://www.stat.duke.edu/datasets/mw/ts_data/all_exrates.html).

Nous avons travaillé avec plusieurs échantillons et avec un nombre de devises différent à chaque fois. Les exemples donnés dans cet exposé concernent cinq devises seulement, à savoir : le dollar canadien, le franc français, le mark allemand, le yen japonais et la livre sterling. Toutes ces devises sont prises relativement au dollar US.

## 6 Conclusion

Nous nous sommes intéressés dans ce travail à l'optimisation de portefeuilles. Une attention particulière a été donnée au cas où ces derniers sont des devises fortes. Un modèle linéaire dynamique a été utilisé pour prédire la prochaine valeur des rendements. Des intervalles de confiance simultanés pour les composantes du vecteur rendement ont été développés. Aussi, pour obtenir le portefeuille optimal, un problème de programmation non linéaire a été résolu. Enfin une étude de sensibilité a été effectuée pour déterminer l'effet des hyperparamètres sur la solution. Il s'est avéré qu'un seul paramètre influait vraiment sur la solution. Plusieurs simulations ont été faites et un intervalle de valeurs possibles du dit paramètre autour du mode de la densité prédictive a été trouvé.

## Références

- [1] Box, G.E.P. et Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis*, Addison-Wesley Publishing Company.
- [2] Gauvin, J. (1995). *Leçons de programmation mathématique*, Éditions de l'École polytechnique de Montréal.
- [3] Johnson, L.J. et Kotz, S. (1972). *Distributions in Statistics : Continuous Multivariate Distributions*, John Wiley.
- [4] Quintana, J.M. et West, M. (1987). *An Analysis of International Exchange Rate using Multivariate DLM'S*, *The Statistician (1987)*, **36** 275–281.
- [5] West, M. et Harrison, J. (1999). *Bayesian Forecasting and Dynamic Models*, Springer-Verlag.