

# CHAPITRE 2

## Modélisation de la volatilité: Modèles MIDAS et données de haute fréquence

Michel LUBRANO

Avril 2011

### Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Rappels</b>	<b>3</b>
2.1	Stationnarité . . . . .	3
2.2	Fonction d'auto-covariance . . . . .	4
2.3	Opérateur retard . . . . .	4
2.4	Modèles à retards échelonnés . . . . .	5
<b>3</b>	<b>Analyse des propriétés dynamiques</b>	<b>6</b>
3.1	Calcul des coefficients de retard individuels . . . . .	7
3.2	Multiplicateurs . . . . .	8
3.3	Retard moyen . . . . .	8
<b>4</b>	<b>Inférence dans les modèles dynamiques</b>	<b>9</b>
4.1	Variables instrumentales . . . . .	10
4.2	La méthode du maximum de vraisemblance . . . . .	11
4.3	Première estimation par MLE . . . . .	13
4.4	Estimation du modèle général . . . . .	14
<b>5</b>	<b>La relation rendement risque</b>	<b>14</b>
5.1	Le CAPM . . . . .	14
5.2	Un estimateur alternatif de la volatilité . . . . .	15
5.3	Estimation du modèle MIDAS . . . . .	16

<i>CONTENTS</i>	2
<b>6 Prévision de la volatilité</b>	<b>17</b>
6.1 Régressions MIDAS . . . . .	18
6.2 Modèles concurrents . . . . .	19
6.3 Outils de comparaison . . . . .	20
6.4 Résultats avec données journalières . . . . .	20
6.5 MIDAS et les high frequencies . . . . .	22
<b>7 Conclusion</b>	<b>22</b>
<b>8 Exercice</b>	<b>23</b>

# 1 Introduction

Il s'agit de modéliser la volatilité conditionnelle

1. Rendements: différence entre deux cours mesurés à des temps différents.

$$r_t = s_t - s_{t-1}$$

2. Volatilité: la variance du processus

$$r_t \sim N(\mu_t, \sigma_t^2)$$

3. Comment estimer ces deux quantités. Doit-on utiliser des données de même fréquence?

Les modèles GARCH de Engle (1982) représentent l'approche la plus simple

$$\begin{aligned} r_t &= \mu_t + h_t \epsilon_t \\ \epsilon_t &\sim N(0, 1) \\ u_t &= h_t \epsilon_t \\ h_t &= \omega + \alpha u_{t-1}^2 + \beta h_{t-1} \end{aligned}$$

Dans ce modèle la volatilité est mesurée par le carré des rendements. On peut se poser la question de savoir si elle ne serait pas mieux mesurée par la valeur absolue des rendements. Ding et al (1993).

Une autre approche considère la Realised Volatility en utilisant des données de très haute fréquence (5 min).

La question que l'on pose ici c'est de savoir si l'on peut mesurer les rendements et la volatilité à la même fréquence ou s'il vaut mieux mélanger les fréquences.

Comment prévoir la volatilité?

## 2 Rappels

Il est utile de commencer par rappeler quelques notions de base sur les séries temporelles.

### 2.1 Stationnarité

On supposera pratiquement tout le temps que les séries sont stationnaires, c'est à dire

- $E(x_t) = \mu < \infty \forall t$
- $\text{Var}(x_t) = \sigma^2 < \infty \forall t$

- $\text{Cov}(x_t, x_{t-k}) = \gamma_k \forall t$

Ces moments sont finis et indépendants de du temps. Cela définit la stationnarité au second ordre. Une définition plus stricte de la stationnarité requiert que la distribution entière soit invariante par translation. Dans le cas de la normalité, les deux définitions sont confondues car la normale est parfaitement caractérisée par ses deux premiers moments.

Le modèle suivant décrit un processus qui est toujours stationnaire

$$x_t = \epsilon_t + \beta\epsilon_{t-1}.$$

En effet la moyenne est toujours nulle et la variance est constante et égale à  $\sigma^2(1 + \beta^2)$  quelle que soit la valeur des paramètres. Il s'agit d'un modèle moyenne mobile ou MA(1).

Il est par contre facile d'avoir un modèle non stationnaire en considérant par exemple

$$x_t = \mu + \delta t + \epsilon_t.$$

C'est un modèle avec tendance déterministe. La moyenne est égale à  $\mu + \delta t$ . Elle n'est pas constante dans le temps. La variance par contre est égale à  $\sigma^2$ .

## 2.2 Fonction d'auto-covariance

La fonction d'auto-covariance va servir à décrire la mémoire du processus. On la définit comme

$$\gamma_k = \mathbf{E}(y_t, y_{t-k}).$$

Il est utile de la normaliser par  $\sigma^2$  et d'en donner le graphique. Elle caractérise bien certains processus. Une transformation de cette fonction s'appelle la densité spectrale. Permet d'étudier les cycles dans le domaine des fréquences.

## 2.3 Opérateur retard

Comme les observations son indexées par le temps, il est commode de définir un opérateur qui permette de faire des opérations sur cet index temporel.

$$L^r y_t = y_{t-r}.$$

Une opération usuelle dans les séries temporelles consiste à prendre la différence première:

$$\Delta y_t = y_t - y_{t-1}.$$

On peut insérer cet opérateur dans des polynômes. On obtiendra alors des polynômes de retard comme par exemple

$$A(L) = 1 - \alpha L$$

qui va permettre d'écrire un modèle auto-régressif:

$$(1 - \alpha L)y_t = \epsilon_t.$$

L'inverse de ce polynôme se calcule facilement:

$$y_t = \frac{\epsilon_t}{A(L)} = \sum_{j=0}^{\infty} \alpha^j \epsilon_{t-j},$$

en utilisant la somme d'une progression géométrique. Cette écriture permet de voir deux choses. Premièrement, l'inversion nécessite que  $|\alpha| < 1$ . Deuxièmement cette condition est une condition de stationnarité pour le modèle auto-régressif. En effet, il faut utiliser cette forme pour calculer les moments et ceux-ci n'existent donc que si  $|\alpha| < 1$ .

## 2.4 Modèles à retards échelonnés

Un modèle à retards échelonnés se note:

$$y_t = \mu + \sum_{i=1}^s \beta_i x_{t-i} + u_t = \mu + B(L) x_t + u_t \quad (1)$$

Les coefficients  $\beta_i$  sont les coefficients de retards. Ils déterminent la façon dont  $y_t$  va répondre à un changement dans  $x_t$ . Dès lors que l'on suppose que les  $u_t$  sont des bruits blancs Gaussiens, il n'a pas de problème statistique particulier pour estimer les coefficients de ce modèle car les hypothèses usuelles des moindres carrés sont satisfaites et en particulier l'indépendance entre les régresseurs et le terme d'erreur. Toutefois une série temporelle évolue lentement à cause des effets mémoire si bien que les différents retards de la variable  $x_t$  auront tendance à être corrélés entre eux. On va donc se heurter à un problème de multi-colinéarité qui va gêner la précision dans l'estimation des coefficients de régression. On va imposer une structure particulière à la forme des coefficients de retard pour diminuer le nombre de paramètres à estimer. On résout le problème de multi-colinéarité en introduisant une information supplémentaire. Plusieurs structures sont possibles:

- Almon (1965): les coefficients sont contraints par un polynôme de degré  $n$  inférieur au nombre de retards, en général 2 ou 3. On aura

$$\beta_i = \sum_{j=1}^n \gamma_j i^j.$$

- retards échelonnés rationnels: la structure des retards est déterminée par le ratio de deux polynômes de retards:

$$y_t = \mu + \frac{B(L)}{A(L)} x_t + u_t$$

On consultera avec profit l'article de Griliches (1967). Ce type de modélisation est extrêmement souple dans la mesure où le cas le plus simple avec  $B(L) = \beta_0 + \beta_1 L$  et  $A(L) = 1 - \alpha L$  permet déjà une grande variété de configuration pour la structure des retards.

- retards géométriques ou de Koyck. C'est un cas particulier du précédent où l'on a posé  $B(L) = 1$  et  $A(L) = 1 - \alpha L$ . Les coefficients de retards décroissent de manière exponentielle avec la longueur du retard.

Examinons en détail ce dernier cas. Dans le modèle initial à retards échelonnés, on impose une structure particulière sur les coefficients avec

$$\beta_i = \delta \alpha^i \text{ avec } |\alpha| < 1 \quad (2)$$

Les valeurs des  $\beta_i$  décroissent très vite avec le temps. Aussi il n'est pas très restrictif de supposer un nombre de retards infini. Cela peut même être très commode pour les calculs. Écrivons le modèle:

$$\begin{aligned} y_t &= \mu + \sum_{i=1}^{\infty} \beta_i x_{t-i} + u_t \\ &= \mu + \delta \sum_{i=1}^{\infty} \alpha^i x_{t-i} + u_t \\ &= \mu + \delta \sum_{i=1}^{\infty} (\alpha L)^i x_{t-i} + u_t \end{aligned}$$

Il peut alors se mettre sous la forme:

$$y_t = \mu + \frac{\delta}{1 - \alpha L} x_t + u_t \quad (3)$$

en utilisant la définition de la somme d'une progression géométrique. En multipliant par  $(1 - \alpha L)$  les deux membres cette expression, on tombe sur un modèle auto-régressif:

$$y_t = (\mu - \alpha \mu) + \delta x_t + \alpha y_{t-1} + u_t - \alpha u_{t-1} \quad (4)$$

Ce modèle semble plus simple que le précédent et semble se prêter à une estimation commode, mais il comporte un terme d'erreurs auto-corrélées. L'estimation par moindres carrés est impossible à cause de la présence d'une variable endogène retardée, ce qui rend les variables dépendantes corrélées avec le terme d'erreur. Il faut avoir recours à une procédure par maximum de vraisemblance ou de variables instrumentales.

### 3 Analyse des propriétés dynamiques

Nous allons ici analyser la caractérisation de la dynamique de la partie systématique d'un modèle à retards rationnels du type

$$y_t = \frac{B(L)}{A(L)} x_t + u_t$$

On s'intéresse au cheminement moyen de  $y_t$  c'est à dire à l'espérance de  $y_t$  conditionnellement aux valeurs présentes et passées des exogènes. Comme les erreurs sont de moyenne nulle, on arrive à

$$E(y_t) = A^{-1}(L) B(L) x_t \quad (5)$$

On voit que cette espérance est une fonction des coefficients de retards d'un modèle à retards échelonnés infinis. Il s'agit donc de déterminer cette suite infinie de coefficients et

de tenter de la résumer au moyen de diverses caractéristiques. Pour cela il faut bien sûr que cette suite soit convergente, ce qui implique entre autres que les racines du polynôme  $A(L)$  soient toutes situées à l'extérieur du cercle unité, c'est à dire que le polynôme  $A(L)$  soit inversible. Les racines de ce polynôme correspondent à l'équation caractéristique  $A(z) = 0$ , c'est à dire

$$A(z) = 1 - \alpha_1 z - \alpha_2 z^2 - \dots - \alpha_r z^r = 0.$$

Une solution d'équilibre pour le modèle est alors possible étant donné que si  $x_t$  est constant,  $y_t$  sera aussi constant et l'on aura:

$$\bar{y} = \frac{\beta_0 + \beta_1 + \dots + \beta_s}{1 - \alpha_1 - \alpha_2 \dots - \alpha_r} \bar{x} = \frac{B(1)}{A(1)} \bar{x} \quad (6)$$

Les valeurs de  $A(1)$  et  $B(1)$  se trouvent en remplaçant l'opérateur  $L$  dans le polynôme par  $z$  et en posant  $z = 1$ . L'expression trouvée correspond à la somme des coefficients du polynôme.

### 3.1 Calcul des coefficients de retard individuels

L'effet total est donc obtenu comme le ratio entre la somme des coefficients de  $B(L)$  et la somme des coefficients de  $A(L)$ , c'est à dire sans calculer la suite des coefficients de retards. Mais nous pouvons parfois avoir besoin de celle-ci. Définissons:

$$D(L) = \frac{B(L)}{A(L)} = \delta_0 + \delta_1 L + \delta_2 L^2 + \dots \quad (7)$$

On peut calculer cette suite simplement par une opération de division de polynômes classique. On peut aussi procéder par identification au moyen de la formule:

$$B(L) = A(L) D(L) \quad (8)$$

Cette opération s'effectue de manière récursive en identifiant les termes des deux cotés. On va commencer par traiter un exemple où les deux polynômes sont de degré 1 avant de donner la formule générale. On part de

$$D(L) = \frac{\beta_0 + \beta_1 L}{1 - \alpha L} = \delta_0 + \delta_1 L + \delta_2 L^2 + \dots$$

d'où l'on tire

$$\begin{aligned} \beta_0 + \beta_1 L &= (1 - \alpha L)(\delta_0 + \delta_1 L + \delta_2 L^2 + \dots) \\ &= \delta_0 + (\delta_1 - \alpha \delta_0)L + (\delta_2 - \alpha \delta_1)L^2 + \dots \end{aligned}$$

L'identification des puissances de  $L$  des deux membres va donner les équations suivantes

$$\begin{aligned} \delta_0 &= \beta_0 \\ \delta_1 - \alpha \delta_0 &= \beta_1 \\ \delta_2 - \alpha \delta_1 &= 0 \\ &\dots \end{aligned}$$

d'où l'on tire la condition initiale  $\delta_0 = \beta_0$  et la solution de la récurrence devient

$$\begin{aligned}\delta_1 &= \beta_1 - \alpha\beta_0 \\ \delta_2 &= \alpha(\beta_1 - \alpha\beta_0) \\ \delta_3 &= \alpha^2(\beta_1 - \alpha\beta_0) \\ &\dots\end{aligned}$$

On en déduit la relation:  $\delta_j = \alpha^{j-1}(\beta_1 - \alpha\beta_0)$ .

Cette suite se généralise facilement pour  $A(L)$  et  $B(L)$  quelconques. Comme dans tous les cas on aura  $\alpha_0 = 1$ , il en résulte que l'on aura toujours la même condition initiale  $\delta_0 = \beta_0$ . Il vient ensuite la récurrence suivante:

$$\delta_j = \sum_{i=1}^{\min(j,r)} \alpha_i \delta_{j-i} + \beta_j \quad \text{si } 1 \leq j \leq s \quad (9)$$

$$\delta_j = \sum_{i=1}^{\min(j,r)} \alpha_i \delta_{j-i} \quad \text{si } j > s \quad (10)$$

### 3.2 Multiplicateurs

La notion de multiplicateur est définie dans le contexte d'un équilibre de long terme. On a

$$\bar{y} = \frac{B(1)}{A(1)} \bar{x}.$$

On envisage maintenant une perturbation unitaire de la valeur d'équilibre  $\bar{x}$ , ce qui fait que l'on passe de  $\bar{x}$  à  $\bar{x} + 1$ . Quel est la nouvelle valeur d'équilibre de  $y$ ? On a:

$$\bar{y}^* = \frac{B(1)}{A(1)} (\bar{x} + 1) = \bar{y} + \frac{B(1)}{A(1)}. \quad (11)$$

Donc  $B(1)/A(1)$  est le multiplicateur total étant donné qu'il représente la somme de tous les coefficients  $\delta_j$ . *Le multiplicateur d'impact* est le premier élément de la suite des  $\delta$ , c'est à dire  $\delta_0$ . *Le multiplicateur d'intérim* à l'ordre  $J$  est défini par la somme des  $J$  premiers termes de la suite des  $\delta_j$ :

$$\delta_J^* = \sum_{j=0}^J \delta_j \quad (12)$$

Il est commode de normaliser les multiplicateurs d'intérim par rapport au multiplicateur total de manière à pouvoir calculer le pourcentage de l'effet après  $J$  périodes, sachant qu'à l'infini l'effet sera total.

### 3.3 Retard moyen

Normalisons maintenant la suite des  $\delta_j$  par leur somme. Si tous les  $\delta_j$  sont positifs, alors la somme des  $\delta_j$  normalisés sera égale à un. Comme il s'agit de nombres positifs et normalisés à un, on peut alors les interpréter comme des probabilités définies sur la suite



entière des retards  $(0, 1, 2, \dots, \infty)$  et calculer certaines caractéristiques comme la moyenne et la médiane. Le retard moyen est alors défini par:

$$\mu_\delta = \frac{\sum_{j=0}^{\infty} j \delta_j}{\sum_{j=0}^{\infty} \delta_j} \quad (13)$$

Il faut prendre cette formule comme une définition, mais non comme une procédure de calcul, car elle nécessite le calcul de tous les  $\delta_j$ . Pour calculer le retard moyen, on lui préfère l'approche initiée par Griliches (1967). Considérons formellement  $D(z)$  avec  $D(1) = 1$  et tous les  $\delta_i$  positifs. Si  $D(z)$  converge pour  $-z_0 < z < z_0$ , alors  $D(z)$  est appelée une fonction génératrice de probabilités. Entre autres propriétés, on a que l'espérance de la variable qui est distribuée selon la distribution générée par  $D(z)$  est égale à la dérivée première de  $D(\cdot)$  calculée au point  $z = 1$ .<sup>1</sup> Appliquons ces formules à notre cas. Le retard moyen que met un choc sur  $x_t$  à se transmettre à  $y_t$  se calcule à partir de

$$\mu_\delta = \left. \frac{D'(L)}{D(L)} \right|_{L=1}.$$

Comme:

$$D'(L) = \frac{A(L) B'(L) - A'(L) B(L)}{A^2(L)}, \quad (14)$$

on arrive à:

$$\mu_\delta = \frac{D'(1)}{D(1)} = \frac{B'(1)}{B(1)} - \frac{A'(1)}{A(1)}. \quad (15)$$

Dans le cas du modèle d'ajustement partiel:

$$y_t = \alpha y_{t-1} + \beta x_t + \epsilon_t \quad (16)$$

on vérifie facilement à partir de (15) que le retard moyen est égal à  $\alpha/(1 - \alpha)$ . On appelle parfois  $\alpha$  le coefficient d'inertie. Quand  $\alpha$  est égal à zéro, il n'y a aucune inertie dans le modèle et l'ajustement se fait de manière instantanée. Pour  $\alpha = 1$ , le retard moyen devient infini.

## 4 Inférence dans les modèles dynamiques

L'inférence dans le modèle auto-régressif

$$y_t = \alpha y_{t-1} + u_t$$

ne pose pas de problème particulier. On peut appliquer les moindres carrés car la variable prédéterminée est indépendante des erreurs. En effet

$$\begin{aligned} \text{plim } \frac{1}{T} \sum y_{t-1} u_t &= \text{plim } \frac{1}{T} \sum (\alpha y_{t-2} + u_{t-1}) u_t \\ &= \text{plim } \frac{1}{T} \sum \alpha y_{t-2} u_t + \text{plim } \frac{1}{T} \sum u_{t-1} u_t \\ &= 0. \end{aligned}$$

<sup>1</sup>La variance fait intervenir la dérivée seconde de  $D(\cdot)$ , notée  $D''(\cdot)$ . Son expression est donnée par  $D''(1) + D'(1) - [D'(1)]^2$ .

Si l'on ajoute des variables exogènes, même retardées, le problème ne change pas. Considérons par exemple le modèle

$$y_t = \alpha y_{t-1} + \beta_1 x_t + \beta_2 x_{t-2} + u_t.$$

Pour estimer ce modèle, il suffit de construire deux matrices d'observations

$$y = [y_t] \quad X = [y_{t-1}, x_t, x_{t-1}]$$

Alors l'estimateur des moindres carrés sera formé par  $\hat{\gamma} = (X'X)^{-1}X'y$  où  $\gamma$  représente l'ensemble des paramètres de régression à estimer.

Par contre, le modèle à retards échelonnés va lui poser des problèmes d'inférence. La façon la plus simple de s'en apercevoir consiste à considérer le cas du modèle à retards géométriques

$$y_t = \beta \sum_{j=0}^{\infty} \alpha^j x_{t-j} + u_t.$$

On a vu que l'on pouvait résoudre ce modèle en utilisant un polynôme de retard en un modèle autorégressif, mais avec erreurs autocorrélées:

$$y_t = \alpha y_{t-1} + \beta x_t + v_t \quad v_t = u_t - \alpha u_{t-1}.$$

L'estimateur des OLS ne sera pas consistant car la variable endogène retardée est cette fois-ci corrélée avec le terme d'erreur. En effet:

$$\begin{aligned} \text{plim} \frac{1}{T} \sum y_{t-1} v_t &= \text{plim} \frac{1}{T} \sum y_{t-1} (u_t - \alpha u_{t-1}) \\ &= 0 - \alpha \text{plim} \frac{1}{T} \sum y_{t-1} u_{t-1} \\ &= -\alpha \text{plim} \frac{1}{T} \sum (y_{t-2} + \beta x_{t-1} + v_{t-1}) u_{t-1} \\ &= -\alpha \sigma^2, \end{aligned}$$

ce qui n'est plus égal à zéro, sauf si  $\alpha = 0$ , c'est à dire le cas statique. On va donc devoir utiliser une méthode d'estimation différente. La méthode la plus simple, mais la moins précise, consiste à utiliser des variables instrumentales. L'autre approche, plus précise, passe par le maximum de vraisemblance.

## 4.1 Variables instrumentales

On peut construire un estimateur par variables instrumentales en utilisant  $x_{t-1}$  comme instrument pour  $y_{t-1}$ . Ce n'est pas très efficace mais sert pour amorcer les itérations d'un MLE. On rappelle que quand on définit les moindres carrés, on suppose que les erreurs sont indépendantes des exogènes. C'est une condition d'identification qui dans le modèle

$$y_t = \alpha y_{t-1} + \beta x_t + v_t$$

s'exprime comme

$$\begin{aligned} E[(y_t - \alpha y_{t-1} - \beta x_t)x_t] &= 0 \\ E[(y_t - \alpha y_{t-1} - \beta x_t)y_{t-1}] &= 0. \end{aligned}$$

En résolvant ces deux équations où l'on a remplacé l'espérance par la moyenne empirique, on trouve l'estimateur des moindres carrés. On a deux équations pour deux inconnues. Si une de ces deux relations n'est plus valide, on ne peut plus calculer l'estimateur des moindres carrés. C'est le cas ici dans le modèle auto-régressif. Pour résoudre cette difficulté, on va remplacer la variable fautive par un instrument qui par conjecture sera corrélé avec  $y_{t-1}$ , mais non corrélé avec le terme d'erreur. Si  $x_{t-1}$  est choisi comme instrument, on va remplacer la deuxième condition par

$$E[(y_t - \alpha y_{t-1} - \beta x_t)x_{t-1}] = 0.$$

On va pouvoir définir l'estimateur par variable instrumentale de la manière suivante. Construisons tout d'abord les matrices d'observations:

$$Z = [x_{t-1}, x_t], \quad X = [y_{t-1}, x_t], \quad y = [y_t],$$

comportant  $T - 1$  lignes. Si l'on définit maintenant le paramètre  $\gamma' = [\alpha, \beta]$ , alors l'estimateur par variable instrumentale est égal à

$$\hat{\gamma}_{IV} = (Z'X)^{-1}Z'y.$$

Cet estimateur représente la version la plus simple de l'estimateur à variable instrumentale, car l'on a remplacé la variable fautive par un seul instrument. Dans le cas général on aura plus d'une variable instrumentale. On ne pourra résoudre aussi facilement les deux équations car la matrice  $Z'X$  ne sera plus carrée et inversible. On aura alors un problème de sur-identification et l'on fera appel à la méthode des variables instrumentales généralisées. Voir par exemple le chapitre 5 de Verbeek (2000) pour plus de détails.

## 4.2 La méthode du maximum de vraisemblance

La méthode du maximum de vraisemblance consiste à partir de la densité d'une observation, à en déduire la densité d'un échantillon complet et ensuite à trouver la valeur du paramètre qui rend maximale la probabilité d'avoir effectivement obtenu cet échantillon. Dans le cas IID, si la densité d'une observation est  $p(y_i; \theta)$  où  $\theta$  est le paramètre qui indexe cette densité, la fonction de vraisemblance se forme en considérant le produit

$$L(y_1, y_2, \dots, y_T) = \prod_{i=1}^T p(y_i; \theta).$$

L'estimateur du maximum de vraisemblance se trouve en résolvant les conditions du premier ordre. Pour des raisons de commodité de calcul, on préfère considérer le logarithme de la vraisemblance et donc:

$$\frac{\partial \log L(y; \theta)}{\partial \theta} = 0.$$

La matrice d'information se définit comme moins l'espérance de la dérivée seconde de la log vraisemblance

$$I(\theta) = -\mathbf{E} \frac{\partial^2 \log L(y; \theta)}{\partial \theta \partial \theta'}.$$

On ne peut facilement calculer cette espérance que dans des cas simples. Dans le cas général, on va se contenter d'une approximation asymptotique

$$IA(\theta) = \lim_{T \rightarrow \infty} \frac{1}{T} I(\theta).$$

Sous certaines conditions de régularité, l'estimateur du maximum de vraisemblance sera asymptotiquement normal avec

$$\sqrt{T}(\hat{\theta} - \theta) \sim \mathbf{N}(0, IA^{-1}(\theta)),$$

ce qui signifie que pour calculer la variance de  $\hat{\theta}$  on utilisera  $IA^{-1}(\theta)/T$ .

Quand les observations sont liées dans le temps, on ne peut plus prendre un simple produit pour trouver la densité de l'échantillon complet. C'est à dire qu'il n'est plus possible d'écrire  $f(y_1, y_2) = f(y_1) \cdot f(y_2)$ , mais par contre on peut toujours utiliser la décomposition  $f(y_1, y_2) = f(y_1|y_2) \cdot f(y_2)$ . On va donc maintenant devoir écrire:

$$\log L(y; \theta) = \sum_{t=2}^T \log l(y_t|y_{t-1}, \dots, y_1) + \log l(y_1).$$

On aura donc une suite de densités conditionnelles, suivie par la densité marginale de la première observation. Harvey (1981) montre comment on peut remplacer la première partie de cette écriture par la densité des erreurs de prévision, c'est à dire la densité des résidus récurrents. Traitons par exemple le modèle autorégressif  $y_t = \alpha y_{t-1} + \epsilon_t$ . La densité de la première observation est

$$f(y_1) = f_N \left( y_1 | 0, \frac{\sigma^2}{1 - \alpha^2} \right),$$

et la densité des résidus récurrents  $e_t$

$$f(e_t) = f_N(e_t | y_t - \alpha y_{t-1}, \sigma^2).$$

La log vraisemblance sera donc

$$\begin{aligned} \log L(y) &\propto \frac{1}{2} \log(1 - \alpha^2) - \frac{1 - \alpha^2}{2\sigma^2} y_1^2 \\ &\quad - \frac{T}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=2}^T (y_t - \alpha y_{t-1})^2. \end{aligned}$$

On voit tout de suite que si l'on néglige la distribution de la première observation, on retombe sur la solution des moindres carrés, c'est à dire la minimisation d'une somme de carrés d'erreurs.

### 4.3 Première estimation par MLE

Considérons en premier le modèle à retards échelonnés de Koyck:

$$y_t = \beta \sum_{j=0}^{\infty} \alpha^j x_{t-j} + \epsilon_t$$

Les erreurs  $\epsilon_t$  étant IID, l'estimation par MLE peut se faire en minimisant une somme de carrés des erreurs, comme si l'on était dans un modèle de régression linéaire à erreurs normales usuelles. Cependant, l'écriture directe est de peu d'utilité car elle fait intervenir une somme infinie de retards. On va décomposer donc la somme infinie en deux parties

$$y_t = \beta \sum_{j=0}^{t-1} \alpha^j x_{t-j} + \alpha^t [\beta \sum_{j=0}^{\infty} \alpha^j x_{0-j}] + \epsilon_t$$

Ce qui peut s'écrire encore

$$y_t = \beta x_t^*(\alpha) + \alpha^t \mathbf{E}(y_0) + \epsilon_t$$

où  $x_t^*(\alpha) = \sum_{j=0}^{t-1} \alpha^j x_{t-j}$  doit être construit de manière récursive en partant de  $x_1^*(\alpha) = x_1$ :

$$x_t^*(\alpha) = x_t + \alpha x_{t-1}^*(\alpha).$$

Il faut régler le problème des conditions initiales, c'est à dire donner une valeur à  $\mathbf{E}(y_0)$ . Trois solutions sont possibles.

- On peut les fixer égales à zero. On minimise alors simplement

$$S(\alpha, \beta) = \frac{1}{T} \sum (y_t - \beta x_t^*(\alpha))^2$$

Il y a bien sûr un léger biais dans l'estimation. Celui-ci s'atténue en grand échantillon ou pour des valeurs faibles de  $\alpha$  car les conditions initiales rentrent au moyen de  $\alpha^t$ .

- On peut estimer les conditions initiales en les traitant comme un paramètre supplémentaire. Le problème d'optimisation comporte alors un paramètre de plus, ce qui rend la recherche de l'optimum plus problématique.
- Enfin, on peut décider de fixer les conditions initiales égales à  $\mathbf{E}(y_1)$  en modifiant les indices de la sommation. Le modèle devient

$$y_t = \beta \sum_{j=0}^{t-2} \alpha^j x_{t-j} + \alpha^{t-1} \mathbf{E}(y_1) + \epsilon_t$$

On peut maintenant poser sans trop d'erreur  $\mathbf{E}(y_1) = y_1$ . L'inconvénient, c'est que l'on a utilisé une observation de moins dans la vraisemblance.

Dans tous les cas, il faut veiller à faire l'optimisation sous la contrainte que  $|\alpha| < 1$ .

#### 4.4 Estimation du modèle général

Le modèle précédent est relativement restrictif car il impose une structure très particulière des retards. Le modèle général se

$$y_t = \frac{B(L)}{A(L)}x_t + \epsilon_t,$$

alors que le modèle de Koyck correspond à  $B(L) = 1$  et  $A(L) = 1 - \alpha L$ . Pour estimer ce modèle général, il serait relativement difficile d'opérer comme précédemment. On préfère circonscrire la question des retards infinis en passant à la forme autorégressive au moyen de

$$A(L)y_t = B(L)x_t + A(L)\epsilon_t.$$

La procédure par maximum de vraisemblance consiste tout d'abord à construire la suite des résidus récursifs. Prenons l'exemple simple où les polynômes  $B(L)$  et  $A(L)$  sont tous les deux de degré 1. On aura la formule de récursion générale:

$$\hat{\epsilon}_t = y_t + \alpha y_{t-1} - \beta_0 x_t - \beta_1 x_{t-1} + \alpha \epsilon_{t-1}.$$

Il faudra amorcer cette récursion en posant par exemple  $\epsilon_1 = 0$  et en la démarrant en  $t = 2$  pour résoudre le problème des valeurs de départ pour  $x$  et  $y$ . On peut alors facilement évaluer la récursion suivante en prenant  $\epsilon_1 = 0$ :

$$\begin{aligned}\hat{\epsilon}_2 &= y_2 + \alpha y_1 - \beta_0 x_2 - \beta_1 x_1 + \alpha \epsilon_1 \\ \hat{\epsilon}_3 &= y_3 + \alpha y_2 - \beta_0 x_3 - \beta_1 x_2 + \alpha \epsilon_2 \\ \dots &= \dots\end{aligned}$$

Si on laisse tomber la question de la distribution de la première observation, la fonction de vraisemblance approchée se réduit à

$$\log L(y_T, \dots, y_2; \theta) \propto -\frac{T}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=2}^T \hat{\epsilon}_t^2.$$

## 5 La relation rendement risque

La finance théorique a mis en avant l'arbitrage entre rendement et risque, le paramètre central étant l'aversion au risque. Un investisseur acceptera un risque plus grand dans son choix de portefeuille si celui-ci lui rapporte en espérance un rendement supérieur. Cette relation classique est une relation statique que l'on a du mal à mettre à jour dans les études empiriques.

### 5.1 Le CAPM

Merton (1973) a proposé un modèle inter-temporel pour le CAPM (Capital Asset Pricing Model) qui propose la relation empirique suivante qui reste à tester:

$$E_t(R_{t+1}) = \mu + \gamma \text{Var}(R_{t+1}).$$

$R_{t+1}$  mesure en  $t + 1$  le rendement du portefeuille de marché. En finance on appelle rendement la variation du cours, c'est à dire  $S_{t+1} - S_t$  si  $S$  est le prix de l'actif financier. L'espérance est calculée en  $t$ . Le coefficient  $\gamma$  représente l'aversion au risque. Le cours du portefeuille de marché est donné par un indice. Le problème dans l'équation du haut c'est que la variance (ou volatilité) n'est pas observée et doit être estimée à partir des rendements passés.

On utilise des observations mensuelles pour mesurer l'espérance des rendements. Utiliser une fréquence plus haute (données journalières) conduirait à trop de bruit. On a donc résolu la question du membre gauche de l'équation. Pour estimer la variance des rendements, on se base en général sur des données journalières de la même variable (22 observations dans le mois). Appelons  $R_t$  les observations mensuelles des rendements et  $r_t$  les observations journalières de cette même variable et  $\hat{V}_t$  un estimateur de la variance future en  $t$ . L'estimateur usuel est basé sur

$$\hat{V}_t = \frac{1}{22} \sum_{d=0}^{22} r_{t-d}^2.$$

Cet estimateur est basé sur le principe que les variations journalières fournissent une bonne idée sur la variance du portefeuille de marché. Toutefois quand on l'utilise pour estimer la relation initiale, on trouve des valeurs pour  $\gamma$  qui sont tantôt positives, tantôt négatives et rarement significatives.

## 5.2 Un estimateur alternatif de la volatilité

Ghysel, Santa-Clara, and Valkanov (2005) ont proposé d'utiliser un estimateur basé sur les modèles à retards échelonnés. L'idée est que premièrement la variance peut être persistante et que donc un mois d'observations n'est pas suffisant et que deuxièmement le poids des observations passées doit décroître régulièrement. Ils proposent donc d'utiliser

$$V_t = 22 \sum_{d=0}^{\infty} w_d r_{t-d}^2$$

où les  $w_d$  sont des poids sur lesquels on va imposer une structure. On a vu que les retards de Koyck imposaient une structure de décroissance exponentielle au moyen d'un seul paramètre. Ghysel, Santa-Clara, and Valkanov (2005) proposent d'utiliser une structure décroissante à deux paramètres

$$w(d, \kappa) = \frac{\exp(\kappa_1 d + \kappa_2 d^2)}{\sum_{i=0}^{\infty} \exp(\kappa_1 i + \kappa_2 i^2)}$$

Ce schéma garantit des poids positifs à cause de l'exponentielle et dont la somme est égale à 1. D'autre part, les poids seront décroissants à partir d'un certain nombre de retards pour  $\kappa_2 < 0$ . Ghysel, Santa-Clara, and Valkanov (2005) choisissent de tronquer la suite infinie à 252, c'est à dire à environ une année d'observations journalières. On se retrouve finalement dans un modèle à retards échelonnés que l'on va estimer dans sa forme initiale en supposant que les erreurs sont indépendantes. Les auteurs font une hypothèse de normalité des erreurs,

basée sur cette hypothèse d'indépendance, mais supposent que la variance des erreurs varie dans le temps selon le schéma suivant:

$$R_{t+1} \sim N(\mu + \gamma V_t(\kappa_1, \kappa_2), V_t(\kappa_1, \kappa_2))$$

ce qui correspond à un modèle de régression à retards échelonnés polynômiaux et erreurs hétérosckélastiques un peu particulier

$$R_t = \mu + \gamma 22 \sum_{d=0}^{252} w(d, \kappa) r_{t-d} + U_t$$

avec  $\text{Var}(U_t) = V_t(\kappa_1, \kappa_2)$ . L'indice  $t$  représente des fréquences mensuelles pour  $R$  et  $U$ , alors qu'il représente des fréquences journalières pour  $r$ . L'hypothèse de normalité permet d'écrire la fonction de vraisemblance comme:

$$\log L(R_T, \dots, R_{12}; \theta) \propto -\frac{1}{2} \sum_{t=12}^T \log V_t(\kappa_1, \kappa_2) - \sum_{t=12}^T \frac{1}{2V_t(\kappa_1, \kappa_2)} \sum_{t=12}^T \hat{U}_t^2,$$

avec

$$\hat{U}_t = R_t - \mu - \gamma 22 \sum_{d=0}^{252} w(d, \kappa) r_{t-d}.$$

### 5.3 Estimation du modèle MIDAS

Ghysel, Santa-Clara, and Valkanov (2005) estiment ce modèle en utilisant des données américaines fournies par le Center for Research in Security Prices (CRSP) de Janvier 1928 à Décembre 2000. Toutes les estimations sont significatives comme l'indiquent les  $t$  tests

Table 1: Estimation par maximum de vraisemblance du modèle risque/rendement

Sample	$\mu$	$\gamma$	$\kappa_1 * 10^3$	$\kappa_2 * 10^5$
1928:01 - 2000:12	6.430	2.606	-5.141	-10.580
	[11.71]	[6.71]	[-4.53]	[-5.24]
1928:01 - 1963:12	11.676	1.547	-0.909	-10.807
	[5.89]	[3.38]	[-3.77]	[-2.11]
1964:01 - 2000:12	3.793	3.748	-6.336	-18.586
	[5.67]	[8.61]	[-7.86]	[-7.71]

Les statistiques de Student sont données entre crochets

donnés entre crochets. L'ordonnée à l'origine  $\mu$  capture l'effets des facteurs non présents dans l'équation. Le coefficient de risque  $\gamma$  est relativement stable à travers les deux périodes



considérées. Les poids décroissent lentement dans le temps. Ils représentent 33% d'effet après un mois, 56% après deux mois et 75% après trois mois. On est donc loin de la pratique courante qui consiste à ne prendre qu'un mois d'observations journalières pour estimer la volatilité. Il y a un phénomène de persistance qui est pris en compte par cette approche MIDAS (mixed data sampling).

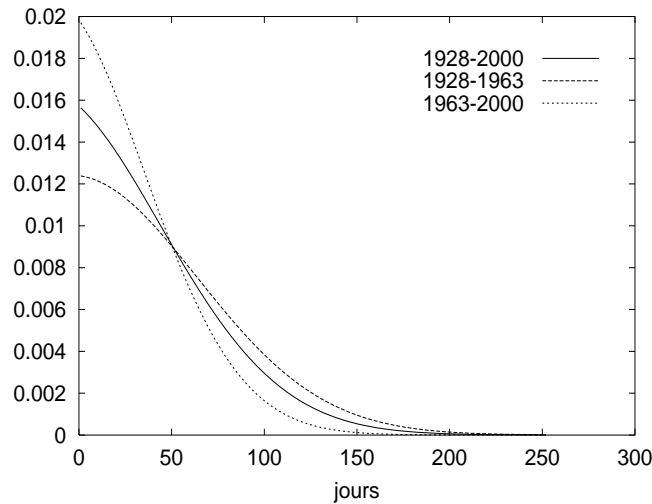


Figure 1: Poids des retards dans l'estimation de la volatilité

## 6 Prévion de la volatilité

On va exploiter le papier de Ghysel, Santa-Clara, and Valkanov (2006) sur la prédiction de la volatilité en utilisant des données avec des fréquences différentes.

Les notations du papier de 2006 sont légèrement différentes.

$$r_{t,t-1} = \log(P_t) - \log(P_{t-1}) \quad \text{daily returns}$$

On veut prédire une mesure de volatilité à horizon  $H$

$$V_{t+H,t}$$

Horizon journalier  $H = 1$ , weekly  $H = 5$ , monthly  $H = 20$ . On veut prévoir la variation quadratique des rendements du processus. Cette quantité n'est pas observable et doit être estimée avec un certain biais de discrétisation. On va utiliser des données intra journalières observées à fréquence  $m$ . On a donc:

1.  $t$  : indice temporel journaliser
2.  $m$  : pas de discrétisation
3.  $t - j/m$  : balayage des observations intra-journalières en nombre  $m$ .  $m = 80$  pour des données échantillonnées toutes les 5 minutes dans une journée d'ouverture des marchés.

La variation quadratique des rendements correspond à la somme des carrés des rendements futurs

$$Q_{t+H,t}^{Hm} = \sum_{j=1}^{Hm} [r_{(t+H)-(j-1)/m, (t+H)-(j-2)/m}]^2$$

Bien regarder la définition de  $r_{t,t-1}$ . Cette mesure de la volatilité est largement acceptée dans la littérature. Idée de Merton (1980) que l'on peut avoir une meilleure estimation de la volatilité au fur et à mesure que l'on réduit le pas d'échantillonnage.

**Merton, R.C., 1980. On estimating the expected return on the market: an exploratory investigation. *Journal of Financial Economics* 8, 323-361.**

## 6.1 Régressions MIDAS

En utilisant des données journalières on aura la modélisation générale suivante

$$V_{t+H,t}^{Hm} = \mu_H + \phi_H \sum_{k=0}^{k^{max}} b_H(k, \theta) X_{t-k, t-k-1} + \epsilon_{Ht}$$

On va mesurer la volatilité  $V_{t+H,t}^{Hm}$  par la variation quadratique des rendements ou le log de cette variation quadratique.

1.  $V_{t+H,t}^{Hm}$  et  $X_{t-k, t-k-1}$  ne sont pas échantillonnés nécessairement avec la même fréquence. On va pouvoir mesurer la volatilité
2. Les polynomes de retard sont parcimonieux
3. le modèle n'est pas nécessairement autorégressif

Une forme particulière des polynomes de retard avec

$$b_H(k, \theta) = \frac{f(k/k_m, \theta_1, \theta_2)}{\sum_{j=1}^{k_m} f(j/k_m, \theta_1, \theta_2)}$$

en utilisant la fonction Beta  $f(z, a, b) \propto z^{a-1}(1-z)^{b-1}$ .

1. Les variables endogènes et exogènes peuvent être échantillonnées à des fréquences différentes. On va donc pouvoir examiner quels sont les fréquences les plus performantes pour prédire la volatilité et faire des comparaisons. Plus particulièrement, on va pouvoir mesurer la volatilité journalière, hebdomadaire, mensuelle. Alors que les  $X$  vont être échantillonnés à une fréquence journalière ou intra journalière. Pour estimer la volatilité mensuelle, on va utiliser par exemple des données journalières.

2. Les polynomes de retard  $b_H$  sont fonction de  $\theta$ , donc sont très parcimonieux. Les poids somment à un. Ils sont tous positifs. On peut augmenter le nombre de retards sans augmenter le nombre de paramètres.
3. Le modèle n'est pas nécessairement autorégressif. C'est un modèle de prévision, plus qu'un modèle de volatilité conditionnelle comme les GARCH. On va pouvoir utiliser des spécifications différentes pour les  $X$  comme les rendements absolus au lieu des rendements au carré.

Modèle autoregressif usuel:

$$Q_{t+H,t}^{Hm} = \mu_H + \phi_H \sum_{k=0}^{k^{max}} b_H(k, \theta) Q_{t-k,t-k-1} + \epsilon_{Ht}$$

On se rapproche des GARCH avec

$$Q_{t+H,t}^{Hm} = \mu_H + \phi_H \sum_{k=0}^{k^{max}} b_H(k, \theta) (r_{t-k,t-k-1})^2 + \epsilon_{Ht}$$

Les regressions MIDAS principales vont être basées sur des mesures alternatives des rendements. On va avoir les valeurs absolues des rendements

$$Q_{t+H,t}^{Hm} = \mu_H + \phi_H \sum_{k=0}^{k^{max}} b_H(k, \theta) |r_{t-k,t-k-1}| + \epsilon_{Ht}$$

ou bien les écarts maximums sur un jour

$$Q_{t+H,t}^{Hm} = \mu_H + \phi_H \sum_{k=0}^{k^{max}} b_H(k, \theta) [high - low]_{t-k,t-k-1} + \epsilon_{Ht}$$

La dernière possibilité consiste à utiliser les **realised power** de Barndorff-Nielsen and Shephard

$$P_{t+1,t}^m = \sum_{j=1}^m |r_{t-(j-1)/m,t-(j-2)/m}|.$$

Donc c'est la même chose que  $Q$  mais on a pris les valeurs absolues.

## 6.2 Modèles concurrents

On va simplement relacher la contrainte sur les paramètres de retards pour adopter

$$Q_{t+H,t}^{Hm} = \mu_H + \phi_H \sum_{k=0}^{k^{max}} b_H(k) Q_{t-k,t-k-1} + \epsilon_{Ht}$$

Quand  $H = 1$ , on a les modèles de realised volatility de Andersen et al (2003).

Andersen, T., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. *Econometrica* 71, 529-626.

These authors consider long memory models to save parameters

$$\left(1 - \sum_{k=1}^5 b(k)L^k\right)(1-L)^d \log Q_{t+1,t}^m = \mu + \epsilon_t$$

This will be the ABDL benchmark to test the MIDAS specification in its forecasting abilities.

On va essayer d'avoir un modèle qui ne soit pas autorégressif et qui ait moins de paramètres. C'est l'argument de MIDAS.

### 6.3 Outils de comparaison

On va utiliser le mean squared error et comparer le MSE de MIDAS avec celui de ABDL.

$$MSE = \frac{1}{n} \sum (y - y^*)^2$$

On va prendre le rapport et si ce rapport

$$MSE_{MIDAS}/MSE_{ABDL}$$

est plus petit que 1, alors, la performance en prévision de MIDAS sera meilleure. Calculs sur l'échantillon et calculs en prévision hors échantillon.

1. Out of sample prediction: on prédit le log de la volatilité future. On prend l'exponentielle et ensuite on calcule le MSE.
2. In sample prediction: on parle alors d'ajustement du modèle. Si MIDAS prédit en logs, on calcule le MSE sur les log et on compare à ABDL. Si MIDAS prédit en niveau, on prend l'exponentielle de ABDL et l'on calcule le MSE.
3. ABDL prédit des volatilités journalières. Ensuite elles sont agrégées. MIDAS peut se spécifier directement à l'horizon de prédiction spécifié. On a donc une préduction directe.

### 6.4 Résultats avec données journalières

Les données utilisées sont le Dow Jones. Intra day returns every 5 minutes: 1993-2003. On a 2669 trading days et 80 observations par jour, ce qui fait 213,520 observations.

1. Si la mesure de la volatilité était le but de l'article, il faudrait filter pour les bruits de microstructure. Ici on ne s'intéresse qu'à la précision de la prévision. Donc pas de filtre pour l'autocorrelation due au non synchronous trading par exemple.

- De même, les données ne concernent que les heures d'ouverture 9:30 - 4:05. On néglige donc la volatilité des WE et de la nuit, ou ce que l'on peut observer avant l'ouverture officielle des marchés.

On va calculer

- daily realised volatility  $Q_{t,t-1}^m = \sum_{j=1}^m (r_{t-(j-1)/m,t-(j-2)/m})^2$
- daily squared returns  $(r_{t,t-1})^2$
- daily absolute returns :  $|r_{t,t-1}|$
- daily ranges :  $[high - low]_{t,t-1}$
- realised power :  $P_{t,t-1}^m = \sum_{j=1}^m |r_{t-(j-1)/m,t-(j-2)/m}|$ , with  $m = 80$

On prendra un lag max de 50 et  $H = 1$  (jour), 5 (semaine), 10, 15, 20. Horizons pour les value at risk et option pricing.

On va d'abord comparer les résultats entre MIDAS et ABDL pour la prévision de la log volatility et de la volatility in sample.

- Les poids: Avec ABDL, les poids sont irréguliers pour les 5 premiers et ensuite décroissent très lentement. Ils sont toujours positifs après 50.
- Les poids : avec MIDAS, les poids décroissent de manière très régulière avec une prédominance des retards faibles. Au delà de 30, ils sont quasiment nuls.
- La structure des poids varie beaucoup en fonction de l'horizon de prévision. Courte pour les horizons courts, plus longue pour les horizons longs.
- Les résultats de MIDAS sont meilleurs au fur et à mesure que l'horizon de prévision augmente.
- $(r_{t,t-1})^2$  n'est pas un bon prédicteur, surtout pour les logs. Noisy measured of past realised volatility.
- Les meilleurs prédicteurs sont  $Q_{t,t-1}^m$  et surtout  $P_{t,t-1}^m$ .
- Parmi les prédicteurs directs,  $[high - low]_{t,t-1}$  est celui qui marche le mieux.
- MIDAS est meilleur pour prévoir directement la volatility que la log volatility.

Passons maintenant aux prévisions hors échantillon (out of sample)

- Realised power est le meilleur prédicteur
- MIDAS continue à bien prévoir par rapport à ABDL. Les gains sont en moyenne de 20 à 30%.
- Log ou prédiction directe donnent des résultats similaires quand on utilise le Realised power.

Les résultats sont données en utilisant toujours la même fréquence d'observation. Donc pour le moment pas de mix frequency.

## 6.5 MIDAS et les high frequencies

L'usage de données haute fréquence permet-il une meilleure prévision. On a vu que les meilleurs prédictors étaient des prédictors composites, c'est à dire agrégeant de manière arbitraire des données haute fréquence. Pourquoi ne pas estimer les poids d'aggrégation en utilisant les possibilités de MIDAS?

On définit un opérateur de retard

$$L^{1/m}r_{t,t-1/m} = r_{t-1/m,t-2/m}$$

et un polynome de retard  $B(L^{1/m})$ . On va alors définir des régressions alternatives à celles considérées jusqu'à présent en se basant sur les squared returns et les absolute returns. C'étaient les régresseurs les plus mauvais, mais ces termes servaient de base pour calculer  $Q$  et  $P$ .

$$Q_{t+H,t}^{Hm} = \mu + \phi B(L^{1/m})[r_{t,t-1/m}]^2 + \epsilon_{Ht}$$

et

$$Q_{t+H,t}^{Hm} = \mu + \phi B(L^{1/m})|r_{t,t-1/m}| + \epsilon_{Ht}$$

Ce sont des généralisations des régressions précédentes qui avaient comme régresseur  $Q$  et  $P$ . Ici on va estimer les poids.

On a donc une variable endogène qui est un index composite calculée pour donner des observations journalières et des prédictors qui maintenant sont sur une base intra day de 5 minutes. Ces dernières ont une forte composante saisonnière (acalmie à midi par exemple) dont il faut tenir compte. Deux manières d'ajuster:

1. On peut enlever la moyenne horaire empirique à l'observation qui appartient à cette heure. C'est équivalent à désaisonnaliser les données.
2. Ajouter à la régression des variables dummy qui correspondent aux heures. On explique la saisonnalité dans le modèle en ajoutant des paramètres.

Les résultats empiriques sont les suivants

1. Il vaut mieux introduire des variables dummy. Utiliser des données filtrées donnent de moins bons résultats.
2. On ne gagne pas beaucoup ou rien du tout à considérer des données de haute fréquence par rapport à la solution précédent où l'on avait calculé de manière arbitraire  $Q$  et  $P$ .

## 7 Conclusion

On a vu dans ce chapitre comment on pouvait estimer la volatilité et que pour cela des données de haute fréquence étaient fort utiles. Le premier modèle considéré examinait la relation rendement risque. On ne mesure pas les rendements et le risque sur la même fréquence.

Le deuxième modèle examine la prédiction de la volatilité et montre comment agréger des données de haute fréquence. Il faut utiliser ce type de données pour pouvoir prévoir la volatilité. Des données journalières ne sont pas suffisantes.

On a employé des modèles économétriques assez simples qui des modèles à retard échelonnés. Il s'estiment par maximum de vraisemblance ou moindres carrés non linéaires. Les modèles à mémoire longue sont aussi utiles, mais beaucoup plus compliqué d'usage et ne donnent pas de meilleurs résultats en prévision.

## 8 Exercice

Considérez la relation usuelle du CAPM qui relie les rendements  $R_{t+1}$  au risque  $V_t$ :

$$R_{t+1} = \mu + \gamma V_t + u_t :$$

- Quelle est l'interprétation du coefficient  $\gamma$ ?
- Comment peut-on estimer la volatilité  $V_t$ ?
- Quel intérêt a-t-on à utiliser des données de fréquence d'observation différentes pour estimer cette relation?
- Comment formulerez vous le modèle à retards échelonnés correspondant?

## References

- ALMON, S. (1965): "The Distributed Lag between Capital Appropriations and Expenditures," *Econometrica*, 33, 178–196.
- GHYSEL, E., P. SANTA-CLARA, AND R. VALKANOV (2005): "There is a risk-return trade-off after all," *Journal of Financial Economics*, 76, 509–548.
- (2006): "Predicting volatility: getting the most out of return data sampled at different frequencies," *Journal of Econometrics*, 131, 59–95.
- GRILICHES, Z. (1967): "Distributed lags: a Survey," *Econometrica*, 35, 16–49.
- HARVEY, A. C. (1981): *The Econometric Analysis of Time Series*. Philip Allan, Oxford.
- MERTON, R. (1973): "An intertemporal capital asset pricing model," *Econometrica*, 41(5), 867–887.
- VERBEEK, M. (2000): *A Guide to Modern Econometrics*. John Wiley and Sons, New York.