

# Chapitre 4

## La Prédiction de la Volatilité

(2020-2021)

Econométrie Financière Avancée

Olivier DARNÉ

## Prévision de la volatilité

- ▶ La modélisation de la volatilité des rentabilités est une clé importante pour la gestion des risques, la tarification des produits dérivés, la couverture du risque (VaR), les participants de marché, la sélection de portefeuille, les décisions, de politique monétaire, et bien d'autres activités financières.
- ▶ Il est possible de sélectionner le modèle optimal à partir des prévisions de la volatilité.
- ▶ Il existe des indicateurs statistiques qui permettent de comparer la volatilité prévue avec la **volatilité observée**.
- ▶ Meilleures sont les prévisions et meilleur est supposé le modèle.

Nous cherchons à comparer la capacité des modèles présentés dans le chapitre précédent à prévoir (*forecast*) la variance conditionnelle  $\hat{\sigma}_{T+h}^2$  pour un horizon de prévision donné  $h > 0$ .

Dans le cas d'un modèle GARCH(1,1) l'équation de prévision générale pour un horizon  $h$  est donnée par :

$$\hat{\sigma}_{T+h|T}^2 = \hat{\omega} + \hat{\alpha}\varepsilon_{T+h-1|T}^2 + \hat{\beta}\hat{\sigma}_{T+h-1|T}^2$$

Néanmoins, il faut prendre en compte le "pas"  $i$  (*step*) de la prévision car la prévision de la volatilité n'est pas calculée de la même manière lorsque  $i = 1$  ou  $i > 1$ .

En effet, la **prévision à un pas** (*one-step ahead forecast*), i.e.  $i = 1$ , pour  $T + 1$  est donnée par

$$\begin{aligned}
 E_T[\sigma_{T+1}^2] &= \hat{\sigma}_{T+1}^2 \\
 &= E_T[\omega + \alpha \varepsilon_T^2 + \beta \sigma_T^2] \\
 &= E_T[\omega] + E_T[\alpha \varepsilon_T^2] + E_T[\beta \sigma_T^2] \\
 &= \hat{\omega} + \hat{\alpha} E_T[\varepsilon_T^2] + \hat{\beta} E_T[\sigma_T^2] \\
 &= \hat{\omega} + \hat{\alpha} E_T[\varepsilon_T^2] + \hat{\beta} E_T[\sigma_T^2] \\
 &= \hat{\omega} + \hat{\alpha} \hat{\varepsilon}_T^2 + \hat{\beta} \hat{\sigma}_T^2
 \end{aligned}$$

- $\hat{\varepsilon}_T^2$  : valeurs ajustées pour l'erreur au carré pour l'échantillon  $t = 1, \dots, T$
- $\hat{\sigma}_T^2$  : valeurs ajustées pour la variance conditionnelle pour  $t = 1, \dots, T$
- $\hat{\omega}, \hat{\alpha}$  et  $\hat{\beta}$  : coefficients estimés sur  $t = 1, \dots, T$
- $T$  : dernière observation de l'échantillon utilisée pour l'estimation (*in-sample*) du modèle

De la même manière on obtient la **prévision à un pas** pour les modèles GARCH asymétriques

- GJR-GARCH

$$\hat{\sigma}_{T+1}^2 = \hat{\omega} + \left[ \hat{\alpha} + \hat{\gamma} \hat{l}(\varepsilon_T) \right] \hat{\varepsilon}_T^2 + \hat{\beta} \hat{\sigma}_T^2$$

- TGARCH model

$$\hat{\sigma}_{T+1} = \hat{\omega} + \left[ \hat{\alpha} + \hat{\gamma} \hat{l}(\varepsilon_T) \right] \hat{\varepsilon}_T + \hat{\beta} \hat{\sigma}_T$$

- EGARCH

$$\hat{\sigma}_{T+1}^2 = \exp(\hat{\omega}) \exp[\hat{g}(z_T)] \hat{\sigma}_T^{2\hat{\beta}}$$

La **prévision à deux pas** (*two-step ahead forecast*), i.e.  $i = 2$ , pour  $T + 2$  est donnée par

$$\begin{aligned} E_T[\sigma_{T+2}^2] &= \hat{\sigma}_{T+2}^2 \\ &= E_T[\omega + \alpha \varepsilon_{T+1}^2 + \beta \sigma_{T+1}^2] \\ &= \hat{\omega} + \hat{\alpha} E_T[\varepsilon_{T+1}^2] + \hat{\beta} E_T[\sigma_{T+1}^2] \end{aligned}$$

Puisque la valeur prévue de  $\hat{\varepsilon}_{T+1}^2$  est inconnue nous devons utiliser l'**espérance** de  $E_T(\varepsilon_{T+1}^2) = \sigma_{T+1}^2$ .

Il faut noter que comme  $\varepsilon_t$  est une erreur alors son espérance conditionnelle est nulle,  $E(\varepsilon_T) = 0$ , et sa variance conditionnelle :  $E_T(\varepsilon_T^2) = \sigma_T^2$

$$\begin{aligned} \hat{\sigma}_{T+2}^2 &= \hat{\omega} + \hat{\alpha} \hat{\sigma}_{T+1}^2 + \hat{\beta} \hat{\sigma}_{T+1}^2 \\ &= \hat{\omega} + (\hat{\alpha} + \hat{\beta}) \hat{\sigma}_{T+1}^2 \end{aligned}$$

De manière similaire, les prévisions avec **un pas supérieur à un** ( $i > 1$ ) peuvent être obtenues, avec  $i = 2, \dots, k$ , par

$$\hat{\sigma}_{T+i}^2 = \hat{\omega} + (\hat{\alpha} + \hat{\beta})\hat{\sigma}_{T+i-1}^2$$

On obtient de la même manière pour les autres modèles :

- GJR-GARCH

$$\hat{\sigma}_{T+i}^2 = \hat{\omega} + \left( \hat{\alpha} + \frac{1}{2}\hat{\gamma} + \hat{\beta} \right) \hat{\sigma}_{T+i-1}^2$$

- TGARCH model

$$\hat{\sigma}_{T+i} = \hat{\omega} + \left( \hat{\alpha} + \frac{1}{2}\hat{\gamma} + \hat{\beta} \right) \hat{\sigma}_{T+i-1}$$

- EGARCH

$$\hat{\sigma}_{T+i}^2 = \hat{C} \exp \left( \hat{\omega} - \hat{\theta}_2 \sqrt{2/\pi} \right) \hat{\sigma}_{T+i-1}^{2\hat{\beta}}$$

Pour le **modèle GJR-GARCH** l'espérance de la variable dichotomique est donnée par  $E[I(\varepsilon_{t-1}) < 0] = P[\varepsilon_{t-1} < 0] = \frac{1}{2}$  si les innovations standardisées suivent une loi Normale.

Pour le **modèle EGARCH**,  $\hat{C}$  est une constante dépendant des paramètres  $\hat{\theta}_1$  and  $\hat{\theta}_2$ .

La **prévision à deux pas** peut se réécrire en substituant  $\hat{\sigma}_{T+1}^2$

$$\begin{aligned}\hat{\sigma}_{T+2}^2 &= \hat{\omega} + (\hat{\alpha} + \hat{\beta})\hat{\sigma}_{T+1}^2 \\ &= \hat{\omega} + (\hat{\alpha} + \hat{\beta}) \left[ \hat{\omega} + (\hat{\alpha} + \hat{\beta})\sigma_T^2 \right] \\ &= \hat{\omega} + \hat{\omega}(\hat{\alpha} + \hat{\beta}) + (\hat{\alpha} + \hat{\beta})^2\sigma_T^2\end{aligned}$$

De manière itérative on en déduit que les prévisions avec **un pas supérieur à un** ( $i > 1$ ) peuvent être obtenues, avec  $i = 2, \dots, k$ , par

$$\begin{aligned}E_T[\sigma_{T+i}^2] &= \hat{\sigma}_{T+i}^2 \\ &= \sum_{j=0}^{i-1} \left[ (\hat{\alpha} + \hat{\beta})^j \hat{\omega} \right] + (\hat{\alpha} + \hat{\beta})^{i-1} \hat{\sigma}_T^2 \\ &= \hat{\omega} \sum_{j=0}^{i-1} \left[ (\hat{\alpha} + \hat{\beta})^j \right] + (\hat{\alpha} + \hat{\beta})^{i-1} \hat{\sigma}_T^2\end{aligned}$$



Cette relation montre que les **prévisions de la variance** sont **additives** sur le temps.

$$E_T[\sigma_{T+i}^2] = \hat{\sigma}_{T+i}^2 = \hat{\omega} \sum_{j=0}^{i-1} [(\hat{\alpha} + \hat{\beta})^j] + (\hat{\alpha} + \hat{\beta})^{i-1} \hat{\sigma}_T^2$$

Cette propriété est intéressante car si on produit des prévisions quotidiennes à un, deux, trois, quatre et cinq pas, cad des prévisions de chaque jour de la semaine alors la prévision de la variance sur l'ensemble de la semaine sera la **somme** des cinq prévisions journalières

Il est possible de montrer pour  $i > 1$  que (voir Annexe) :

$$\hat{\sigma}_{T+i}^2 = \bar{\sigma}^2 + (\hat{\alpha} + \hat{\beta})^{i-1} [\hat{\sigma}_T^2 - \bar{\sigma}^2]$$

- $\bar{\sigma}^2 = \hat{\omega} / (1 - \hat{\alpha} - \hat{\beta})$  : la variance non conditionnelle de  $\varepsilon_t$
- Comme dans le modèle GARCH(1,1)  $\alpha + \beta < 1$  lorsque  $i \rightarrow \infty$  alors  $\hat{\sigma}_{T+i}^2 \rightarrow \bar{\sigma}^2$

Les **prévisions à plusieurs pas** convergent alors vers la variance non conditionnelle quand l'horizon tend vers l'infini

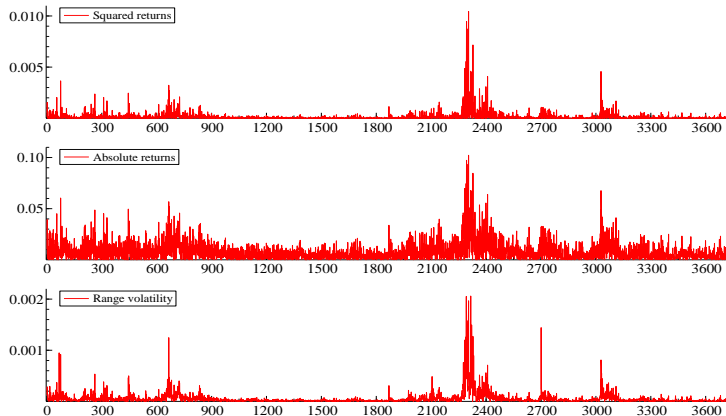
## Approximation de la volatilité

Pour évaluer la prévision de la volatilité (variance) nous devons utiliser un **proxy de la volatilité** car elle n'est pas observable et doit être approximée :

- les **rentabilités au carré** :  $r_t^2$
- les **rentabilités absolues** :  $|r_t|$
- **volatilité basée sur le rang** (*range volatility*) :

$$rng = \frac{1}{4 \ln 2} [\ln(p_{high,t}) - \ln(p_{low,t})]^2$$

- $p_{high,t}$  : le prix le plus haut du jours  $t$
- $p_{low,t}$  : le prix le plus bas du jours  $t$

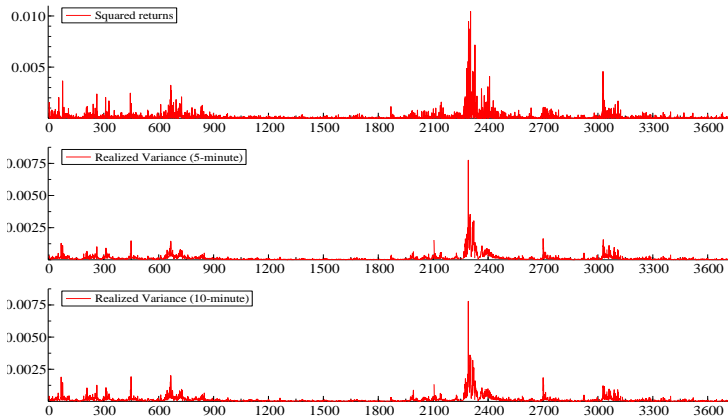


► L'utilisation des **données intra-journalières** est reconnue comme étant la "meilleure" approximation de la volatilité (quand elles sont disponibles).

► Cette volatilité est appelée la **volatilité réalisée** (*realized volatility*) :

$$RV = \sum_{i=2}^{n(\delta)} [\ln(p_{i,t}) - \ln(p_{i,t-1})]^2$$

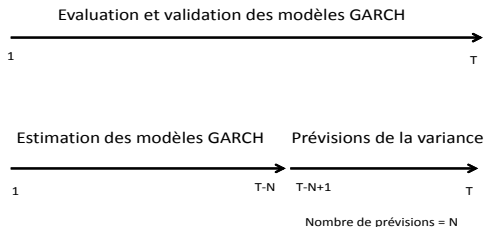
- $p_{i,t}$  : le  $i$ ème prix du jour  $t$  à la fréquence  $\delta$ , tel que  $i = 1, \dots, n(\delta)$
- $n(\delta) = n_{sec}/\delta$
- $n_{sec}$  : le nombre de secondes dans le jour ouvré



Soit  $r_t$  les rentabilités sur la période  $t = 1, \dots, T$ .

Pour évaluer la **performance de précision** des modèles en prévision il est nécessaire de comparer les prévisions à des valeurs observées (connues)

- 1 On enlève un nombre  $N$  de valeurs de l'échantillon ("que l'on met de côté !") ;
- 2 On estime le modèle sur le nouvel échantillon ( $T - N$ ) ;
- 3 On prévoit  $N$  valeurs à partir du modèle.



## Critères de comparaison de prévision

*All models are false, but some are useful (G. Box, 1979).*

On définit la variable d'**erreur de prévision** à l'horizon  $h$  définie comme l'écart entre les valeurs observées (proxy de la volatilité,  $\sigma_t^2$ ) et les valeurs prévues ( $\hat{\sigma}_t^2$ ) à la date  $t$  pour l'horizon  $h$  :

$$e_{t+h} = \sigma_{t+h}^2 - \hat{\sigma}_{t+h}^2 \quad \text{avec } h = 1, \dots, N$$



La caractéristique principale d'un « bon » prédicteur  $\hat{\sigma}_{t+h}^2$  est de **minimiser** cette erreur de prévision au sens d'une **fonction de perte**

La **perte associée avec la prévision**  $h$  est supposée être une fonction de l'erreur de prévision  $e_{t+h}$  et est notée

$$g(e_{t+h}) = L(\sigma_{t+h}^2, \hat{\sigma}_{t+h}^2)$$

Cette fonction  $g(\cdot)$  est une **fonction de perte** telle que

- $g(\cdot) = 0$  quand aucune erreur n'est faite
- $g(\cdot) \geq 0$
- elle est croissante en taille lorsque les erreurs deviennent plus importantes en magnitude

Les **mesures de précision** permettent d'évaluer et de comparer la **précision des prévisions** obtenues à partir de différents modèles.

Le meilleur modèle en prévision est celui qui **minimise** des mesures qui se différencie par rapport à la **fonction de perte** retenue

$$L(\sigma_{t+h}^2, \hat{\sigma}_{t+h}^2)$$

*Mean Squared Error (MSE) :*

$$MSE = \frac{1}{N} \sum_{h=1}^N (\sigma_{t+h}^2 - \hat{\sigma}_{t+h}^2)^2$$

*Root Mean Squared Error (RMSE) :*

$$RMSE = \sqrt{\frac{1}{N} \sum_{h=1}^N (\sigma_{t+h}^2 - \hat{\sigma}_{t+h}^2)^2}$$

*Mean Absolute Error (MAE) :*

$$MAE = \frac{1}{N} \sum_{h=1}^N (|\sigma_{t+h}^2 - \hat{\sigma}_{t+h}^2|)$$

*Gaussian Likelihood (QLIKE)* ou *Gaussian Likelihood Loss Function (GLLS)* :

$$QLIKE = \frac{1}{N} \sum_{h=1}^N \left[ \frac{\sigma_{t+h}^2}{\widehat{\sigma}_{t+h}^2} - \ln \left( \frac{\sigma_{t+h}^2}{\widehat{\sigma}_{t+h}^2} \right) - 1 \right]$$

*Heteroskedasticity-adjusted Mean Squared Error (HMSE)* :

$$HMSE = \frac{1}{N} \sum_{h=1}^N \left( \frac{\sigma_{t+h}^2}{\widehat{\sigma}_{t+h}^2} - 1 \right)^2$$

*Heteroskedasticity-adjusted Mean Absolute Error (HMAE)* :

$$HMAE = \frac{1}{N} \sum_{h=1}^N \left| \frac{\sigma_{t+h}^2}{\widehat{\sigma}_{t+h}^2} - 1 \right|$$

## Comparaison $R^2_{OOS}$

Il est possible de comparer les prévisions de différents modèles par rapport à un **modèle de référence** (benchmark) avec le  $R^2$  *out-of-sample* :

$$R^2_{OOS} = 1 - \frac{MSE_i}{MSE_0}$$

avec  $MSE_i$  le MSE du modèle  $i$ , et  $MSE_0$  le MSE du benchmark

Si  $R^2_{OOS} > 0 \Rightarrow$  modèle  $i >$  modèle de référence

## L'approche par la régression de Mincer et Zarnowitz

Mincer et Zarnowitz (1969) introduisent une [approche par la régression](#) comme une alternative à l'utilisation des fonctions de perte pour évaluer les prévisions de la volatilité.

Pour une prévision à un pas  $\hat{\sigma}_{t+1}^2$  et une fonction de perte  $L(\sigma_{t+1}^2, \hat{\sigma}_{t+1}^2)$ , il est attendu que l'espérance de l'erreur de prévision conditionnelle par rapport à l'information disponible la plus récente soit nulle.

$$E \left[ \frac{\partial L(\sigma_{t+1}^2, \hat{\sigma}_{t+1}^2)}{\partial \hat{\sigma}_{t+1}^2} \middle| I_t \right] = 0$$

Mincer et Zarnowitz (MZ, 1969) proposent une **régression simple** de l'espérance conditionnelle :

$$\frac{\partial L(\sigma_{t+1}^2, \hat{\sigma}_{t+1}^2)}{\partial \hat{\sigma}_{t+1}^2} = \beta_0 + \beta_1 \hat{\sigma}_{t+1}^2 + \varepsilon_{t+1}$$

avec  $\varepsilon_{t+1} \sim N(0, \sigma^2)$

En considérant le cas d'une **fonction de perte quadratique**

$$L(\sigma_{t+1}^2, \hat{\sigma}_{t+1}^2) = (\sigma_{t+h}^2 - \hat{\sigma}_{t+h}^2)^2$$

on obtient :

$$\frac{\partial L(\sigma_{t+1}^2, \hat{\sigma}_{t+1}^2)}{\partial \hat{\sigma}_{t+1}^2} = -2(\sigma_{t+h}^2 - \hat{\sigma}_{t+h}^2)$$

La **régression MZ** devient alors

$$\begin{aligned}(\sigma_{t+1}^2 - \hat{\sigma}_{t+1}^2) &= \beta_0 + \beta_1 \hat{\sigma}_{t+1}^2 + \varepsilon_{t+1} \\ \Leftrightarrow \sigma_{t+1}^2 &= \beta_0 + (\beta_1 + 1) \hat{\sigma}_{t+1}^2 + \varepsilon_{t+1}\end{aligned}$$

L'**hypothèse nulle** testée est alors la suivante :

- $H_0 : \beta_0 = (\beta_1 - 1) = 0$
- $H_1 : \beta_0 \neq 0$  ou  $\beta_1 \neq 1$

Dans ce cas la régression doit être estimée en prenant en compte un **estimateur cohérent avec l'hétéroscédasticité** (*Heteroscedastic Consistent*, HC) comme celui de White (1980).



Pour les prévisions à **plusieurs pas**, la régression MZ est donnée par

$$\sigma_{t+h}^2 = \beta_0 + (\beta_1 + 1)\hat{\sigma}_{t+h-1}^2 + \varepsilon_{t+h-1}$$

Dans ce cas la régression doit être estimée en prenant en compte un **estimateur cohérent avec l'hétéroscédasticité et l'autocorrélation** (*Heteroscedastic and Autocorrelated Consistent*, HAC) comme celui de Newey et West (1987).

Des modifications de la régression MZ ont été proposées, notamment pour diminuer sa sensibilité aux valeurs extrêmes

- Pagan et Schwert (1990) avec un **transformation logarithmique**

$$\ln(\sigma_{t+h}^2) = \beta_0 + (\beta_1 + 1)\ln(\hat{\sigma}_{t+h-1}^2) + \varepsilon_{t+h}$$

- Jorion (1995) en prenant plutôt l'**écart-type**

$$\sigma_{t+h} = \beta_0 + (\beta_1 + 1)\hat{\sigma}_{t+h-1} + \varepsilon_{t+h}$$

## Tests de précision des prévisions

Comment savoir lorsque  $MSE_1 > MSE_2$  si les prévisions du modèle 1 sont ou pas **statistiquement** différentes de celles du modèle 2 ?

Pour obtenir de telles conclusions, des **tests statistiques** ont été développés, fournissant des informations plus fiables.

Un certain nombre de tests statistiques ont été développés comme

- (a) les tests fondés sur la **capacité prédictive égale** (EPA, *equal predictive ability*) qui comparent deux par deux les performances prédictives (Diebold et Mariano, 1995 ; Clark et McCracken, 2001, 2007 ; Giacomini et White, 2006)
- (b) les tests fondés sur la **capacité prédictive supérieure** (SPA, *superior predictive ability*) qui permettent une comparaison multiple entre les modèles par rapport à un modèle de référence (White, 2000 ; Hansen, 2005)

## (a) Les tests de capacité prédictive égale (EPA)

### Le test de Diebold et Mariano (1995)

Diebold et Mariano (1995) ont proposé un test (DM) basé sur l'**hypothèse d'égalité prédictive** :

$$MSE_1 = MSE_2 \quad \Leftrightarrow \quad g(e_{1,t}) = g(e_{2,t}) \quad \Leftrightarrow \quad L_{1,t} = L_{2,t}$$

Le **test DM** est basé sur le **différentiel de pertes** d'erreurs de prévision (ponctuelle) :

$$d_{12,t} = g(e_{1,t}) - g(e_{2,t}) = L_{1,t} - L_{2,t}$$

- $g(\cdot)$  : une fonction de perte
- $e_{1,t}$  et  $e_{2,t}$  : erreurs de prévision

L'**hypothèse nulle de capacité prédictive égale** (*equal predictive ability*, EPA) correspond à

$$H_0 : \mu_1 = E[d_{12,t}] = E[g(e_{1,t})] - E[g(e_{2,t})] = 0$$

$$H_1 : \mu_1 = E[d_{12,t}] \neq 0$$

Le **test statistique DM** est

$$DM_{12} = \frac{\bar{d}_{12}}{\widehat{\sigma}_{\bar{d}_{12}}} \rightarrow N(0, 1)$$

avec

- $\bar{d}_{12} = T^{-1} \sum_{t=1}^T d_{12,t}$  : moyenne empirique des différentiels de pertes
- $\widehat{\sigma}_{\bar{d}_{12}}$  : estimation de l'écart-type de long-terme de  $\bar{d}_{12}$
- sous l'hypothèse que le différentiel de pertes soit stationnaire en covariance
- le test n'est applicable que pour des **modèles non emboîtés**
- McCracken (2007), Clark and McCracken (2001) et Clark et West (2007) proposent les tests des **modèles emboîtés** (*nested models*)

Puisque les erreurs de prévision, et donc les différentiels de pertes, peuvent être sérielement corrélées, DM ont proposé une **version robuste** de leur test :

$$DM_{12} = \frac{\bar{d}_{12}}{\hat{\sigma}_{\bar{d}}} \rightarrow N(0,1)$$

avec  $\bar{d} = \sqrt{\hat{g}(0)T^{-1}}$ , où  $\hat{g}(0)$  est un estimateur du spectre à la fréquence zéro du différentiel de pertes.

L'**hypothèse nulle EPA sera rejetée** au seuil  $\alpha$  si

$$|DM_{12}| > z_{\alpha/2}$$

avec  $z_{\alpha/2}$  la valeur d'une loi Normale centrée réduite.

Le test est aussi possible en unilatéral avec  $H_1 : \mu_1 > 0$  ou  $H_1 : \mu_1 < 0$

## Le test de Harvey, Leybourne et Newbold (1997)

Harvey et alii (1997) (HLN) montrent que le test DM est **biaisé en petits échantillons** et proposent d'améliorer ses propriétés en petits échantillons en

- rajoutant une correction du biais à la statistique de test DM
- comparant la statistique corrigée avec une loi de Student à  $(T - 1)$  ddl plutôt qu'une loi Normale

La **statistique corrigée** est définie par

$$HLN - DM = mDM = \sqrt{\frac{T + 1 - 2h + h(h - 1)}{T}} DM$$

## Le test de Mariano et Preve (2012)

Mariano et Preve (2012) proposent une version **multivariée** du test de Diebold-Mariano (MDM)

Ils testent l'hypothèse nulle entre  $(k + 1)$  modèles alternatifs :

$$H_0 : \mu_j = E[d_{j,t}] = E[g(e_{j,t})] - E[g(e_{j+1,t})] = 0 \quad \text{pour tout } j = 1, \dots, k$$

$$H_1 : \mu_j = E[d_{j,t}] \neq 0$$

## Le test de Clark et West (2007)

Clark et West (2007) proposent un test pour les **modèles emboîtés** (*nested model*), à savoir qu'un modèle A est emboîté dans un modèle B si les paramètres du modèle A sont un sous-ensemble des paramètres du modèle B.

Ils proposent une statistique ajustée pour la différence entre les MSE des 2 modèles sous l'hypothèse nulle

$$d_{12,t}^* = g(e_{1,t}) - g(e_{2,t}) + (\hat{y}_{1,t} - \hat{y}_{2,t})^2$$

La statistique *CW* prend alors la forme

$$CW_{1,2} = \sqrt{N} \frac{\bar{d}_{12}^*}{\hat{\sigma}_{\bar{d}_{12}^*}}$$

- $\bar{d}_{12}^* = N^{-1} \sum_{t=1}^N d_{12,t}^*$  le différentiel de MSE ajusté
- $\hat{\sigma}_{\bar{d}_{12}^*}$  une estimation convergente de l'écart-type de  $\bar{d}_{12}^*$
- $\hat{y}_{1,t}$  et  $\hat{y}_{2,t}$  les prévisions des modèles 1 et 2



## (b) Les tests de capacité prédictive supérieure (SPA)

### Le test de White (2000)

White (2000) propose un test de comparaison de précision de prévision basé sur la SPA, noté RC (*Reality Check*).

Ce test RC peut être interprété comme une **version non standardisée** du test SPA de Hansen (2005)

$$RC T_{\max} = \max_{j \in \mathcal{M}} n^{1/2} \bar{d}_j$$

La **distribution du test**  $RC T_{\max}$  sous l'hypothèse nulle  $\mu_j \leq 0$  est une distribution Normale

$$n^{1/2}(\bar{d}_j - \mu) \xrightarrow{d} N(0, \Omega)$$

Le problème est que sous  $H_0$   $\mu$  n'est pas unique.

Le test de White suit la convention ... d'utiliser la **configuration la moins favorable** (LFC, *Least Favorable Configuration*) pour résoudre la problème :  $\mu = 0$

Sous la LFC on a alors

$$n^{1/2} \bar{d}_j \xrightarrow{A} N(n^{1/2} \mu, \Omega)$$

White (2000) propose deux manières pour calculer les valeurs critiques : soit par **simulation** soit par **bootstrap**

Le test RC de White pose des problèmes de test  $\Rightarrow$  **modification** du test par Hansen (2005)

## Le test de Hansen (2005)

Hansen (2005) propose un test de comparaison de précision de prévision basé sur la **capacité prédictive supérieure** (SPA, *superior predictive ability*).

Ce test permet une **comparaison multiple** entre les modèles par rapport à un **modèle de référence**.

Soit le **différentiel de pertes d'erreurs de prévision** :

$$d_{0j,t} = L_{0,t} - L_{j,t},$$

- $t = 1, \dots, n$  et  $j = 1, \dots, m$ , avec  $j \in \mathcal{M}_0$ ,
- $\mathcal{M}_0$  : un ensemble de modèles à comparer
- $d_{0j,t}$  : représente la performance relative du modèle  $j$  par rapport au modèle de référence (indiqué par 0),
- $L_{j,t}$  : une fonction de perte

L'**hypothèse nulle** est exprimée par rapport au meilleur modèle alternatif :

- $H_0 : \mu_j = E[d_{0j,t}] \leq 0 \quad \forall j \in \mathcal{M}_0$
- $H_1 : \mu_j = E[d_{0j,t}] > 0 \quad j = 1, \dots, m$

La **statistique** de test est alors donnée par :

$$SPA \ T_{\max} = \max_{j \in \mathcal{M}} \frac{n^{1/2} \bar{d}_j}{\hat{\omega}_j}$$

- $\bar{d}_j = n^{-1} \sum_{t=1}^n d_{0j,t}$
- $\hat{\omega}_j^2 = \widehat{var}(n^{1/2} \bar{d}_j)$  : un estimateur cohérent de la variance asymptotique de  $(n^{1/2} \bar{d}_j)$  obtenu par bootstrap

La **distribution du test** SPA  $T_{\max}$  sous l'hypothèse nulle est une distribution Normale

$$SPA\ T_{\max} \sim N(\hat{\mu}, \Omega)$$

où  $\hat{\mu}$  est un estimateur choisi pour  $\mu$ .

Comme différents estimateurs pour  $\hat{\mu}$  sont possibles, impliquant ainsi des  $p$ -values différentes du test, Hansen (2005) propose **trois estimateurs**

$$\hat{\mu}^l \leq \hat{\mu}^c \leq \hat{\mu}^u$$

Les  **$p$ -values du test** sont alors notées  $SPA_c$ , et les limites inférieure (*lower*) et supérieure (*upper*) respectivement par  $SPA_l$  et  $SPA_u$ .

Une  $p$ -value significative ( $SPA_c$ ) signifie que l'hypothèse nulle est rejetée, à savoir que le modèle de référence n'est pas amélioré.

## Différence entre les tests EPA et SPA

Les tests EPA et SPA portent sur des **différentiels de fonctions de pertes**

$$\begin{aligned}d_{0j,t} &\equiv L_{0,t} - L_{j,t} & j = 1, \dots, m \\ \mu_j &\equiv E[d_{0j,t}] & t = 1, \dots, n\end{aligned}$$

Si  $\mu_j > 0$  (ou  $\mu_j < 0$ ) alors la  $j$ ième prévision est **meilleure** (plus mauvaise) que le **benchmark**

La différence entre les tests EPA et SPA est sur l'**hypothèse nulle**

- Test EPA :  $H_0 : \mu_1 = 0 \quad (m = 1)$
- Test SPA :  $H_0 : \mu_j \leq 0 \quad j = 1, \dots, m$

## Le test de Hansen, Lunde et Nason (2011)

Un inconvénient avec les tests précédents, notamment les tests SPA, est que l'on doit **spécifier un modèle de référence** pour faire la comparaison.

Hansen et alii (2011) ont alors proposé la procédure de l'**ensemble de confiance de modèles** (MCS, *model confidence set*) qui ne requiert pas de spécifier (ou de choisir) au préalable un modèle de référence.

Cette procédure permet de déterminer un **sous-ensemble de modèles**, noté  $\mathcal{M}^*$ , (équivalent en termes de capacité prédictive), qui sont supérieures en termes de SPA aux autres modèles concurrents à partir d'un **ensemble de modèles à comparer**, noté  $\mathcal{M}_0$ .

La procédure MCS donne un **ensemble de confiance de modèles**, noté  $\widehat{\mathcal{M}}^*$ , contenant **les meilleurs modèles** avec un niveau de confiance donné  $\alpha$ .

Cet ensemble MCS permet d'obtenir **plusieurs modèles** affichant des **performances de prévision équivalentes** et donne donc de la robustesse à l'exercice de prévision plutôt que de fonder l'analyse de prévision sur un seul modèle.



La **statistique de test** est définie par :

$$T_{\max \mathcal{M}} = \max_{i \in \mathcal{M}} t_i \quad \text{avec} \quad t_i = \frac{\bar{d}_i}{\sqrt{\widehat{\text{var}}(\bar{d}_i)}}$$

- $\widehat{\text{var}}(\bar{d}_i)$  représente l'estimation de  $\text{var}(\bar{d}_i)$
- $\bar{d}_i = m^{-1} \sum_{j \in \mathcal{M}} \bar{d}_{ij}$ ,
- $\bar{d}_{ij} = n^{-1} \sum_{t=1}^n d_{ij,t}$ ,
- $d_{ij,t} = L_{i,t} - L_{j,t} \quad \forall i, j \in \mathcal{M}_0$ ,
- $L_{i,t}$ : une fonction de perte

On peut remarquer que :

- $\bar{d}_{ij}$  mesure la perte relative entre le  $i$ -ème modèle et le  $j$ -ème modèle
- $\bar{d}_i$  la perte du  $i$ -ème modèle par rapport à la moyenne des différents modèles  $\mathcal{M}$

La statistique de test  $T_{\max \mathcal{M}}$  est associée à l'**hypothèse nulle de capacité prédictive égale** (EPA)

- $H_{0,\mathcal{M}} : E(\bar{d}_i) = 0 \quad \forall i \in \mathcal{M}, \text{ où } \mathcal{M} \subset \mathcal{M}_0$
- $H_{1,\mathcal{M}} : E(\bar{d}_i) \neq 0$

Le MCS est une **procédure de test séquentielle** :

- élimination à chaque étape du plus "mauvais" modèle de  $\mathcal{M}$ , jusqu'à ce que l'hypothèse nulle d'EPA soit acceptée pour tous les modèles
- si l'hypothèse nulle d'EPA est rejetée pour  $\mathcal{M} = \mathcal{M}_0$ , alors le modèle le moins performant est exclu de l'ensemble  $\mathcal{M}$ .

Le choix du plus "mauvais" modèle à éliminer est fondé sur la règle d'élimination suivante :

$$e_{\max \mathcal{M}} = \arg \max_{i \in \mathcal{M}} t_i$$

- la procédure itérative s'arrête lorsque l'hypothèse nulle d'EPA des modèles encore inclus dans l'ensemble ne peut pas être rejetée.
- si  $H_{0,\mathcal{M}}$  est accepté au seuil de  $\alpha$  alors le MCS est l'ensemble défini par  $\widehat{\mathcal{M}}_{1-\alpha}^*$ .

Les  $p$ -values des statistiques  $T_{\max \mathcal{M}}$  sont calculées par la méthode du bootstrap.

- La fonction `dm.test` du package R [forecast](#) implémente les tests EPA de Diebold et Mariano (1995) et de Harvey et alii (1997)
- La fonction `DM.test` du package R [multDM](#) implémente les tests EPA de Diebold et Mariano (1995) et de Harvey et alii (1997)
- La fonction `MDM.test` du package R [multDM](#) implémente le test multivarié de Diebold et Mariano (1995) proposé par Mariano et Preve (2012)
- La fonction `MCSprocedure` du package R [MCS](#) implémente le test MCS de Hansen et alii (2011)

La fonction `LossVol` du package **MCS** permet de calculer différentes fonctions de pertes

$$MSE_1(SE1) = \frac{1}{H} \sum_{h=1}^H (\sigma_{t+h} - \hat{\sigma}_{t+h})^2$$

$$MSE_2(SE2) = \frac{1}{H} \sum_{h=1}^H (\sigma_{t+h}^2 - \hat{\sigma}_{t+h}^2)^2$$

$$QLIKE = \frac{1}{H} \sum_{h=1}^H [\ln(\hat{\sigma}_{t+h}^2) + \sigma_{t+h}^2 \hat{\sigma}_{t+h}^{-2}]$$

$$R2LOG = \frac{1}{H} \sum_{h=1}^H [\ln(\sigma_{t+h}^2 \hat{\sigma}_{t+h}^{-2})]^2$$

$$MAE_1(AE1) = \frac{1}{H} \sum_{h=1}^H |\sigma_{t+h} - \hat{\sigma}_{t+h}|$$

$$MAE_2(AE2) = \frac{1}{H} \sum_{h=1}^H |\sigma_{t+h}^2 - \hat{\sigma}_{t+h}^2|$$

`msegarch=LossVol(y[,1], y[,2], which = "SE2")`

`y[,1]`: mesure de la volatilité réalisée

`y[,2]`: mesure de la volatilité prévue

La fonction `MCSprocedure` du package **MCS** implémente le test MCS de Hansen et alii (2011)

```
loss=cbind(msegarch,mseegarch,msegjr,msetgarch)
summary(loss)
MCS <- MCSprocedure(Loss=loss[,1:4],alpha=0.1,B=5000,statistic='Tmax',cl=NULL)
```

Les fonctions `DM.test` et `MDM.test` du package **multDM** implémentent les tests EPA de Diebold et Mariano (1995) et de Harvey et alii (1997), et le test DM multivarié de Mariano et Preve (2012)

```
DM <- DM.test(y[,2],y[,3],y[,1],loss.type="SE",1,c=FALSE,H1="same")
```

```
MDM <- MDM.test(realized=y[,1],evaluated=y[,2:4],q=2,statistic="Sc",loss.type="SE")
```