

PRÉVISION D'UN INDICE DES CHOCS DU MARCHÉ AVEC DES MODÈLES HYBRIDES HMM-MLP

Madalina OLTEANU, Joseph RYNKIEWICZ,
SAMOS-MATISSE, Université Paris1

19 septembre 2003

Résumé

L'indice de crise IMS (Index of Market Shocks) a été introduit pour les marchés financiers comme analogue de l'échelle de Richter. Cet outil permet de mesurer l'amplitude et l'impact des crises financières. L'objet de cette étude est de lui ajuster un bon modèle et d'essayer d'en faire une bonne prévision. Dans ce papier, après avoir considéré des modèles classiques (linéaire et non-linéaire), on étudie un modèle hybride intégrant des perceptrons multi-couches et des chaînes de Markov cachées qui fournira, éventuellement, une séparation des états entre les crises et les périodes stables du marché.

Mots-clés : crises financières ; HMM ; réseaux de neurones

1 INTRODUCTION

Les “krachs” ou les “crises” financières dont on entend parler dans les médias sont traduites sur les marchés par des larges mouvements des prix. Qu’il s’agisse de la Grande Dépression qui a commencé en 1929, de la chute dans une seule journée de 22.6% des actions américaines en octobre 1987 ou des attaques terroristes du 11 septembre 2001, les marchés sont passés par des grandes turbulences depuis leur création.

Jusqu’à récemment, il n’y avait pas de moyen pour quantifier les crises et mesurer leur amplitude. Les mesures de risque traditionnelles, comme la volatilité implicite ou la volatilité historique, présentent trop de désavantages pour pouvoir être considérées comme de bons outils. En suivant l’approche initiale de Zumbach et al (2000), Maillet et Michel (2003) proposent le “Index of Market Shocks” (IMS), une mesure facilement calculable et interprétable qui est, en même temps, une échelle pour l’amplitude et l’étendu des crises financières et un outil de comparaison entre les différentes périodes de fortes turbulences. L’IMS est calculé de manière à intégrer les horizons des différents opérateurs intervenant sur le marché.

Brièvement, sa construction est une traduction financière de l’échelle de Richter qui mesure l’intensité des tremblements de terre. Le premier pas est de calculer des volatilités historiques, de

la plus basse à la plus haute fréquence disponibles, après, une analyse en composantes principales est faite sur ces volatilités et une fonction croissante de “l’énergie dissipée” (la volatilité, dans ce cas) est appliquée aux facteurs retenus.

Une fois l’indice construit, on peut aussi chercher à le prédire, une bonne prédiction étant, par exemple, très utile dans la gestion du risque systématique. Le but de ce papier est de chercher la meilleure méthode pour modéliser et prédire l’IMS. L’idée principale est de trouver un modèle hybride intégrant des perceptrons multi-couches et des chaînes de Markov cachées qui pourrait fournir, en plus d’un bon ajustement des données et une bonne prévision, une séparation des états qui va caractériser le comportement du marché pendant les crises et pendant les périodes plus stables.

2 MODÈLES AUTORÉGRESSIFS À CHANGEMENTS DE RÉGIME MARKOVIENS

Avant de présenter les résultats, faisons une courte description des modèles autorégressifs intégrant des chaînes de Markov cachées.

Les historiques des séries financières, considérés sur des longues périodes de temps, montrent des ruptures importantes dans le comportement des séries. Ces ruptures sont dues à plusieurs raisons, comme les grandes faillites et les moments de panique sur les marchés, les changements des politiques gouvernementales ou les guerres. Une méthode pour la modélisation de ces séries temporelles qui tient compte des changements de régime est l’utilisation des chaînes de Markov cachées.

Considérons $(y_t)_{t \in \mathbb{N}}$ les valeurs de la série observée et $(x_t)_{t \in \mathbb{Z}}$ une chaîne de Markov homogène définie par l’espace de ses états $E = \{e_1, \dots, e_N\}$, $N \in \mathbb{N}^*$ et par la matrice de transition A avec les éléments $a_{ij} = P(x_{t+1} = i \mid x_t = j)$, $i, j = 1, \dots, N$. En supposant, sans perte de généralité, que l’espace d’états de la chaîne est la base canonique de \mathbb{R}^N et en posant $v_{t+1} = x_{t+1} - E(x_{t+1} \mid x_t, x_{t-1}, \dots)$, un modèle autorégressif à changements de régime markoviens a la forme suivante :

$$\begin{cases} x_{t+1} = Ax_t + v_{t+1} \\ y_{t+1} = F_{x_{t+1}}(y_{t-p+1}^t) + \sigma_{x_{t+1}}\varepsilon_{t+1} \end{cases}$$

où y_{t-p+1}^t définie le vecteur (y_{t-p+1}, \dots, y_t) , $F_{x_{t+1}} \in \{F_{e_1}, \dots, F_{e_N}\}$ est une fonction autorégressive d’ordre p , $\sigma_{x_{t+1}} \in \{\sigma_{e_1}, \dots, \sigma_{e_N}\}$ est un réel strictement positif, $(\varepsilon_t)_{t \in \mathbb{N}}$ sont indépendants, identiquement distribués $N(0, 1)$.

Une étude d’Hamilton [3] sur la croissance du GNP (Gross National Product) aux États Unis est une des premières applications des modèles HMM (Hidden Markov Models), la série étant supposée un mélange de AR classiques. Une extension de ce modèle avec des fonctions autorégressives non linéaires comme le perceptron multicouches est immédiate.

3 ÉTUDE DE LA SÉRIE IMS (INDEX OF MARKET SHOCKS)

Les données utilisées sont les valeurs de l'IMS, calculé pour l'indice MSCI (Morgan Stanley Capital International) World Equities, depuis le 25 Juin 1998 jusqu'au 25 Juillet 2002 (1066 observations). On suppose que la série contient toute l'information sur elle-même et on n'ajoute pas d'autres variables explicatives. Afin d'éviter les problèmes numériques, on a choisi de travailler sur les données centrées et normées.

3.1 Le modèle linéaire

La série étant fortement autocorrélée, on essaie d'abord d'ajuster un modèle linéaire classique de type ARMA sur les données. On décide d'étudier tous les modèles ARMA(p,q), pour p et q dans des rangs spécifiés, et garder celui qui minimise le critère d'information BIC. La procédure est basée sur l'estimation du maximum de vraisemblance et les valeurs initiales pour les coefficients AR sont calculées en utilisant les équations de Yule-Walker.

Le modèle final retenu est un AR(6) avec le BIC égal à -0.99202 et avec les coefficients suivants :

$$\begin{aligned}ims_t = & 0.65312 \cdot ims_{t-1} + 0.28499 \cdot ims_{t-2} - 0.02787 \cdot ims_{t-3} - 0.13745 \cdot ims_{t-4} \\ & - 0.182 \cdot ims_{t-5} + 0.24094 \cdot ims_{t-6} + e_t\end{aligned}$$

et une variance des résidus $\{e_t\}$ égale à 0.3566.

3.2 Le modèle MLP

Le prochain pas est d'estimer et d'identifier un perceptron multicouches et vérifier si la non-linéarité améliore les résultats obtenus auparavant avec un AR(6). On utilise, pour cela, l'algorithme "Statistical Stepwise"[2], développé dans le logiciel REGRESS[5].

Le modèle final retenu a une couche cachée avec deux unités et douze paramètres à estimer. L'erreur moyenne quadratique est 0.34191 et le BIC est égal à -0.98778 :

$$\begin{aligned}ims_t = & 4.26771 \cdot \varphi(0.27390 \cdot ims_{t-2} - 0.14530 \cdot ims_{t-4} - 0.05999 \cdot ims_{t-5} - 0.50868) \\ & + 1.96573 \cdot \varphi(0.40941 \cdot ims_{t-1} - 0.44074 \cdot ims_{t-2} + 0.23265 \cdot ims_{t-4} \\ & + 0.15018 \cdot ims_{t-6} - 0.50759) + 2.79593\end{aligned}$$

où φ est la fonction tangente hyperbolique.

Si on compare les résultats dans les cas linéaire et non-linéaire, on voit que, même si on obtient une amélioration de 4% avec le perceptron (0.3566 et, respectivement, 0.34191), le BIC ne décroît pas pour le modèle non-linéaire (-0.99202 et, respectivement, -0.98778).

Pour tester les prévisions faites par le perceptron contre celles du modèle linéaire, on a divisé la série en deux ensembles : un ensemble d'apprentissage contenant les 800 premières observations sur lequel on fait les estimations d'un AR(6) et d'un perceptron à deux unités cachées et un ensemble de validation avec le reste de 266 valeurs et qu'on utilise pour tester les prévisions des deux modèles. Sur l'ensemble d'apprentissage, on a des erreurs moyennes de 0.33965

pour le modèle linéaire et de 0.33989 pour le perceptron. Les erreurs de prévision, calculées sur l'ensemble de validation, sont 0.34685 et, respectivement, 0.34989.

Vus les résultats obtenus et le nombre de paramètres à estimer (6 et, respectivement, 12), on préfère, pour l'instant, l'emploi d'un modèle linéaire AR(6) pour l'IMS.

3.3 Recherche d'un modèle MLP-HMM

On utilise un modèle hybride intégrant des perceptrons multi-couches et une chaîne de Markov cachée pour étudier s'il y a des changements de régime dans la série et si on peut distinguer entre différents comportements du marché. On a regardé plusieurs architectures possibles et on a retenu deux configurations, sur la base des critères suivants : la première est celle qui minimise l'erreur moyenne quadratique et la deuxième est celle qui a la "meilleure" matrice de transition. Le choix de la "meilleure" matrice de transition se fait en maximisant la trace de la matrice divisée par sa dimension car, en maximisant ce critère, on a observé, empiriquement, qu'on a alors la meilleure segmentation de la série.

Tout d'abord, il faudrait faire quelques remarques sur la procédure utilisée. L'estimation est faite en maximisant la log-vraisemblance avec un algorithme de second ordre (BFGS). Pour éviter les minimas locaux, on a fait plusieurs initialisations différentes de l'algorithme, avec 200 itérations pour chacune des estimations. On suppose qu'au début on n'a pas d'information sur les experts et donc on ne favorise aucun d'entre eux, en prenant les probabilités égales dans la matrice de transition initiale. Les poids des perceptrons sont initialisés à 1 (sauf pour ceux qui correspondent aux retards manquants et qu'on initialise à 0), les fonctions d'activation sont des tangentes hyperboliques et les innovations sont supposées gaussiennes.

Le modèle qui minimise l'erreur moyenne quadratique

Les premiers résultats qu'on présente ici correspondent au modèle qui minimise l'erreur moyenne quadratique. Pour chacune des configurations, on a fait dix apprentissages sur toute la série des 1066 observations. Le modèle retenu a trois experts et la matrice de transition estimée est :

$$\hat{A} = \begin{pmatrix} 0.68352 & 0.24899 & 0.27525 \\ 0.27959 & 0.60700 & 0.16191 \\ 0.03689 & 0.14401 & 0.56284 \end{pmatrix}$$

Les deux premiers experts ont trois unités cachées et les variances associées sont 0.03616 et 0.20769. Le troisième perceptron a deux unités cachées et une variance associée de 0.47321.

Pour comparer les prévisions faites par ce modèle avec celles fournies par les modèles linéaire et non-linéaire (MLP), on divise la série en un ensemble d'apprentissage comprenant les 800 premières observations et un ensemble de test qui contient 266 valeurs. On fait une nouvelle estimation, en prenant des valeurs initiales égales aux paramètres estimés antérieurement. On obtient une erreur moyenne quadratique de 0.32796 sur l'ensemble d'apprentissage et de 0.37429 sur l'ensemble de validation. Ces deux valeurs suggèrent un phénomène de sur-apprentissage, même si on était tenté de dire que ce modèle ajustait mieux les données.

Le modèle avec la “meilleure” matrice de transition

Le deuxième modèle retenu par la procédure correspond à la “meilleure” matrice de transition. Il a deux experts et la matrice de transition estimée est :

$$\hat{A} = \begin{pmatrix} 0.90808 & 0.14065 \\ 0.09192 & 0.85935 \end{pmatrix}$$

Le premier expert a trois unités cachées avec un bruit associé de variance 0.43045 et le deuxième est en fait un modèle linéaire (pas d’unités cachées) avec une variance associée 0.05637. L’erreur moyenne quadratique sur toute la série est égale à 0.34789.

La figure 1 représente la série (afin d’éviter une lecture difficile, on a choisi de garder seulement les 200 premières valeurs), les probabilités conditionnelles du premier expert (les probabilités pour le deuxième expert s’en déduisent immédiatement) et les résidus. Une première inspection du graphique indique le fait que le perceptron pourrait correspondre aux périodes de turbulences sur le marché.

Pour tester les prévisions obtenues avec ce modèle, on fait une nouvelle estimation en divisant la série dans un ensemble d’apprentissage contenant les 800 premières observations et un ensemble de test de 266 valeurs. On fait dix apprentissages différents avec 200 itérations chaque fois et on prend comme valeurs initiales les paramètres estimés auparavant. On obtient une erreur moyenne de 0.35228 sur l’ensemble d’apprentissage et 0.36626 sur l’ensemble de validation, ces résultats n’étant pas meilleurs que ceux fournis par le modèle linéaire.

Même si l’utilisation des modèles hybrides n’a pas amélioré l’erreur de prédiction pour l’IMS, l’étude des probabilités conditionnelles est intéressante puisqu’on a une séparation entre deux états qui signifient deux comportements différents de l’indice et du marché. D’une part, on a un premier expert pour les périodes de crises et de fortes turbulences et d’autre part, on a le deuxième expert pour les périodes plus stables. Si on représente graphiquement les probabilités conditionnelles du premier expert, on voit qu’elles se rapprochent beaucoup de 1 pendant les crises.

La figure 2 représente les probabilités conditionnelles du premier expert pendant la crise russe d’août 1998. Les principales causes ont été la forte dévaluation de la rouble (de l’ordre de quatre par rapport au dollar américain) et le défaut de paiement du pays, combinés avec les échos de la crise de l’Asie de l’Est en 1997. Sur le graphique, on voit que les pics correspondent aux dates les plus importantes de la crise : le 11 août, le marché des actions atteint son plus bas niveau pendant que le marché des obligations à court terme s’effondre, le 17 août, la Banque Centrale élargie l’intervalle d’échange rouble/dollar et un moratoire de 90 jours sur le paiement de la dette internationale est annoncé, le 27 août, le plus grand marché d’échange de Russie, le MICEX, arrête les transactions sur devises.

Le cas suivant qu’on a représenté graphiquement est la crise brésilienne de janvier 1999, due principalement à l’important déficit fiscal et au taux d’inflation élevé. On peut aussi lier les pics du graphique aux moments décisifs du krach : le 7 janvier, le gouverneur du troisième plus grand état annonce un moratoire de 90 jours sur le paiement de la dette au gouvernement fédéral, le 13 janvier, l’intervalle d’échange pour le réal est élargi et le 15 Janvier, la Banque Centrale abandonne sa défense du réal malgré les importantes réserves financières en dollars américains.

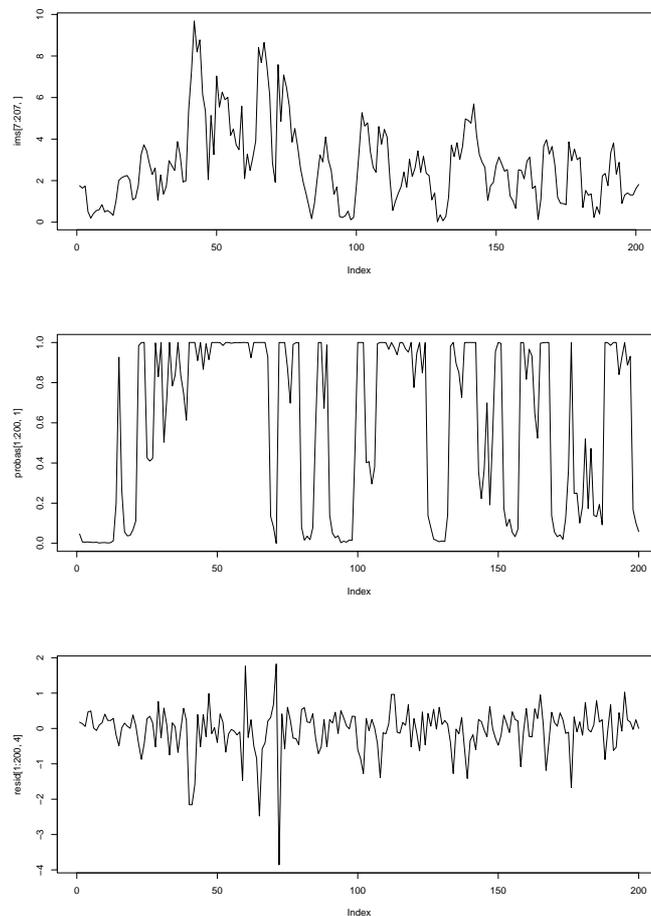


FIG. 1 – Le modèle avec la “meilleure” matrice de transition

Le dernier exemple qu’on propose ici c’est la crise de septembre 2001. Même avant les attaques terroristes, il y avait des signes de récession de l’économie américaine. Sur le graphique, on a les deux pics, le premier après le 6 septembre quand les cours des actions commencent à baisser de manière accélérée et le deuxième correspondant à l’ouverture retardée après le 11 Septembre.

4 CONCLUSION

L’étude d’un modèle hybride pour la série IMS nous a fourni une séparation des états avec deux experts, un pour les crises et un pour les périodes stables. Quand on représente graphiquement les probabilités de transition conditionnelles pour le premier expert, on peut extraire les dates correspondantes aux décisions ou aux événements générant des fluctuations sur le marché. En ce qui concerne les prédictions, il n’y a pas d’amélioration par rapport au modèle linéaire

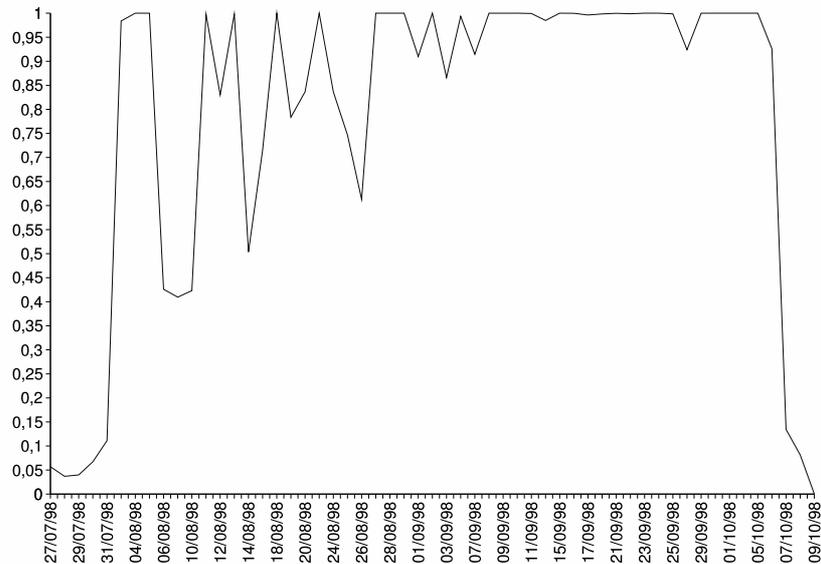


FIG. 2 – La crise russe

qu'on préfère pour sa simplicité et son utilisation facile.

Références

- [1] Brockwell P.J., Davis R.A. (2002), Introduction to Time Series and Forecasting, Second Edition, Springer-Verlag, New York
- [2] Cottrell M. and al (1995), Neural Modeling for Time Series : a Statistical Stepwise Method for Weight Elimination, IEEE Transaction on Neural Networks, 6, 1355-1364
- [3] Hamilton J.D. (1994), Time Series Analysis, Chapter 22 Modeling Time Series with Changes in Regime, Princeton University Press, 677-703
- [4] Maillet B., Michel T. (2003), An Index of Market Shocks Based on Multiscale Analysis, Quantitative Finance 3, 88-97
- [5] Rynkiewitz J. (2000), Modèles hybrides intégrant des réseaux de neurones artificiels à des modèles de chaînes de Markov cachées : applications à la prédiction des séries temporelles, Thèse de doctorat
- [6] Venables W.N., Ripley B.D. (2000), S Programming, Springer-Verlag, New York
- [7] Zumbach G., Dacorogna M., Olsen J., Olsen R. (2000), Measuring Shocks in Financial Markets, International Journal of Theoretical and Applied Finance 3, 347-355

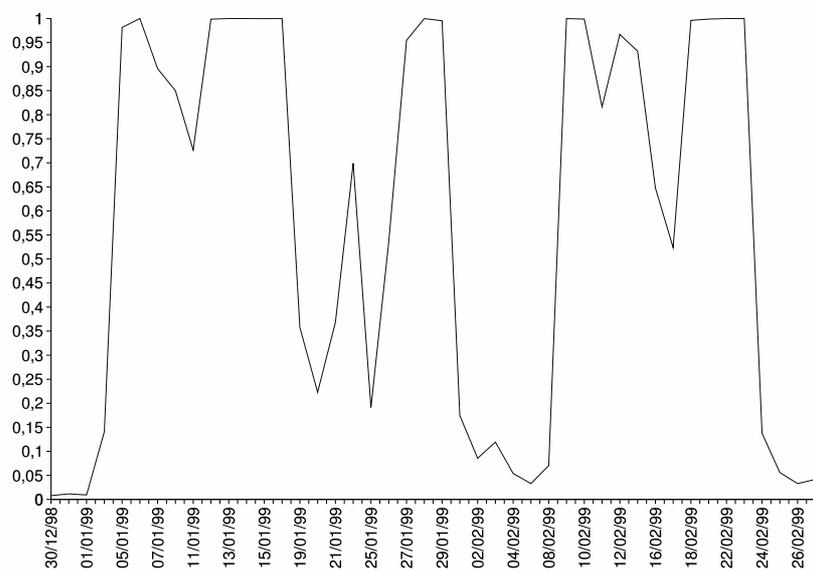


FIG. 3 – La crise brésilienne

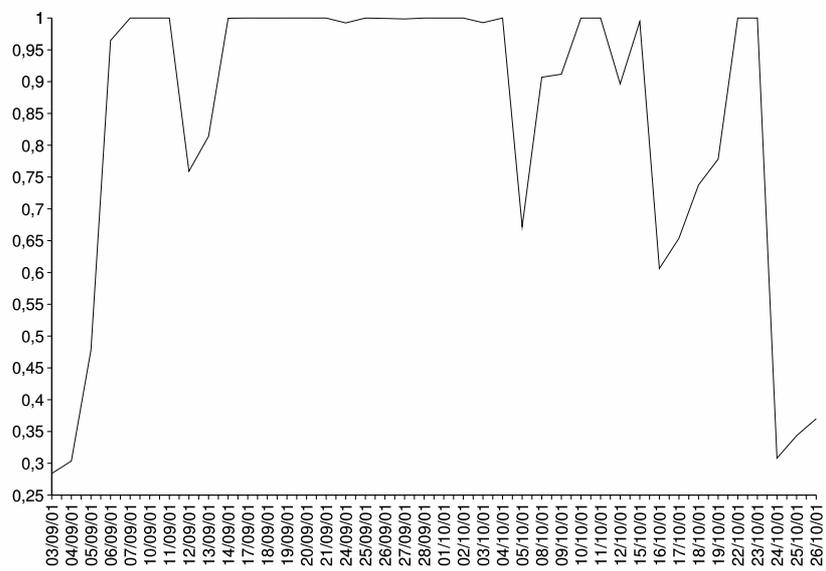


FIG. 4 – La crise de septembre 2001