



UNIVERSITÉ DE NANTES



IAE NANTES
ÉCONOMIE & MANAGEMENT

ÉCONOMÉTRIE DES DONNÉES À HAUTE-FRÉQUENCE

High-Frequency Financial Econometrics

Modeling and forecasting realized volatility



Teodoro MOUNIER TEBAS

Enseignant : M. SÉVY
Année universitaire : 2020-2021

Master 2 Économétrie et Statistiques, parcours Économétrie Appliquée

Table des matières

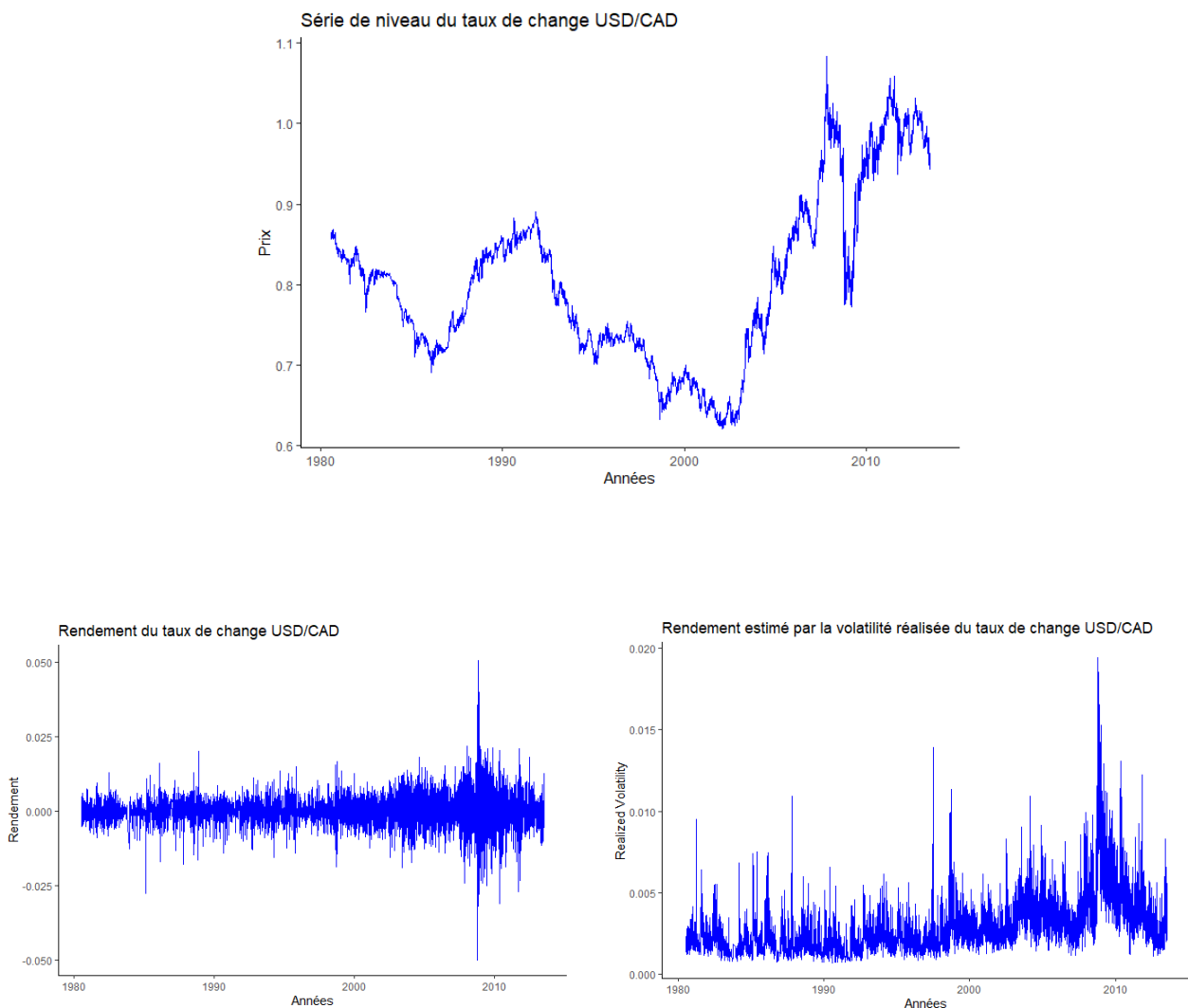
I	Propriétés de la série volatilité réalisée	2
I.1	Les données	2
I.1.1	Représentations graphiques de la série	2
I.1.2	Distribution des séries	3
I.1.3	Graphiques de corrélations des rentabilités	4
I.1.4	Statistiques descriptives de la série corrigée	4
II	Modélisation de la volatilité réalisée	5
II.1	Modélisations HAR	5
III	Forecasting volatility	6
IV	Annexe	7

I Propriétés de la série volatilité réalisée

I.1 Les données

L'intérêt de ce développement est d'estimer l'évolution du taux de change¹ entre le dollar américain et le dollar canadien depuis juillet 1980 à août 2013 dans le but de faire de la prédiction. Pour l'estimation et la prédiction de séries financière, plusieurs méthodes sont envisageables en fonction des données et des objectifs. Dans notre cas nous allons estimer la **Variance réalisée** (RV), la **Volatilité réalisée** (Rvol) et enfin le **Logarithme de la volatilité réalisée** (LogRVol) par une spécification **HAR**². Les données sont journalières ce qui nous fait 6869 observations pour la période correspondant à 33 années. Après avoir retiré les valeurs manquantes et traitements de la base pour les modélisations nous sommes restées à 6825 observations. Les graphiques ci-dessous représentent la série en niveau puis les rendements et enfin la volatilité réalisée.

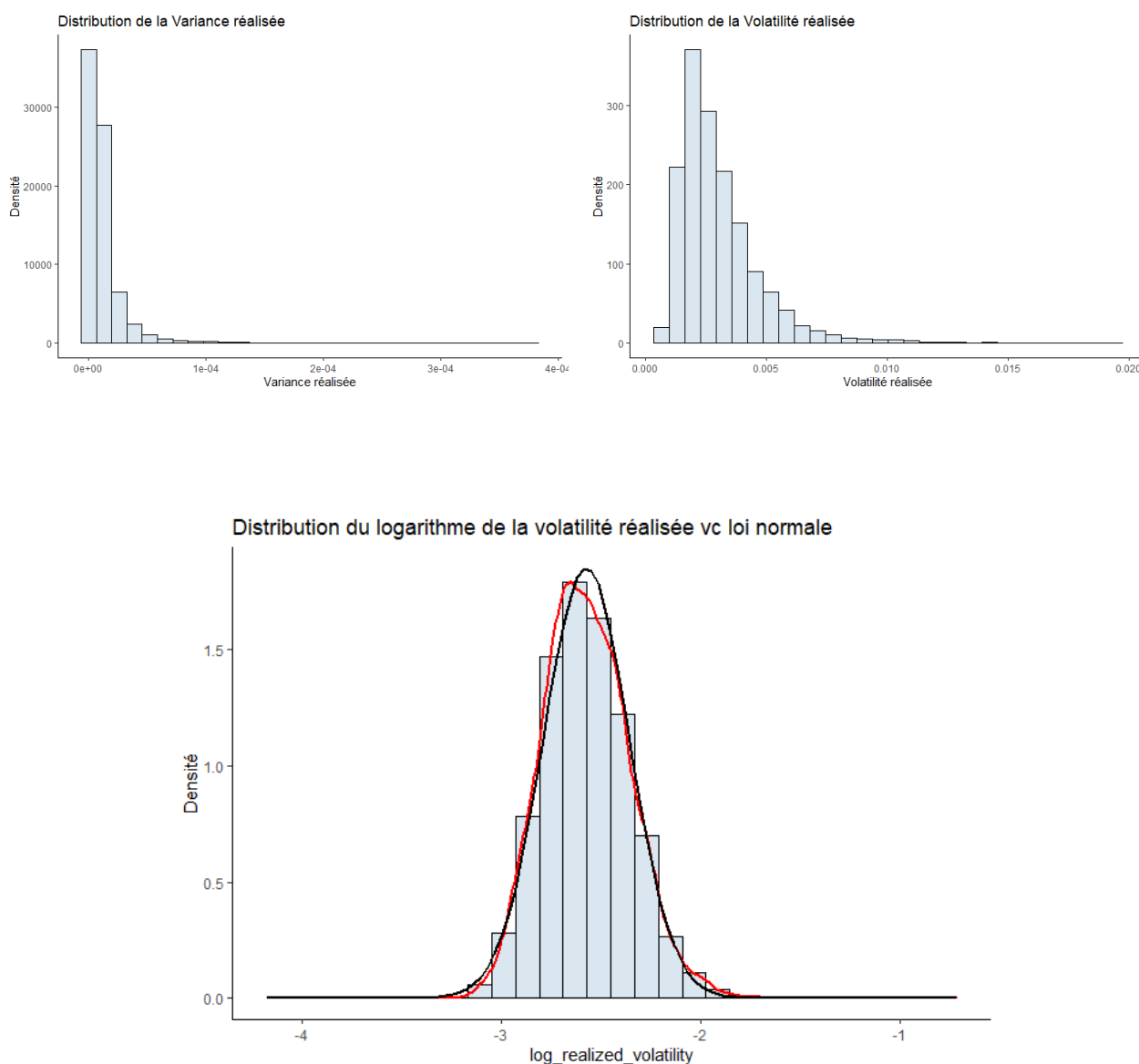
I.1.1 Représentations graphiques de la série



1. C'est le prix d'une devise (monnaie) dans une autre devise. Dans notre cas le taux de change USD/CAN c'est le prix du dollar américain en dollars canadien.

2. heterogeneous autoregressive model for realised volatility

I.1.2 Distribution des séries



Les distributions représentées ci-dessus nous permettent d'observer la non-normalité de la distribution de la variance réalisée ainsi que de la volatilité réalisée. Cependant en ce qui concerne le log de la volatilité réalisée (courbe rouge) nous avons une distribution très proche de la normale (courbe noire). La courbe rouge est légèrement asymétrique à gauche ce qui fait référence à une asymétrie négative ce qui correspond à des mouvements plus fort à la baisse qu'à la hausse. Nous allons maintenant observer l'**autocorrélation**³ de ces différentes mesures.

3. L'autocorrélation d'une série fait référence au fait que dans une série temporelle, la mesure d'un phénomène à un instant t peut être corrélée aux mesures précédentes (t_1 , t_2 , etc.). Une série autocorrélée est ainsi corrélée à elle-même, avec un décalage (lag) donné.

I.1.3 Graphiques de corrélations des rentabilités

Nous voyons ci-dessous les fonctions d'autocorrélation représentées par un corrélogramme pour la variance, volatilité et log de la volatilité réalisée. Nous pouvons constater d'emblée une autocorrélation très forte, peu importe la spécification. En effet, il faut spécifier un retard de k égal 2000 (jours) pour que l'autocorrélation ne soit plus significative.

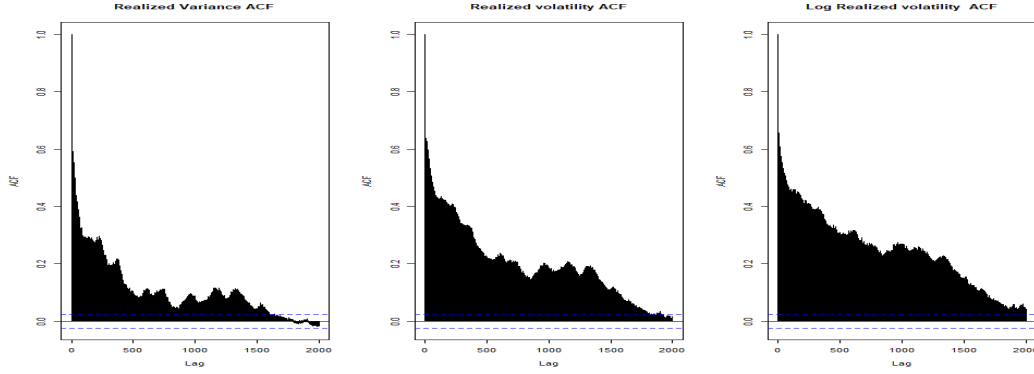


Figure 1 – Autocorrélation de RV, RVol, LogRVol

I.1.4 Statistiques descriptives de la série corrigée

Avant de passer aux modélisations nous avons résumé l'information dans un tableau descriptif de chacune des séries ainsi que des **Jump**⁴. La table ci-dessous résume la valeur minimum, maximum et moyenne de chaque série. Les coefficients d'asymétrie (**Skewness**) et d'aplatissement (**Kurtosis**) qui nous permettent de confirmer ce que nous avons observé précédemment dans les plots de distribution des séries. Enfin, les tests Ljung-Box basés sur l'hypothèse nulle de la non-autocorrélation confirment la présence d'autocorrélation des séries puisque les p-values sont toutes significatives.

Table 1 – Statistiques descriptives

	RV_t	$RV_t^{1/2}$	$\log(RV_t^{1/2})$	J_t	$J_t^{1/2}$	$\log(J_t + 1)$
Mean	0.000012	0.00303	-2.574492	0.000012	0.003522	0.000005
St. dev	0.000018	0.001718	0.215552	0.000774	0.012895	0.000336
Skewness	6.625	2.205	0.272	1.531	3.77	1.471
Kurtosis	76.292	8.719	-0.001	52.416	14.105	51.995
Min.	0.000000	0.000663	-3.178590	-0.008298	0.000000	-0.003619
Max.	0.000378	0.019444	-1.711217	0.013647	0.116820	0.005887
LB ₁₀	22205***	28953***	28527***	10.981	7.2681	11.145

4. "Saut" en français, représente les changements brusques dans le niveau moyen d'une série temporelle

II Modélisation de la volatilité réalisée

II.1 Modélisations HAR

Les statistiques descriptives nous ont permis de mieux comprendre les séries, en mettant en avant de l'autocorrélation forte ainsi que de l'**hétéroscédasticité**⁵ lorsque l'on observe les représentations graphiques de la série. Ces résultats nous confirment la pertinence d'estimer la série par une spécification HAR qui prendrait en compte l'hétéroscédasticité ainsi que l'autocorrélation de la série.

Nous définissons la volatilité réalisée moyenne sur la période $t + 1, t + h$ par :

$$RV_{t+1,t+h} = h^{-1}[RV_{t+1} + RV_{t+2} + \dots + RV_{t+h}]$$

Nous pouvons donc écrire le modèle HAR de la manière suivante :

$$RV_{t+1,t+h} = \beta_0 + \beta_1 RV_t + \beta_5 RV_{t-1,t-4} + \beta_{22} RV_{t-5,t-21} + \epsilon_t$$

Au final, 9 modèles ont été estimés par les MCO pour le taux de change USD/CAD. En effet, nous avons estimé à trois horizons différents chacune des versions de la volatilité. Pour $h = 1, 5$ et 22 jours dans sa "version brute" (RV = variance réalisée), sa version volatilité réalisée (racine carrée de la variance réalisée) et sa version logarithme de la variance réalisée. Les estimateurs obtenus sont résumés dans le tableau ci-dessous.

Table 2 – Daily, Weekly, and Monthly Regression

	$RV_{t,t+h}$			$(RV_{t,t+h})^{1/2}$			$\log((RV_{t,t+h})^{1/2})$		
h	1	5	22	1	5	22	1	5	22
β_0	6.464e-07 (1.991e-07)	8.791e-07 (1.218e-07)	1.538e-06 (1.037e-07)	1.279e-04 (3.043e-05)	1.796e-04 (1.908e-05)	3.051e-04 (1.686e-05)	-1.1924e-01 (0.02380)	-1.63996e-01 (0.015452)	-2.85821e-01 (0.013866)
β_D	1.190e-01 (1.410e-02)	8.458e-02 (8.550e-03)	4.733e-02 (7.346e-03)	1.454e-01 (1.424e-02)	1.013e-01 (9.081e-03)	5.564e-02 (7.893e-03)	1.7164e-01 (0.01434)	1.21563e-01 (0.009524)	6.7464e-02 (0.008356)
β_W	1.895e-01 (2.969e-02)	9.486e-02 (1.781e-02)	3.226e-01 (1.547e-02)	2.998e-01 (2.824e-02)	2.352e-01 (1.764e-02)	2.948e-01 (1.565e-02)	3.5856e-01 (0.02699)	3.01557e-01 (0.017562)	2.57973e-01 (0.015730)
β_M	6.399e-01 (2.839e-02)	7.538e-01 (1.702e-02)	5.057e-01 (1.479e-02)	5.133e-01 (2.589e-02)	6.079e-01 (1.615e-02)	5.504e-01 (1.435e-02)	4.2338e-01 (0.02433)	5.12514e-01 (0.015804)	5.63248e-01 (0.014175)
β_J	-7.033e-04 (1.949e-04)	-1.490e-05 (7.096e-06)	-6.437e-04 (2.340e-04)	-5.170e-02 (1.644e-02)	-2.557e-03 (6.300e-04)	-5.112e-02 (2.099e-02)	-4.53756 (2.07124)	-3.85967e-01 (0.082697)	-5.881968 (2.780640)
$R^2_{HAR-RV-J}$	0.54	0.7573	0.7876	0.6251	0.804	0.8255	0.6215	0.7862	0.8027
R^2_{HAR}	0.5392	0.7572	0.7874	0.6246	0.8036	0.8254	0.6213	0.7856	0.8026

À partir des résultats du tableau, nous pouvons constater que tous les coefficients sont significatifs, quel que soit l'horizon ou la version estimée. L'inclusion des **jumps** dans les estimations est non seulement significatif, mais améliore la qualité d'ajustement des modèles, car les $R^2_{HAR-RV-j}$ sont toujours supérieur aux R^2_{HAR} c'est-à-dire les spécifications n'incluant pas les jumps.

Enfin, nous pouvons conclure que la version s'ajustant le mieux à notre série temporelle est la version où l'on estime la **volatilité réaliser mensuelle** avec la prise en compte des jumps. Nous obtenons un $R^2 = 0.8255$ ce qui est très bon comme ajustement. Cependant, dans la section suivante nous allons faire de la prévision uniquement à partir des estimations à la journée.

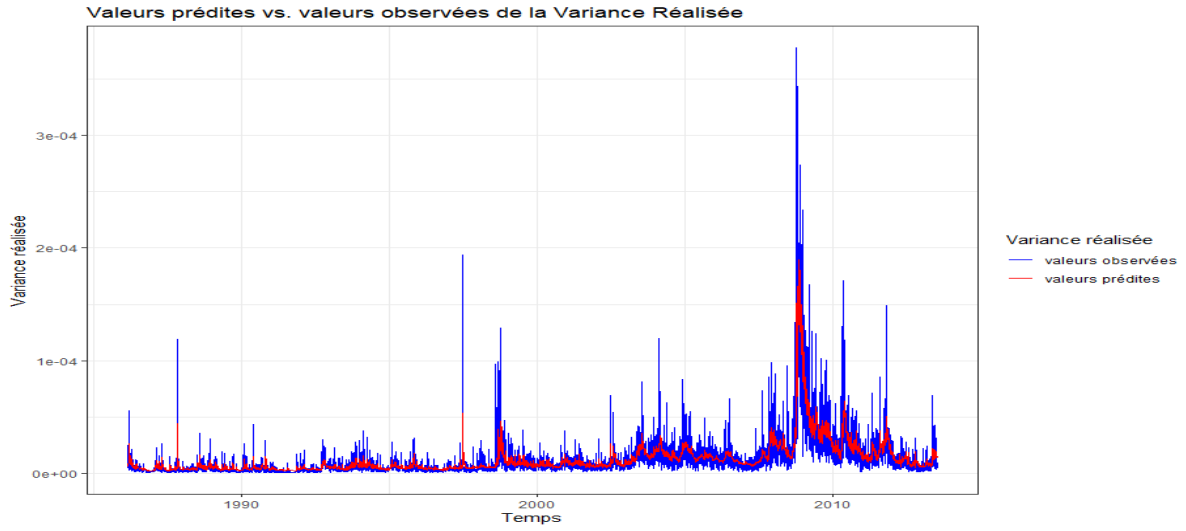
5. Dans les séries temporelles l'hétéroscédasticité correspond à des clusters de volatilité c'est-à-dire des successions de périodes de forte et de faible volatilité.

III Forecasting volatility

Dans cette section, nous avons utilisé les modèles estimés ci-dessus pour faire de la prévision de la Variance, volatilité et log de volatilité réalisée à un horizon d'une journée. Nous calculerons ensuite la **MSE**⁶ et nous retiendrons le modèle qui a la meilleure MSE.

Concrètement, nous réestimerons les modèles sur une fenêtre roulante de 1000 jours puis nous réaliserons nos prévisions pour une journée et ainsi de suite jusqu'à arriver aux 5825 jours prédit. Une fois les modèles estimés, nous comparerons les valeurs prédites aux valeurs réalisées afin d'en retirer les MSE.

Enfin, ci-dessous nous avons représenté graphiquement les prévisions (en rouge) ainsi que les valeurs réelles (en bleu) pour la variance réalisée. Visuellement, nous observons que la ligne rouge s'ajuste bien à la courbe bleue même si elle est moins volatile que cette dernière. En annexe se trouvent les représentations graphiques pour la volatilité réalisée ainsi que le Log de la volatilité réalisée.



Pour terminer, voici les résultats de la MSE pour chacune des prédictions. On constate que les prédictions qui se rapprochent le plus des valeurs réelles sont celles de la variance réalisée avec le carré des erreurs moyennes égales à $1.820035e-10$. Nous pouvons conserver le modèle estimé avec la variance réalisée comme notre meilleure modélisation.

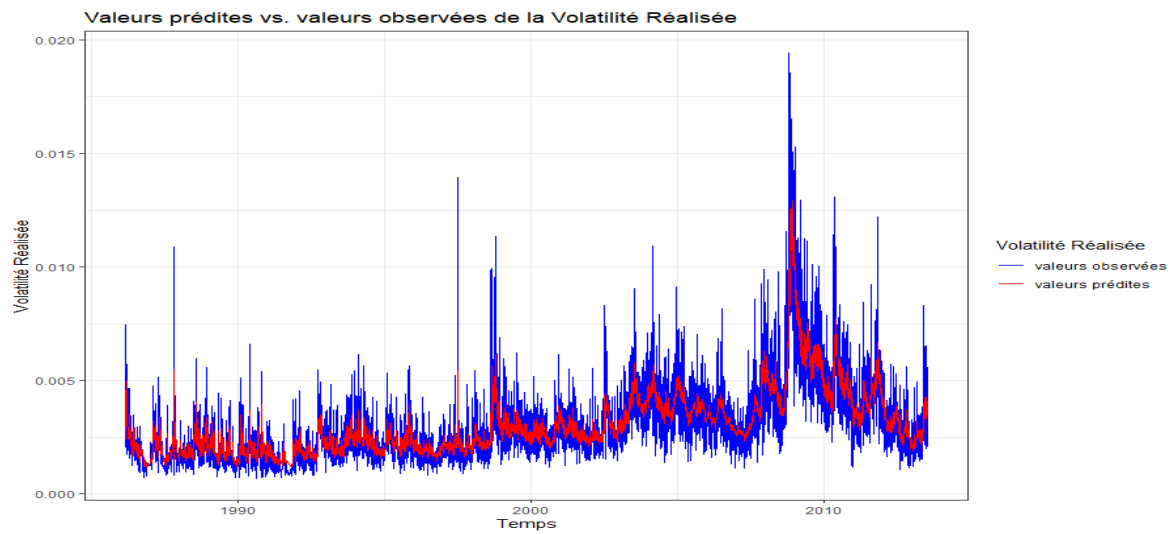
Table 3 – Comparing Estimations

	RV_t	$RV_t^{1/2}$	$\log(RV_t^{1/2})$
MSE	1.820035e-10	1.205095e-06	0.01777948

6. Mean Squared Error, c'est la moyenne arithmétique des carrés des écarts entre prévisions du modèle et observations. C'est donc une valeur à minimiser.

IV Annexe

Annexe n°1 : Prévisions de la volatilité réalisée



Annexe n°2 : Prévisions du Log de volatilité réalisée

