



UNIVERSITÉ DE NANTES



**IAE NANTES**  
ÉCONOMIE & MANAGEMENT

**Econométrie Haute Fréquence:**  
Estimation d'un modèle HAR sur la volatilité du  
USD/CAD

Aurouet Lucas

Présenté pour le master II EKAP  
sous la direction de Mr. Sevi

Master EKAP  
IAE Nantes  
01/01/2020

# 1 Introduction

Il est désormais reconnu que la volatilité est extrêmement persistante et présente des auto-corrélations significatives jusqu'à plusieurs semaines, voire mois. Beaucoup de travaux ont été entrepris pour tenter de modéliser la volatilité d'un actif tirant profit de cette persistance. Les modèles de la famille GARCH (Generalized Auto Regressive Heteroskedasticity) sont probablement les exemples le plus célèbres. En revanche cette famille de modèles s'utilise avec des données dont la fréquence est relativement faible (journalière tout au plus). Avec l'explosion du volume des données financières, nous pouvons désormais avoir accès à des données de haute fréquence, avec plusieurs points de données dans la même journée. L'augmentation de la fréquence des données requiert des nouveaux modèles, adaptés à ces nouveaux volumes, qui savent prendre en compte l'information fournie par les mouvements de prix à l'intérieur d'une même journée. Dans cette optique, Corsi développe le HAR (Heterogenous Auto Regressive) en 2002<sup>1</sup> avec l'intention de tirer profit des ces nouvelles données intra-journalières. De plus le modèle HAR s'intéresse à l'idée que les objectifs des acteurs sur les marchés financiers ne sont pas forcément définis pour le même horizon. En fonction de l'information qui les intéresse, les acteurs ne se soucient que d'une fréquence en particulier. Elle peut être journalière, hebdomadaire, mensuelle ou autre, mais HAR repose sur le postulat que tout les individus ne vont pas forcément s'intéresser à la même fréquence. Chaque horizon aura un impact différent sur la volatilité et chaque mesure de la volatilité remet en cause les objectifs aux différents horizons. HAR combine l'intérêt des données haute fréquence et les décisions hétérogènes<sup>2</sup> des acteurs sur les marchés pour estimer la volatilité des actifs. Ici nous nous emploierons donc à estimer puis prédire la volatilité du taux de change USD/CAD à l'aide du modèle HAR. Nous présenterons brièvement les données utilisées puis nous passerons à l'estimation du modèles sur les données brutes et transformées<sup>3</sup> et enfin nous utiliserons les modèles estimés pour prédire la volatilité sur la quasi-totalité de la période.

## 2 Données

Les données sont celles de la volatilité quotidienne du taux de change USD/CAD du 21/07/1980 au 16/08/2013, pour un total de 6869 observations. Comme nous l'avons mentionné plus haut, la volatilité est estimée à partir de données intra-journalières. Précisément, la volatilité est estimée par la somme des logarithmes des rendements intraday. Cette mesure de la volatilité, appelée Realized Variance (RV dans la suite) est le meilleur proxy de la volatilité disponible à l'heure actuelle. En parallèle, les données contiennent également les jumps quotidiens que nous inclurons dans le modèle.

Le modèle HAR se base sur le principe que la volatilité est un processus

---

<sup>1</sup>A Simple Long Memory Model if Realized Volatility - F.Corsi - 2002

<sup>2</sup>D'où son nom

<sup>3</sup>Nous reviendrons plus tard sur l'intérêt de transformer les variables

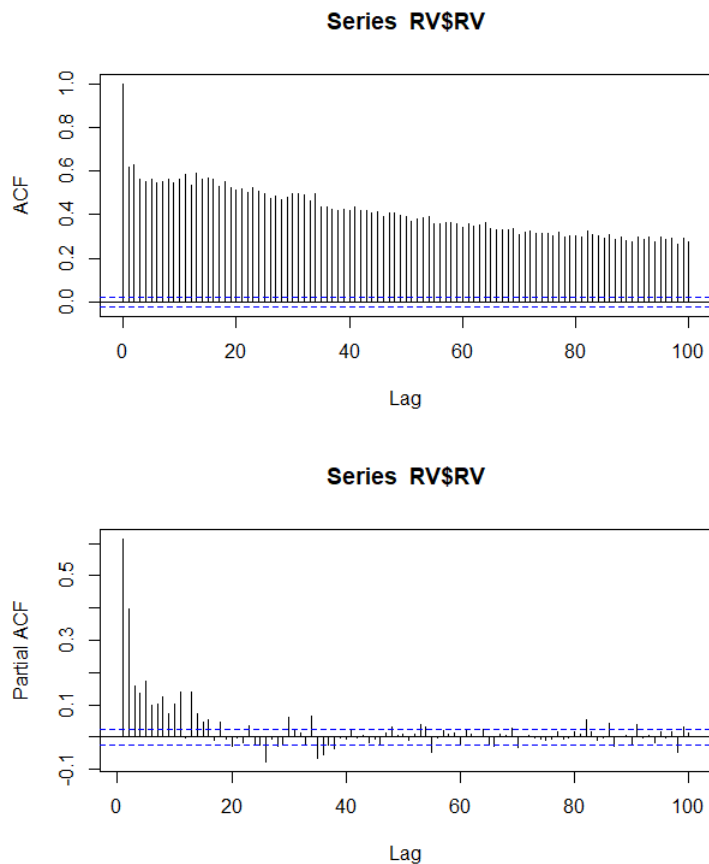


Figure 1: ACF et PACF de la RV

à mémoire longue. Autrement dit la volatilité est très persistante et les auto-corrélations sont significatives jusqu'à des horizons lointains. Il convient donc dans un premier temps de vérifier si la volatilité est belle et bien persistante, de sorte à ce que l'estimation par un modèle auto régressif soit pertinente. Les graphiques ci-dessous représente les fonctions d'auto-corrélations totale et partielle.

Aux vues des ACF et PACF, la volatilité semble bien être un processus à mémoire longue. Les auto-corrélations sont significatives jusqu'à des horizons très lointains (plus de 100 jours). La fonction d'auto-corrélation partielle est significative jusqu'à une vingtaine de jours. Nous notons la possibilité que, de ce fait, le coefficient associé à la RV mensuelle ne soit pas significatif. Quoiqu'il en soit, modéliser la variance réalisée par un processus auto-régressif semble être judicieux.

Nous savons que les rendements sont rarement distribués selon une loi nor-

Série	Skewness	Excess Kurtosis
Brute	6.62	76.16
$\text{Log}(\text{Brute})$	0.27	0.01
$\text{Brute}^{1/2}$	2.20	8.71

Table 1: Statistiques descriptive de la série

male, bien que cette hypothèse soit particulièrement avantageuse lors de l'estimation de certains modèles économétriques. En particulier ici, le modèle HAR sera estimé par MCO<sup>4</sup> et l'efficacité du modèle repose partiellement sur cette hypothèse. Si les rendements ne sont pas distribués normalement, il y a de fortes chances que la RV ne le soit pas non plus. Si la RV n'est pas distribuée normalement, les paramètres estimés par MCO ne seront pas sans biais ni efficaces. Il convient donc de vérifier la distribution de nos données avant de passer à l'estimation de modèles. Nous sommes donc particulièrement intéressés par la skewness et l'excess kurtosis qui nous aideront à déterminer la distribution de notre variable. Dans le cas où les données ne sont effectivement pas distribuées normalement, une transformation (logarithme ou racine carrée) pourrait permettre de résoudre le problème.

La série brute a une skewness positive de 6.62, qui indique que la distribution de la volatilité est étalée à droite et fortement asymétrique. L'excès de kurtosis est lui aussi largement positif (76.16), les queues de distributions sont bien plus épaisses que celles d'une loi normale. Transformer les données à l'aide de la fonction racine carrée semble donner une distribution de la volatilité plus proche de celle de la loi normale. Bien que la distribution de données soit toujours asymétrique et leptokurtique, les valeurs associées à la skewness et au kurtosis sont moins extrêmes que pour la série brute. Passer la série en logarithme permet de se débarrasser des queues de distribution (excess kurtosis inférieur à 0.1) et rend la distribution des données quasiment symétrique (skewness proche de 0). On constate bien que la série brute ne respecte pas parfaitement toutes les conditions qui rendent l'estimation par MCO efficace. Nous nous attendons à avoir de meilleures performances des modèles estimés sur les données transformées, notamment, la série passée en logarithme. Nous pouvons visualiser les éléments que nous venons de passer en revue en superposant les distributions des différentes séries contre une distribution normale.

On remarque bien l'excès de kurtosis et l'asymétrie de la série brute qui disparaissent légèrement avec la transformation en racine carrée. La distribution de la série transformée en logarithme coïncide presque parfaitement avec une distribution normale.

---

<sup>4</sup>Moindres Carrées Ordinaires

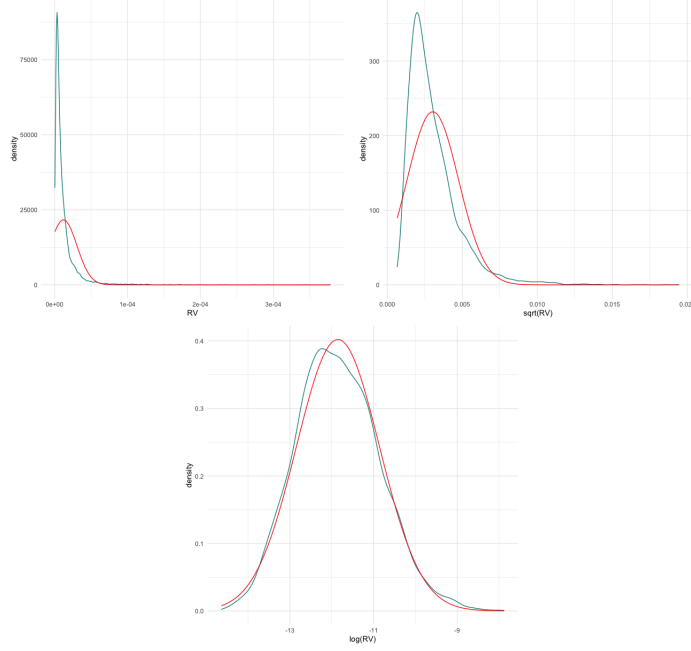


Figure 2: Distribution de la RV

### 3 Estimations

Nous pouvons désormais estimer nos différents modèles. Comme précisé plus haut, l'estimation du modèle HAR se fera par la méthode des MCO. Nous construisons un modèle pour expliquer la RV à différents horizons; 1 jour, une semaine et un mois. Les valeurs hebdomadaires et mensuelles étant des moyennes arithmétiques des 5 et 22 jours précédents. Le modèle HAR s'articule de la manière suivante:

$$RV_t = \beta_0 + \beta_D RV_{t-1}^D + \beta_W RV_{t-1}^W + \beta_M RV_{t-1}^M \quad (1)$$

Avec,

$$RV^W = \frac{1}{5} \sum_{i=1}^5 RV_{t-i}, RV^M = \frac{1}{22} \sum_{i=1}^{22} RV_{t-i} \quad (2)$$

Nous présentons les résultats dans le tableau suivant:

Série	RV			$RV^{1/2}$			$\log(RV)$		
	1	5	22	1	5	22	1	5	22
Coefficient									
$\beta_0$	$6.4e^{-07}(**)$	$8.2e^{-07}(***)$	$1.4e^{-06}(***)$	$1.3e^{-04}(***)$	$2.2e^{-04}(***)$	$3.6e^{-04}(***)$	$-0.6^{(***)}$	$-0.9^{(***)}$	$-1.6^{(***)}$
$\beta_D$	$1.2e^{-01}(***)$	$8.2e^{-02}(***)$	$1.9e^{-01}(***)$	$1.7e^{-01}(***)$	$1.1e^{-01}(***)$	$2.2e^{-01}(***)$	$0.2^{(***)}$	$0.1^{(***)}$	$0.2^{(***)}$
$\beta_W$	$1.9e^{-01}(***)$	$9.8e^{-02}(***)$	$3.0e^{-01}(***)$	$2.5e^{-01}(***)$	$5.2e^{-01}(***)$	$2.2e^{-01}(***)$	$0.3^{(***)}$	$0.5^{(***)}$	$0.2^{(***)}$
$\beta_M$	$6.4e^{-01}(***)$	$7.5e^{-01}(***)$	$4.06e^{-01}(***)$	$5.2e^{-01}(***)$	$2.9e^{-01}(***)$	$4.6e^{-01}(***)$	$0.4^{(***)}$	$0.3^{(***)}$	$0.4^{(***)}$
$\beta_J$	$-7.0e^{-04}(***)$	$-2.0e^{-04}(.)$	$-3.0e^{-04}(**)$	$-3.6e^{-03}(***)$	$4.7e^{-03}(***)$	$-2.2e^{-03}(***)$	$-22.5^{(*)}$	$-10.7^{(}$	$-5.2^{(***)}$

Table 2: Résultats des estimations

Nous pouvons remarquer que la majorité des coefficients sont significatifs. Les coefficients associés aux RV journalière, hebdomadaire et mensuelle sont tous significatifs à 0.1%, et ce, quelque soit l'horizon considéré. La constante est également significative à 0.1% à tous les horizons sauf pour la RV brute à un horizon de 1 jour, où le coefficient associé à la constante est significatif à 1%. Le coefficient associé aux jumps est significatif dans la plupart des cas. Pour la RV hebdomadaire brute, le jump n'est significatif qu'à 5% et n'est pas significatif pour la RV hebdomadaire transformée en log. D'une manière générale la significativité des jumps est moins avérée dès lors que la série a été passée en logarithme. D'autre part, ces jumps semblent avoir un impact moins important dans la détermination de la RV hebdomadaire comparativement à la RV quotidienne ou mensuelle.

Comme nous l'avons mentionné plus haut, le modèle HAR est estimé par MCO. Il convient donc de vérifier que les estimateurs sont non biaisés et efficaces. Pour cela, nous devons nous assurer que les conditions de Gauss-Markov sont respectées. Les résidus doivent être indépendants normaux, de moyenne nulle et de variance constante. Plusieurs tests statistiques permettent de vérifier le respect des hypothèses de Gauss-Markov; le test de Jarques-Bera pour vérifier la normalité des résidus, le test de Durbin-Watson pour l'auto-corrélation à l'ordre 1 et le test de Breusch-Pagan pour l'hétéroscédasticité. L'ensemble de ces tests seront appliqués aux résidus des modèles et nous présenterons également des graphiques qui permettent de visuellement diagnostiquer les résidus. Ces graphiques et les résultats des tests sont présentés ci-dessous.

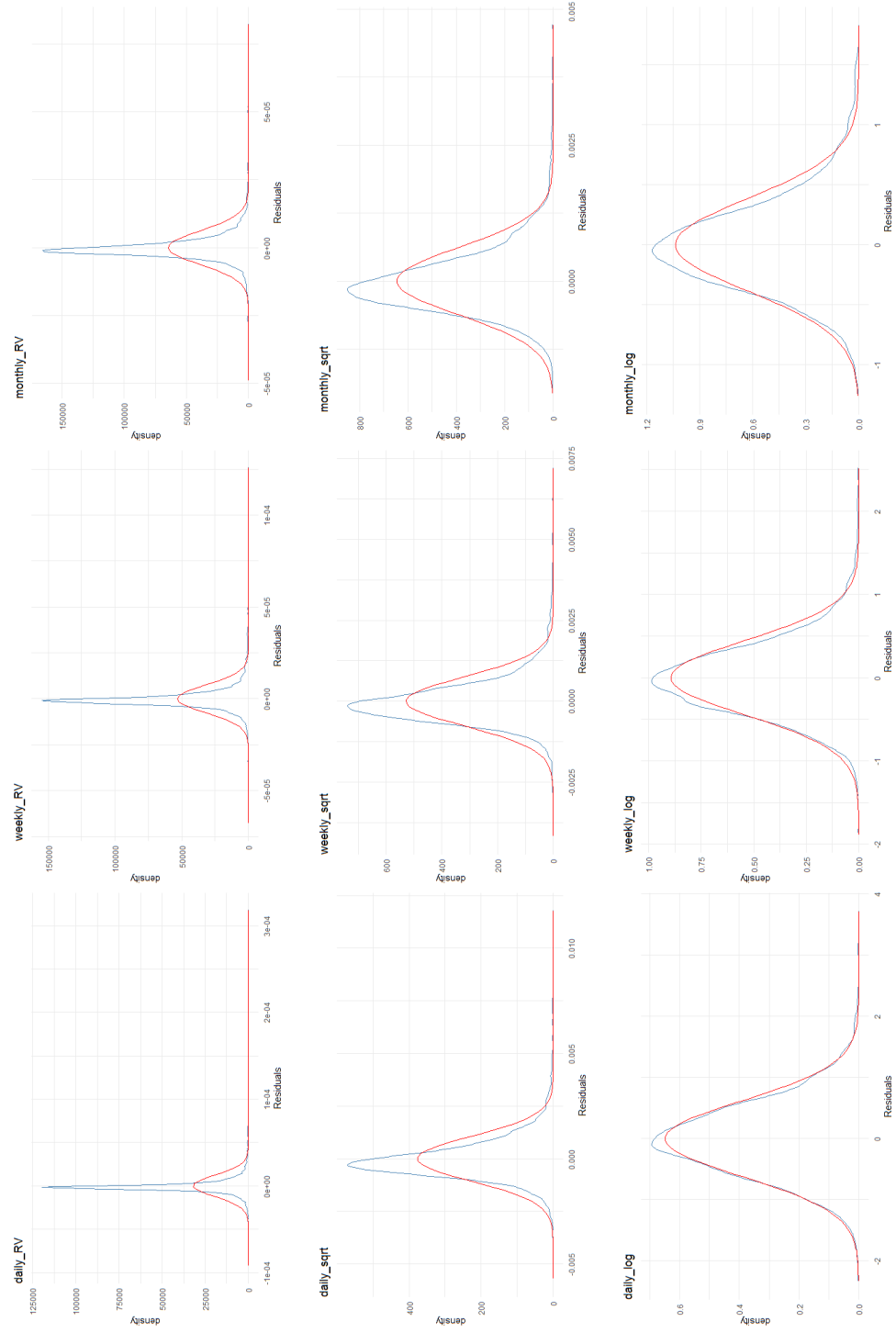


Figure 3: Distribution des résidus



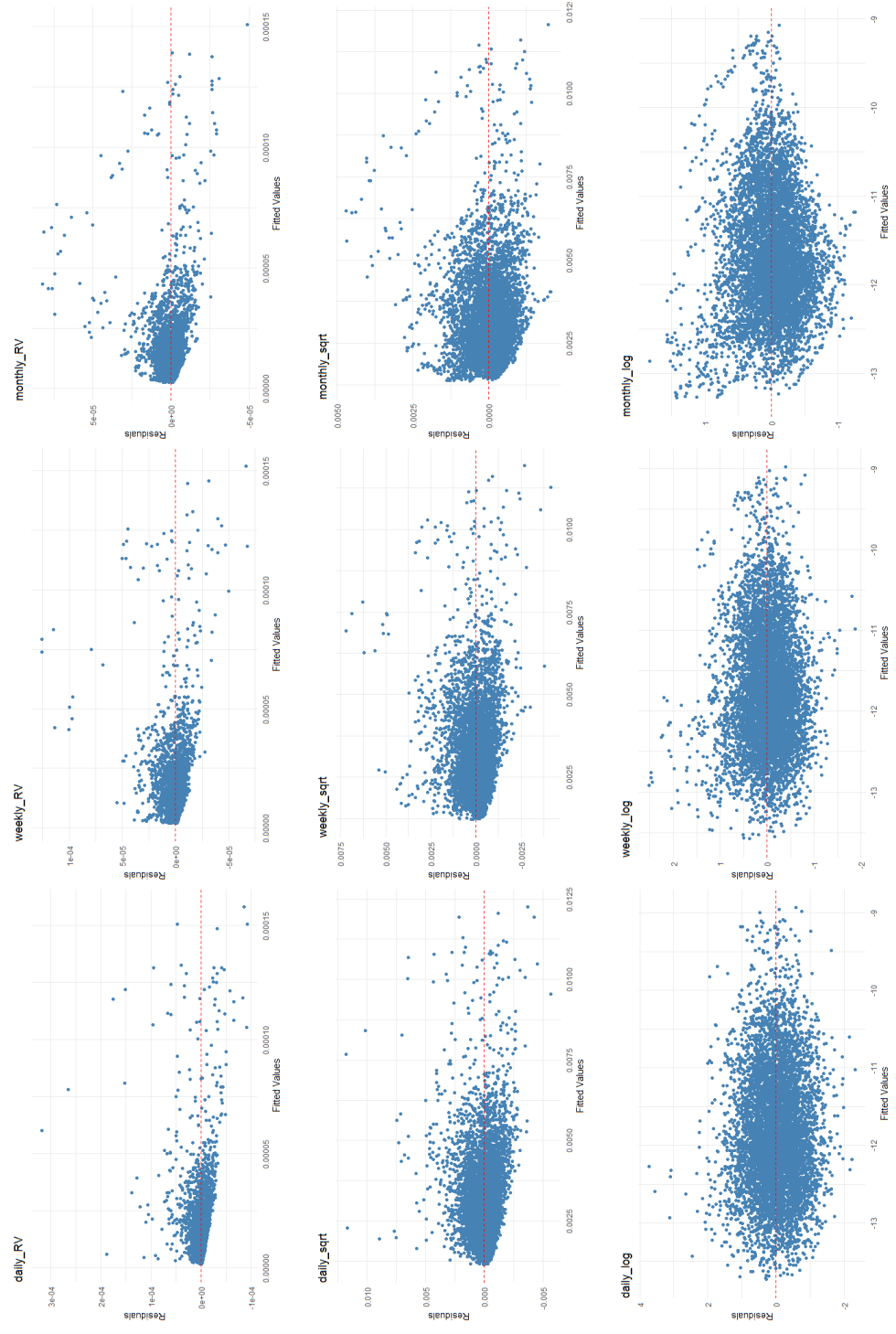


Figure 4: Résidus

Les graphiques ci-dessus montrent que la distribution des résidus des modèles estimés sur les variables transformées est plus proche de la distribution normale que celle des résidus estimés sur la série brute. Les résidus des modèles estimés sur la série brute sont très leptokurtiques et légèrement asymétriques. La transformation des données ne parvient pas à débarrasser les résidus de leur asymétrie, en revanche le kurtosis diminue largement lorsque l'on estime des modèles sur la racine carrée de la RV. Le coefficient d'aplatissement diminue d'autant plus lorsque l'on passe au logarithme de la RV. Avec cette dernière transformation, les résidus semblent (visuellement) être distribués normalement. L'horizon ne semble pas impacter drastiquement la distribution des résidus, les fonctions de densité sont semblables, peu importe que l'on travaille avec la RV quotidienne, hebdomadaire ou mensuelle.

Les graphiques des résidus contre les valeurs prédites nous donnent des indications quant à l'hétéroscédasticité de ces derniers. On peut distinguer une faible relation négative entre ces deux grandeurs pour la série brute. Les résidus semblent être corrélés négativement avec les valeurs prédites. Cette relation disparaît dès lors que l'on utilise une série transformée. D'une manière générale, à part pour la série brute, les résidus paraissent homoscedastiques. Pour confirmer ces intuitions, nous procédons aux trois tests statistiques cités plus haut. nous reportons les p-values pour chacun d'entre eux dans le tableau ci-dessous.

Série	RV			$RV^{1/2}$			$\log(RV)$		
Test	1	5	22	1	5	22	1	5	22
Jarque-Bera	0	0	0	0	0	0	0	0	0
Durbin-Watson	0.27	0	0	0.22	0	0	0.14	0	0
Breusch-Pagan	0	0	0	0	0	0	0.002	0	0

Table 3: Diagnostics des résidus

L'hypothèse  $H_0$  de normalité du test de Jarque-Bera est rejetée à 0.01% pour tous les horizons, pour toutes les séries. Les résidus ne sont pas distribués normalement, bien qu'ils semblaient l'être visuellement, dès lors que l'on utilisait une série transformée. Il en va de même pour le test de Breusch-Pagan, l'hypothèse  $H_0$  d'homoscedasticité des résidus est largement rejetée. La variance des résidus n'est pas constante. Ici encore, nos hypothèses basées sur les graphiques ci-dessus semblent fausses. Enfin l'hypothèse  $H_0$  de non-corrélation des résidus entre eux est rejetée pour les horizons hebdomadaire et mensuel, pour toutes les séries. Pour la RV journalière, nous ne pouvons pas rejeter l'hypothèse  $H_0$  au seuil de 10%. Contrairement aux observations faites plus haut, les résidus ne respectent pas les conditions de Gauss-Markov, il est donc fortement probable que les coefficients soient non-efficaces. Puisqu'une transformation des données ne permet pas de retrouver ces conditions, nous ne pourrions pas obtenir de résidus i.i.d normaux, homoscedastiques. Nous devons donc garder à l'esprit que les coefficients sont non-efficaces. Bien qu'allant à l'encontre des observations faites plus haut, nous n'avons pas d'autre choix que de nous en remettre aux tests statistiques présentés ici.

La prochaine étape consiste à éliminer les modèles qui ont une moins bonne performance afin de ne conserver que le meilleur d'entre eux. La comparaison des modèles se fera à l'aide de l'erreur quadratique moyenne (MSE) et du pourcentage de variance expliquée ( $R^2$ ). A l'instar de D.Buncic, nous verrons également comment l'inclusion des jumps impacte la qualité des modèles<sup>5</sup>, afin de savoir s'il est pertinent de les conserver dans le modèle.

Série	RV		
Métrique	1	5	22
$R^2$	0.540	0.757	0.808
MSE	$1.554e^{-10}$	$5.585e^{-11}$	$3.804e^{-11}$

Série	$RV^{1/2}$		
Métrique	1	5	22
$R^2$	0.621	0.759	0.817
MSE	$1.838e^{-11}$	$3.834e^{-12}$	$1.353e^{-12}$

Série	$\log(RV)$		
Métrique	1	5	22
$R^2$	0.616	0.733	0.762
MSE	$1.654e^{-10}$	$6.965e^{-11}$	$4.498e^{-11}$

Table 4: Qualité d'ajustement des modèles avec les jumps

Quelle que soit la transformation considérée, les modèles avec un horizon mensuel sont constamment mieux ajustés que les modèles avec un horizon hebdomadaire ou journalier. Quel que soit l'horizon considéré, l'estimation de la volatilité réalisée ( $\sqrt{RV}$ ) est la mieux ajustée, les  $R^2$  sont systématiquement plus importants et les MSE, systématiquement inférieures. D'une manière générale, le meilleur modèle est donc celui avec un horizon mensuel, estimé sur des données transformées par une racine carrée. Vis-à-vis de l'horizon journalier, le meilleur modèle est donc celui sur les données transformées en racine, suivi par le modèle sur les données transformées en logarithme et enfin le modèle sur les données brutes.

Nous effectuons les mêmes démarches en retirant la composante jumps des estimations.

Le retrait des jumps entraîne une très légère augmentation des MSE et une diminution du  $R^2$  pour certains horizons. En revanche, les changements sont négligeables. Dans notre cas, les jumps semblent améliorer le modèle mais la précision supplémentaire est infinitésimale. Les diagnostics des résidus sans

<sup>5</sup>The role of jumps and leverage in forecasting volatility in international equity markets - D.Buncic - 2017

Série	RV		
Métrique	1	5	22
$R^2$	0.539	0.757	0.808
MSE	$1.556e^{-10}$	$5.587e^{-11}$	$3.809e^{-11}$

Série	$RV^{1/2}$		
Métrique	1	5	22
$R^2$	0.621	0.757	0.817
MSE	$1.844e^{-11}$	$3.872e^{-12}$	$1.358e^{-12}$

Série	$\log(RV)$		
Métrique	1	5	22
$R^2$	0.616	0.733	0.762
MSE	$1.662e^{-10}$	$6.986e^{-11}$	$4.504e^{-11}$

Table 5: Qualité d’ajustement des modèles sans les jumps

les jumps donnent les mêmes résultats; aucun modèle ne respecte les conditions de Gauss-Markov.

## 4 Prédictions

Nous allons désormais vérifier que les observations faites plus haute se maintiennent out-of-sample. Nous allons réaliser des prédictions à un pas avec des modèles estimés sur une fenêtre roulante de 1000 observations. Nous utiliserons la RV quotidienne, bien que la RV mensuelle semble être d’avantage pertinente aux vues des qualités d’ajustement reportées dans la section précédente. Nous obtiendrons donc trois ensembles de prédictions différents, un pour chaque transformation appliquée aux données. Nous pourrons ensuite comparer les MSE pour chacune des transformations. Vis-à-vis des résultats précédents, nous nous attendons à ce que la volatilité réalisée soit la série la mieux prédite.

	RV	$\sqrt{RV}$	$\log RV$
MSE	$1.567e^{-10}$	$1.562e^{-10}$	$1.635e^{-10}$

Table 6: Qualité de prédiction

Nous pouvons constater que out-of-sample, la transformation en racine carrée permet une nouvelle fois d’obtenir les meilleurs résultats

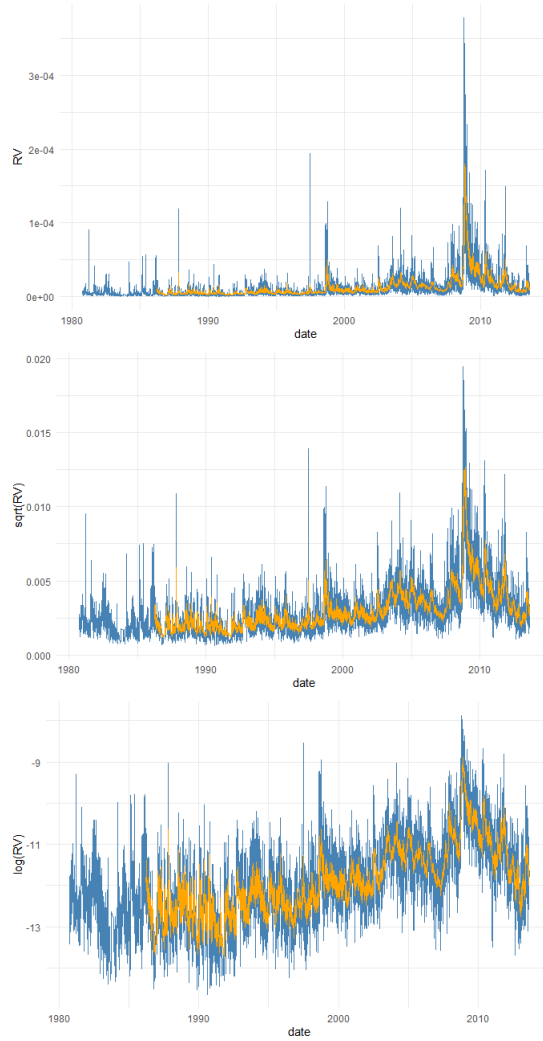


Figure 5: Prédications à un pas

## 5 Conclusions

Nous avons tenté d'estimer et de prédire la volatilité du taux de change USD/CAD à l'aide de différents modèles HAR. Les modèles ont été estimés à différents horizons (1, 5 et 22 jours) et sur des données ayant subi diverses transformations (racine carrée, logarithme naturel) dans le but de les normaliser. Nous avons constaté que les transformations aidaient grandement à obtenir une distribution de la variance réalisée à peu près normale et ce pour tous les horizons. Après avoir estimé tous les modèles, nous avons observé que les résidus ne respectaient

pas les conditions de Gauss-Markov, il est alors impossible de prouver que les coefficients obtenus sont BLUEs. L'ajout des jumps dans le modèle augmente très légèrement la précision, pour tous les horizons et toutes les transformations mais cette augmentation est aisément négligeable. Il semble cependant que la volatilité réalisée offre les meilleurs résultats avec le  $R^2$  le plus important ainsi que la MSE la plus faible, pour tous les horizons. En revanche l'horizon mensuel semble être le plus fiable, pour toutes les transformations. Nous avons ensuite réalisé des prévisions à un pas avec des modèles ré-estimés sur une fenêtre roulante de 1000 observations. A la suite de ces prévisions il s'avère que la racine carrée est la transformation qui permette d'obtenir les meilleurs résultats. Bien que les résidus ne soient pas i.i.d normaux, nous obtenons des qualités d'ajustement satisfaisantes, ce qui nous laisse à penser que le biais introduit par l'auto-corrélation et la non-normalité des résidus n'est pas particulièrement problématique. L'objectif de ce genre de modèle étant principalement prédictif, nous accorderons généralement moins d'importance au biais des coefficients qu'à leur variance. Il n'est donc pas dramatique d'obtenir des coefficients biaisés dès lors que les prédictions sont d'une qualité suffisante.