

# AWALEE NOTES



Utilisation de modèles ARIMA saisonniers  
pour la prédiction des flux de change

Jules Chabran



**Nombreuses** sont les banques qui optent pour la mise en place d'algorithmes destinés à la gestion automatique de leurs activités de market making dès lors que la rentabilité par trade pour la classe d'actifs considérée ne permet plus la présence de personnes dédiées (nominal par trade non significatif et spread trop faible) ou que le volume de requêtes clients est trop élevé pour être géré à l'échelle humaine. Le trader algorithmique prend alors la responsabilité d'émettre, par l'intermédiaire de son algorithme, des bids-offers qu'il souhaite les plus compétitifs possible afin d'accroître ses parts de marché et ainsi satisfaire les demandes de ses clients. Il propose par ailleurs un nouveau service intégrable sur des plateformes digitales, pouvant ainsi toucher une audience plus large.

Economiquement, pour répondre à ces besoins tout en conservant un niveau suffisant de rentabilité, le trader algorithmique peut analyser et prédire les flux intra-journaliers globaux (entre dealers ou venant de ses clients) par l'intermédiaire de techniques statistiques (séries temporelles dans notre analyse mais aussi plus généralement techniques d'apprentissage supervisé / machine learning), à l'image des modèles de centres d'appels optimisant la gestion des flux client au cours d'une journée. Il peut ainsi se permettre d'être plus compétitif et réduire son spread lorsque les signaux prédictifs lui indiquent des flux de trades importants à venir, lui offrant donc la possibilité de déboucler ses positions plus rapidement. À l'inverse, il se devra d'être plus conservateur si ces mêmes indicateurs anticipent une baisse des flux, l'obligeant le cas échéant à conserver ses positions sur une période de temps plus importante ou potentiellement traiter à perte sur le marché interbancaire. S'il parvient à extraire des indicateurs performants de flux à venir, le trader algorithmique peut alors lisser au cours de la journée son flux de trades et ainsi permettre une réduction de la variance de son P&L.

Notre étude se focalise sur le marché des changes, avec pour objectif de construire un modèle prédictif de flux intra-journaliers.

## CARACTÉRISTIQUES DU MARCHÉ ET MODÉLISATION

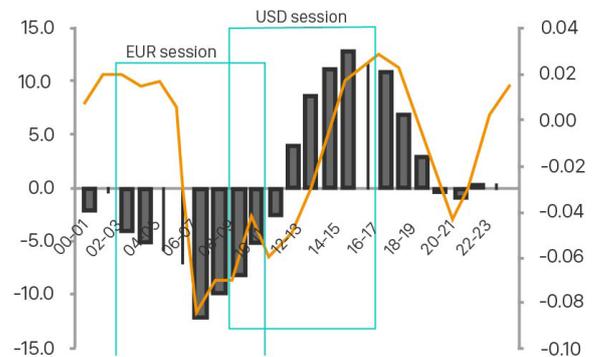
### A. Saisonnalité

Le marché des changes est ouvert en continu durant la semaine. Néanmoins, les flux de change associés sont impactés par les ouvertures globales des marchés asiatiques (Tokyo), américains (New York) et européens (Londres).

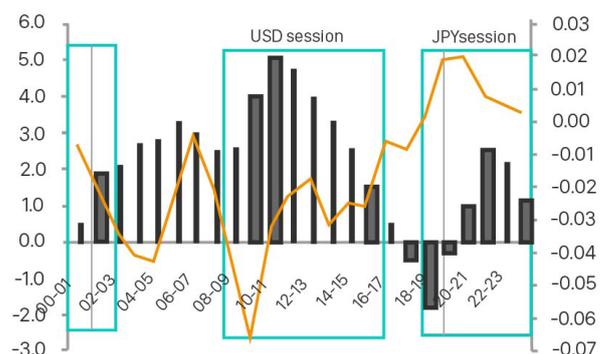
En se plaçant dans un référentiel parisien, le marché global (incluant tous les actifs financiers) ouvre à 09h00 heure locale et se traduit par un premier pic en terme de volumes qui perdure jusqu'aux alentours de 11h00. Il s'en suit une période de creux jusqu'à l'ouverture des marchés US à 14h00. Ces phénomènes intra-journaliers se traduisent par une cyclicité dans l'arrivée des trades dans le book algorithmique. Cette cyclicité à l'échelle du book algorithmique reflète la saisonnalité intra-journalière des flux d'ordres de l'ensemble du marché FX mentionnée dans [Intraday patterns in FX returns and order flow](#) de Francis Breedon et Angelo Ranaldo (2013) avec une illustration graphique du phénomène décrit pour les principales paires de devises.

Les graphiques ci-dessous mettent en évidence l'évolution des flux d'ordres et des rendements en moyenne quotidienne par paire de devises. Les colonnes montrent les flux d'ordres cumulatifs, ie le nombre d'ordres d'achat moins le nombre d'ordres de vente (échelle de gauche). La ligne présente les log-rendements annualisés cumulatifs (échelle de droite). Les colonnes en gras indiquent que les flux d'ordre horaires sont significativement différents de zéro au seuil de 5% selon le test de Student.

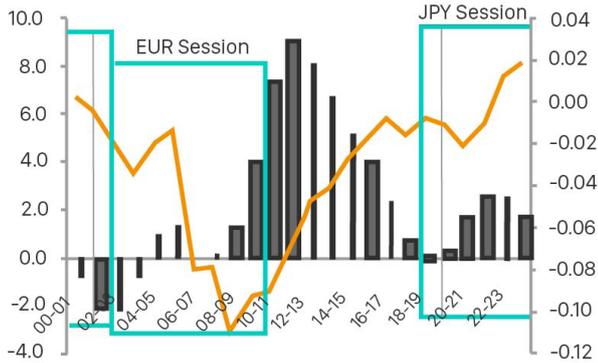
Cumulative order flow and returns on an average day (New York time) **EUR/USD**



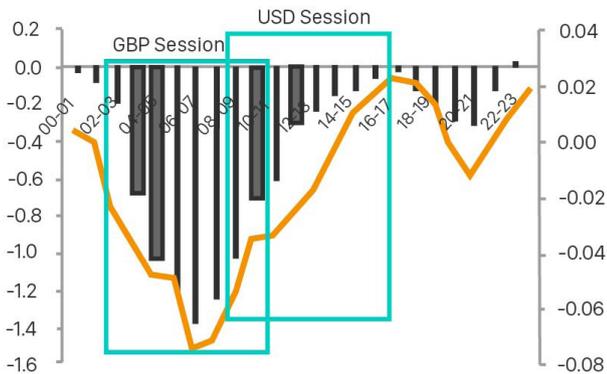
Cumulative order flow and returns on an average day (New York time) **USD/JPY**



Cumulative order flow and returns on an average day (New York time) **EUR/JPY**



Cumulative order flow and returns on an average day (New York time) **GBP/USD**



Toute modélisation du flux intra-journalier doit donc couvrir le phénomène de saisonnalité décrit et observé ci-dessus.

Notre book algorithmique exhibe des flux d'ordres similaires à ceux observés ci-dessus pour la paire EUR/USD avec une saisonnalité en M qui couvre près de 70% des flux globaux du book algorithmique.

### B. Modélisation du processus de comptage et définition de la série temporelle à prédire

L'objectif retenu est de modéliser le nombre de trades entrant dans un book algorithmique au cours de la journée. Cette hypothèse peut paraître réductrice, car elle ne prend par exemple pas en compte la pondération associée au notionnel de chaque trade. Elle contient néanmoins un niveau d'information pertinent, la série sous-jacente pouvant être modélisée par un processus de comptage (Poisson) permettant une analyse statistique des temps d'inter-arrivée (niveau de granularité de l'information le plus fin pour un processus de comptage). Cette modélisation s'inspire fortement de celle adoptée pour les flux enregistrés dans les centres d'appels et décrite dans Bayesian Forecasting of an Inhomogeneous Poisson Process with Applications to Call Center Data (Weinberg et al., 2007).

Elle va néanmoins différer dans la suite car notre analyse va appliquer un modèle SARIMA sur les paramètres estimés du processus de Poisson inhomogène. D'un point de vue financier, les hypothèses fondamentales du modèle considéré traduisent les phénomènes économiques résumés ci-dessous :

- Sur des intervalles de temps très court, on suppose que les clients traitent de manière aléatoire (on néglige donc ici l'effet de corrélation de trades à la milliseconde pouvant être généré par du trading propriétaire côté client).

On veut tester cette hypothèse en analysant la distribution des temps d'arrivée entre trades (le nombre de trades sur cet intervalle de temps ne représentant qu'une seule observation, il est nécessaire d'affiner le niveau de granularité des observations en étudiant les temps d'inter-arrivée). On peut ainsi modéliser le processus de comptage sur ces intervalles de temps comme un processus de comptage aléatoire à intensité constante  $\lambda$ . En effet, si le processus de comptage  $N_t$  est un processus de Poisson d'intensité constante  $\lambda$ , alors le processus des temps d'inter-arrivée  $S_k$  (contenant  $N_{t-1}$  observations) suit une loi exponentielle :

$$T_n = \inf(t > 0, N_t \geq n)$$

$$S_k = T_k - T_{k-1}$$

$$P(S_k < t) = 1 - \exp(-\lambda t)$$

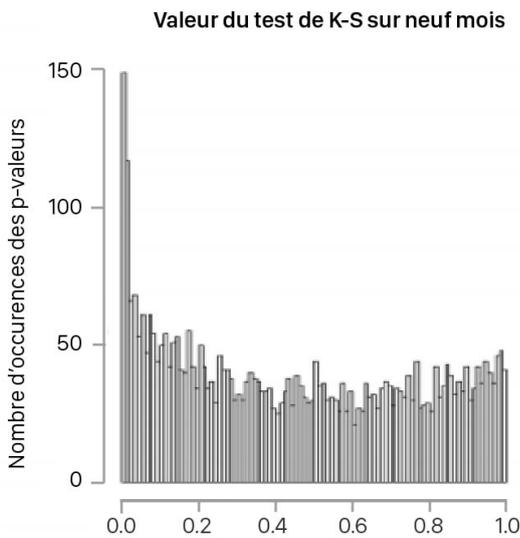
- Le jeu de données utilisé prend en compte les temps d'inter-arrivée de trades FX au cours de l'année 2011-2012.
- Sous l'hypothèse nulle, les temps d'inter-arrivée des trades suivent une loi exponentielle de paramètre  $\lambda$ . Les p-valeurs des tests de Kolmogoroff-Smirnoff sont ainsi enregistrées sur l'ensemble des intervalles de temps disponibles pour une période de neuf mois à raison de 22 statistiques de test journalières : on dispose exactement de 4202 observations de p-valeurs.

Sous l'hypothèse nulle, chaque p-valeur associée à la statistique de test de Kolmogoroff-Smirnoff suit une loi uniforme sur l'intervalle  $\{0;1\}$  :

- Soit  $S$  la statistique de test de Kolmogoroff-Smirnoff
  - p - valeur =  $P(Loi_{KS,\lambda} \leq S)$  avec  $S$  suivant la  $Loi_{KS,\lambda}$  sous l'hypothèse nulle
  - Ceci implique sous l'hypothèse nulle que :
- p-valeur  $\rightarrow Loi_{uniforme \{0,1\}}$

- En appliquant le test d'adéquation à la loi uniforme des 4202 p-valeurs calculées et issues de la statistique de test de Kolmogoroff-Smirnoff pour chaque intervalle de trade, on ne rejette pas l'hypothèse nulle de loi uniforme.

Qualitativement, ce non-rejet se matérialise par la répartition quasi-uniforme des occurrences des p-valeurs observées ci-dessous :



- Sur un intervalle de temps de trente minutes, on en déduit que le test d'adéquation de Kolmogoroff-Smirnoff appliqué au temps d'inter-arrivée ne rejette pas l'hypothèse nulle d'adéquation à la loi exponentielle de paramètre estimé  $\lambda$ .

- Sur des intervalles de temps supérieurs à trente minutes, le test d'adéquation est rejeté et la distribution des p-valeurs n'est plus uniforme, justifiant à posteriori l'agrégation des données sur trente minutes.

- Ce découpage par tranche de trente minutes est par ailleurs assez commun dans la littérature scientifique associée au marché des changes FX ajusté sur des données big data (tick des trades FX), on le retrouve notamment dans Modeling High-Frequency FX Data Dynamics (Jorda & Marcellino, 2003).

- Le processus de comptage associé au nombre de trades est supposé à accroissements dépendants entre intervalles de temps successifs (le lambda de la période des 30 minutes qui précède a potentiellement un pouvoir prédictif sur celui des trente minutes actuelles). Le modèle prédictif développé ci-après a précisément pour but de modéliser cette hypothèse de dépendance temporelle. Celle-ci peut avoir des raisons multiples comme par exemple des annonces économiques entraînant une augmentation des volumes de trades sur des intervalles de temps significativement supérieur à 30 minutes ou encore la persistance de périodes de faible activité sur une durée prolongée.

La série temporelle à prédire correspond donc à la série des  $\lambda_t$  estimés sur des intervalles de 30 minutes (comptage évalué de 07 :30 à 18 :30).

- La série temporelle à prédire correspond donc à la série des  $\lambda_t$  estimés sur des intervalles de 30 minutes (comptage évalué de 07 :30 à 18 :30). L'estimateur des  $\lambda_t$  utilisé est celui du maximum de vraisemblance justifié par le test d'adéquation de Kolmogoroff-Smirnoff évoqué ci-dessus. La période  $\{t_i; t_{i+1}\}$  est considérée:

- Soit  $N_{\{t_i; t_{i+1}\}}$  le nombre de trades sur  $\{t_i; t_{i+1}\}$
- $\forall t \in \{t_i; t_{i+1}\}, \lambda_t = \lambda_{\{t_i; t_{i+1}\}}$
- $T_k = \inf(t > 0, N_t \geq k)$
- $S_k = T_k - T_{k-1}$
- On dispose de  $(N_{\{t_i; t_{i+1}\}} - 1)$  tirages indépendants de temps d'inter-arrivée suivant une loi exponentielle de paramètre constant  $\lambda_{\{t_i; t_{i+1}\}}$  sur l'intervalle de temps  $\{t_i; t_{i+1}\}$

Après dérivation de la vraisemblance (non conditionnelle, on est ici en présence de tirages indépendants), l'estimateur du maximum de vraisemblance issu de l'échantillon ci-dessus s'avère être égal à l'inverse de la moyenne des temps d'inter-arrivée sur l'intervalle de temps  $\{t_i; t_{i+1}\}$  (L vraisemblance (likelihood) fonction des tirages  $S_1, S_2, \dots, S_m$  avec  $m = N_{\{t_i; t_{i+1}\}} - 1$ ) :

$$L(S_1, S_2, \dots, S_m) = \prod_{i=1}^m \lambda_{\{t_i; t_{i+1}\}} e^{-\lambda_{\{t_i; t_{i+1}\}} S_i}$$

$$L(S_1, S_2, \dots, S_m) = \lambda_{\{t_i; t_{i+1}\}}^m e^{-\lambda_{\{t_i; t_{i+1}\}} \sum_{i=1}^m S_i}$$

En maximisant la log vraisemblance (log(L) suivant la terminologie ci-dessus) on obtient facilement l'estimateur:

$$\forall t \in \{t_i; t_{i+1}\}, \hat{\lambda}_t = \widehat{\lambda_{\{t_i; t_{i+1}\}}} = \frac{1}{S_m}$$

- Les modèles de série temporelle retenus ajustés directement sur  $\hat{\lambda}_t$  peuvent retourner des valeurs négatives. Afin de rendre compatible notre série temporelle avec ces modèles, l'opérateur logarithme est appliqué.

### C. Choix du modèle SARIMA

#### 1. Ajustement de la saisonnalité

La série  $\lambda_t$  choisie dans le paragraphe B présente une forte saisonnalité intra-journalière s'expliquant par l'impact de l'ouverture asynchrone des marchés financiers.

Cet effet de saisonnalité est par ailleurs quantifiable du fait de la faible décroissance des autocorrélations journalières (période égale à 22 car le nombre de trades est évalué sur 11h par tranche de 30 minutes comme expliqué ci-après) :

## 2. Modélisation Log - SARIMA

Étant donné le volume significatif de données intra-journalières on se limite aux données collectées au cours des six derniers mois (novembre 2011 – mai 2012) : ce choix permet de diminuer le risque de changement de régime capturé par des données plus anciennes (anciens clients ne traitant plus avec la banque par exemple). Par ailleurs le processus de comptage commence chaque jour à 07h30 et se termine à 18h30, comptabilisant donc 22 observations par jour. Le modèle final choisi est déterminé en deux étapes :

La première étape consiste à présélectionner et estimer les paramètres des meilleurs modèles Log – SARIMA à partir d'un jeu de données in sample (optimisation des critères AIC / BIC / R<sup>2</sup> mais aussi choix des modèles dont les estimateurs de paramètre présentent à la fois un écart type faible et dont le test de nullité pour chaque paramètre est rejeté). L'ajustement en saisonnalité dans notre choix de modèle final ne prend en compte qu'une différenciation sur une période de saisonnalité (paramètres S(0,1,0) du modèle (S)ARIMA) : cela signifie donc que les modèles SARIMA choisis dans notre analyse sont en bijection avec des modèles ARIMA que l'on ajuste sur la série  $Y_t = \log(\hat{\lambda}_t) - \log(\hat{\lambda}_{t-24h})$ . Ce choix de modélisation se trouve validé dans notre analyse car la série des  $Y_t$  rejette l'hypothèse de racine unitaire lors du test de Phillips-Perron. L'analyse se concentre par la suite sur l'estimation des paramètres du modèle ARIMA sous-jacent (permettant ainsi l'utilisation de librairies R natives dont le temps de calcul est optimisé pour estimer les paramètres du modèle ARIMA).

La deuxième étape consiste à comparer les performances out of sample via deux métriques différentes (MAPE - mean absolute percentage error et RMSE – root mean square error). Ces performances sont évaluées chaque jour durant un mois, permettant ainsi de comparer les RMSEs et MAPEs des modèles présélectionnés:

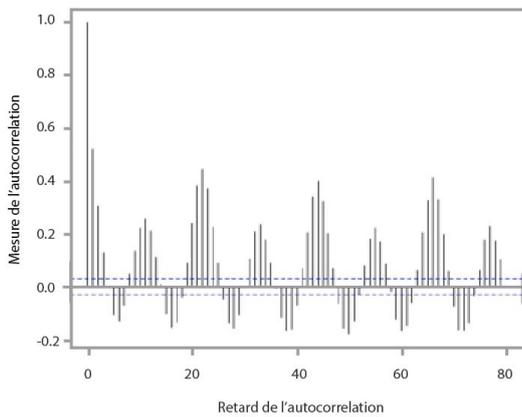
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i)^2}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|X_i - \hat{X}_i|}{X_i}$$

À l'issue de cette double filtration le modèle optimal choisi est défini ci-dessous, pour lequel tous les paramètres rejettent le test de nullité et présente un écart type non significatif:

$$(1 - \varphi_1 * L^1 - \varphi_2 * L^2 - \varphi_3 * L^3 - \varphi_4 * L^4) \cdot (1 - \varphi_{22} * L^{22}) \cdot Y_t = (1 + \beta_2 * L^2) \cdot \varepsilon_t$$

Autocorrellogramme à trois jours de la série des lambda



Cette saisonnalité intra-journalière peut être corrigée en appliquant l'opérateur différence sur la série  $\lambda_t$  entre deux journées. Ainsi, la série désaisonnalisée  $Y_t$  sera égale à :

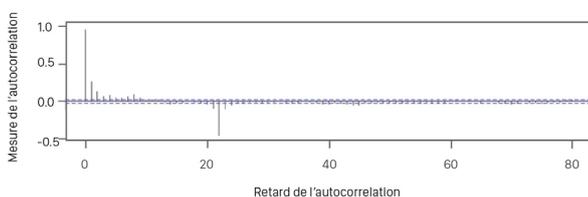
$$Y_t = \log(\hat{\lambda}_t) - \log(\hat{\lambda}_{t-24h})$$

Il est alors nécessaire de tester la stationnarité de la série  $Y_t$  en utilisant le test de racine unitaire de Phillips-Perron (hypothèse nulle de racine unitaire):

Hypothèse nulle	Racine unitaire
Statistique de PP	-3 218
P-valeur	0.01

La P-valeur étant de l'ordre de 0.01 on a de très fortes présomptions contre l'hypothèse nulle de racine unitaire que l'on rejète ici, et on accepte donc la stationnarisation de la série désaisonnalisée  $Y_t$ . L'analyse qualitative de l'auto-corrélogramme appliqué sur la série  $Y_t$  présente maintenant une décroissance exponentielle et confirme donc qualitativement le test de PP :

Autocorrellogramme à trois jours de la série stationnarisée des lambda



La différenciation opérée ci-dessus est caractéristique de la modélisation SARIMA (seasonal ARIMA) qui permet de corriger l'effet de saisonnalité sans pour autant estimer des paramètres supplémentaires (indicatrices statiques de saisonnalité intra-journalières rajoutant un niveau de complexité non nécessaire et coûteux dans la calibration du modèle de série temporelle finale).

Quelques commentaires résultent de ce modèle optimum choisi :

- Bien que la dessaisonnalisation soit effectuée en amont dans la série  $Y_t$ , le paramètre  $\varphi_{22}$  démontre qu'il existe encore une dépendance autorégressive saisonnière à un jour dans le modèle final choisi.

On note par ailleurs dans la partie autorégressive un pouvoir prédictif qui perdure sur les deux heures précédentes (montré par AR(4)).

- Les coefficients significatifs  $\varphi_i$  résultant correspondent bien aux pics significatifs perdurant dans l'autocorrélogramme de la série stationnarisée des lambda (cf. graphe ci-dessus): il est donc consistant avec nos observations qualitatives en amont de l'établissement du modèle SARIMA.

Afin de compléter notre analyse et valider à posteriori la modélisation sélectionnée ci-dessus, il reste à confirmer via le test de Portemanteau que les résidus  $\varepsilon_t$  correspondent bien à un bruit blanc. La statistique de teste de Portemanteau confirme bien cette hypothèse et ne rejette pas l'hypothèse nulle de blancheur des résidus:

Modèle	Hypothèse nulle	Statistique de test	P-valeur
$\lambda$ -SARIMA	Blancheur des résidus	0.2685	0.6045

Nous notons cependant que les tests associés à la normalité des résidus estimés se trouvent rejetés pour le modèle considéré, et limitent donc la portée statistique induite par l'optimisation de la vraisemblance dans le cadre ici d'un modèle SARIMA à résidus non normaux.

Cependant, le modèle ajusté surperforme sur des échantillons significatifs out of sample les indicateurs benchmark de flux (indicateurs classiques ayant pour seule hypothèse l'indépendance des flux désaisonnalisés).

Il en découle donc que la minimisation de la vraisemblance conditionnelle issue des modèles SARIMA à résidus normaux, équivalente numériquement à la minimisation des carrés des résidus, doit être interprétée ici comme la minimisation d'une métrique in sample permettant de capturer un pouvoir prédictif out of sample.

## CONCLUSION ET PERSPECTIVE FUTURE

L'étude menée sur les séries temporelles de flux clients dans cette note s'est focalisée dans un premier temps sur le choix d'agrégation des données de flux sur un intervalle de trente minutes. Ce choix de modélisation nous a permis de capturer les deux phénomènes suivants:

- Extraire des estimateurs statistiquement justifiés du paramètre  $\lambda$  (considéré constant sur ces intervalles de temps) et matérialisant la stabilité des temps d'inter-arrivés sur des échelles de temps plus fins.

- Capturer l'autocorrélation des estimateurs  $\lambda$  intra-journaliers pour ajuster un modèle prédictif prenant à la fois en compte la saisonnalité intra-journalière mais aussi l'interdépendance temporelle constatée qualitativement lors de l'étude de l'autocorrélogramme intra-journalier de la série des  $\lambda$ .

L'ajustement et la pertinence du modèle log SARIMA finalement estimé ont été matérialisés par les points ci-dessous :

- Consistance entre performances out of sample et métriques d'ajustement in sample.

- Validation de l'indépendance des résidus estimés du modèle grâce au test de Portemanteau, posant ainsi les limites quant à l'existence d'un modèle de prédiction univarié capturant plus d'auto-corrélations.

- La surperformance du modèle retenu face à des estimateurs benchmark, tout en reconnaissant que l'amplitude du bruit final évalué reste non négligeable (ordre de grandeur du MAPE égal à 30% contre des benchmarks compris entre 40% et 45%).

Cette étude s'est concentrée sur un modèle non vectoriel (à une seule dimension) de séries temporelles appliqué au flux de trades, néanmoins la portée de cette étude dépasse la simple analyse de flux clients en termes de compréhension des facteurs globaux dirigeant le marché FX. Une analyse plus détaillée des interdépendances entre séries de volumes et séries de volatilité intra-journalière (non développée dans cette note) fait ressortir une forte corrélation entre celles-ci. Les premiers tests effectués semblent démontrer qu'intégrer les séries de volatilité dans des modèles multidimensionnels de prédiction n'améliore pas significativement le modèle univarié développé ci-dessus. Des pistes de réflexion restent encore ouvertes sur ce point et nécessitent de plus amples tests (causalité, cointégration par exemple) afin de déterminer qui capture en premier dans le temps le pouvoir prédictif déterminant l'évolution des flux et niveaux de volatilité future.

## REFERENCE

Tsenga, F-M. & Tzeng , G-H., 2002.

A fuzzy seasonal ARIMA model for forecasting. Fuzzy Sets and Systems, Vol 126(3), pp 367-376. Jorda, Ò, &

Marcellino, M., 2003.

Modeling High-Frequency FX Data Dynamics. Macroeconomic Dynamics, Vol 7(4), 618-635.

Weinberg, J., Brown, L.D., & Stroud, J.R., 2007.

Bayesian Forecasting of an Inhomogeneous Poisson Process with Applications to Call Center Data. Journal of the American Statistical Association, Vol 102 (480), pp1185-1198.

Breedon, F. & Ronaldo, A., 2013.

Intraday patterns in FX returns and order flow. Journal of Money, Credit and Banking, Vol 45(5), pp 953-965.



### A PROPOS D'AWALEE

Awalee est un cabinet de conseil indépendant spécialiste du secteur de la Finance, créé en 2009 et qui compte plus de 70 collaborateurs.

Nous sommes en mesure à la fois d'adresser des sujets relatifs à l'expertise des métiers de la Finance (Consulting) et de conduire des projets d'organisation et de transformation (Advisory). Et nous le faisons grâce à la synergie agile de ces deux savoir-faire.

Nos expertises s'exercent dans la conformité réglementaire, la finance quantitative, la fonction finance, la gouvernance des outils & systèmes, le management des risques et les marchés financiers. Au-delà de ce que nous faisons, il y a comment nous le faisons : viser l'excellence et repousser nos limites tout en cultivant la convivialité et en favorisant l'esprit d'équipe.

Nous sommes Awalee : nous sommes AWARE & AWESOME.

Awalee consulting  
59 avenue Marceau  
75016 Paris

 [www.awaleeconsulting.com](http://www.awaleeconsulting.com)

 [twitter.com/awaleeconsulting](https://twitter.com/awaleeconsulting)

 [linkedin.com/awaleeconsulting](https://linkedin.com/awaleeconsulting)