



Les coûts cachés de la latence dans le calcul des cours de référence

Examen et analyse des opérations

Organisme canadien de réglementation du commerce des valeurs mobilières

9 mars 2016

Auteurs :

Lisa Anderson, chef de l'équipe de l'analytique, OCRCVM

Baiju Devani, directeur de l'analytique, OCRCVM

Yifan Zhang, scientifique des données, OCRCVM

Remerciements :

Nous remercions Corey Garriott et Adrian Walton de la Banque du Canada et Andreas Park de l'Université de Toronto pour leurs commentaires et suggestions. Toutes erreurs et omissions sont les nôtres.

1. Résumé des constatations

La présente étude explore l'ampleur de la latence dans le calcul des cours de référence sur les plateformes de négociation opaques au Canada et révèle que certains marchés affichent régulièrement une telle latence. Il en résulte des cours d'exécution périmés et des latences assez importantes pour qu'un nombre démesurément élevé d'opérations effectuées par des négociateurs à haute vitesse tirent profit de ces cours périmés. Même si l'impact économique de ces latences est faible, les participants peuvent ne pas se rendre compte que certains marchés affichent des latences qui se répercutent sur leurs coûts de négociation.

Nous constatons que 4 % des opérations invisibles exécutées au cours médian sont exécutées à un cours périmé, la latence moyenne globale étant estimée à 4,8 millisecondes. Les latences observées ne se limitent pas à un marché unique. Cependant, différents marchés présentent des variations tant du point de vue de la durée de la latence que de celui de son impact économique. Le coût moyen attribuable à cette inefficience est de 2 969 \$ par jour pour les titres composant l'indice S&P/TSX 60 (soit 748 188 \$ par an) sur la période récente et a déjà atteint 5 000 \$ par jour par le passé. Ces coûts absolus se concentrent sur deux marchés. En termes relatifs, on observe des coûts relatifs importants dus à la latence sur quatre marchés.

Si ces latences n'étaient pas assez importantes pour être exploitables, nous nous attendrions à ce que les coûts soient répartis équitablement entre les participants actifs et passifs. Cependant, nous constatons que les coûts de latence ne sont pas répartis équitablement, 88 % des opérations affectées par la latence favorisant le participant actif. De plus, dans toutes les opérations affectées par la latence qui avantagent le côté actif, les négociateurs à haute vitesse sont les bénéficiaires dominants et obtiennent 83 % de l'avantage total. Lorsqu'on calcule les coûts de latence des opérations exécutées à un cours périmé en pourcentage du coût prévu du demi-écart de toutes les opérations exécutées par un groupe de participants, nous constatons que le fait d'interagir avec un ordre actif saisi de façon invisible par un négociateur à haute vitesse dans le cadre d'une opération affectée par la latence coûte 13,1 % de plus que le coût prévu du demi-écart (coût de latence relatif). La majorité des opérations invisibles exécutées au cours médian ont lieu entre des négociateurs à basse vitesse et, par conséquent, les coûts de latence relatifs globaux ne représentent que 3,1 % du coût total prévu du demi-écart pour tous les participants.

L'étude nous éclaire aussi sur les sources de latence potentielles. Nous constatons que le nombre de messages d'ordres qui précède immédiatement une opération est corrélé avec la durée de la latence. En outre, le recours par les marchés aux services de co-implantation est associé à une réduction de la latence.

2. Introduction

Une faible latence, soit le calcul du meilleur cours acheteur et vendeur national en fonction des données provenant des marchés, est essentielle pour que les participants et les marchés puissent réagir à l'évolution rapide de la conjoncture. Cependant, les participants n'observent pas tous le meilleur cours acheteur et vendeur national le plus à jour. Le temps nécessaire pour qu'un participant puisse observer

le meilleur cours acheteur et vendeur national dépend de la technologie qu'il utilise pour recueillir et traiter les données et de sa proximité physique des autres bourses.

Le type d'arbitrage de latence que nous étudions est celui qui est basé sur les cours d'exécution des opérations invisibles. Le cours d'une opération invisible exécutée au cours médian est calculé d'après l'observation par le marché opaque du meilleur cours acheteur et vendeur national, qui se base sur les données fournies soit directement par les marchés visibles, soit par l'agence de traitement de l'information. En cas de retard dans l'obtention ou le traitement des données ou le calcul du cours d'exécution, cela peut entraîner une latence sur le marché opaque. Contrairement à l'arbitrage de latence traditionnel, qui est possible du fait que les données transmises par l'agence de traitement de l'information parviennent directement aux négociateurs à basse vitesse avec un certain retard, ce type d'arbitrage résulte des inefficacités du marché. La lenteur du marché a en outre un impact sur l'interaction entre les négociateurs à haute et à basse vitesse. Les négociateurs à haute vitesse qui peuvent observer le meilleur cours acheteur et vendeur national le plus récent (qui n'a pas encore été observé ou traité par un marché opaque) peuvent faire exécuter leurs ordres à un cours périmé avant que les cours évoluent en leur défaveur. Pour tirer profit de la latence, les négociateurs à haute vitesse doivent à la fois déterminer le meilleur cours acheteur et vendeur national plus rapidement que le marché et être conscients de la latence sur les marchés. Ces opérations à cours erroné peuvent avoir un impact significatif sur certains participants aux marchés opaques. Ce phénomène a été discuté dans le contexte des opérations effectuées par un client institutionnel unique sur les marchés opaques américains (Biais, Foucault et Moinas, 2014; Easley, Hendershott et Ramadorai, 2013; Foucault, Hombert et Rosu, à paraître) et signale des problèmes potentiels de latence systémique.

Afin d'étudier le phénomène de la latence dans le cadre des opérations invisibles exécutées au cours médian :

1. Nous déterminons la durée de la latence et l'ampleur de son impact économique.
2. Nous précisons les considérations économiques propres aux participants et examinons les preuves d'interaction systématique avec les inefficacités du marché associées à la latence dans le cadre des opérations invisibles exécutées au cours médian.
3. Nous analysons les facteurs potentiels à l'origine de cette latence, par exemple le recours aux services de co-implantation et le degré d'activité du marché avant une opération affectée par la latence.
4. Nous proposons une perspective historique de la durée de la latence et de son impact économique (sur une période de trois ans), et mettons en évidence les périodes marquées par une détérioration importante.

3. Détails relatifs aux données et aux institutions

Tous les marchés canadiens des titres de capitaux propres fournissent à l'OCRCVM des données de négociation dans un format standardisé, le protocole Financial Information eXchange (FIX), au moyen

d'un système de transmission réglementaire en temps réel. Ces données comprennent tous les messages de négociation, d'ordres et de cotation et contiennent tant des champs de données accessibles au public que des désignations réglementaires confidentielles. Les données réglementaires sont utilisées par le système de surveillance en temps réel de l'OCRCVM et sont également stockées dans l'Entrepôt de données des titres de capitaux propres aux fins de la surveillance et de l'analyse après les opérations.

En date du 15 octobre 2012, tous les messages d'événement sont horodatés à la milliseconde près. Un numéro séquentiel propre au marché leur est également attribué, ce qui permet de trier correctement les événements survenus sur un marché au cours d'un intervalle d'une milliseconde. Cependant, l'ordre des événements qui surviennent sur différents marchés au cours du même intervalle d'une milliseconde ne peut être déterminé.

La période d'échantillonnage va du 15 octobre 2012 au 31 août 2015. Nous procédons à une analyse historique de la latence sur les marchés opaques au cours de la période d'échantillonnage. Nous définissons un sous-échantillon comprenant les opérations effectuées en juillet et août 2015 pour déterminer les coûts économiques et les bénéficiaires de cette latence et faire la lumière sur ses sources potentielles. Il est possible de soumettre ce sous-échantillon à une analyse approfondie, parce que les opérations exécutées au cours médian peuvent être repérées de façon plus fiable sur la plupart des marchés à partir de juillet 2015 en raison de l'inclusion d'un nouveau champ du message de négociation dans l'Entrepôt de données des titres de capitaux propres.

Notre analyse porte sur les titres composant l'indice S&P/TSX 60, qui représentent 57,7 % de la valeur négociée de tous les titres compris dans le sous-échantillon d'opérations effectuées en juillet et août. Nous nous concentrons sur les opérations invisibles qui ont lieu entre 9 h 35 et 15 h 55. Les applications intentionnelles, les opérations visant des lots irréguliers et les opérations effectuées alors que le marché est « croisé » sont exclues. De plus, toutes les opérations résultant d'un ordre au cours du marché sont exclues car elles n'ont pas de côté actif.

Au Canada, les types d'ordres invisibles au cours médian offerts sur les marchés des titres de capitaux propres le sont sur les marchés entièrement opaques et sur les marchés opaques et visibles (se reporter au tableau 1). Pour les besoins de la présente étude, nous avons masqué l'identité des six marchés qui utilisent le cours médian et négocient les titres de la TSX (appelés M1 à M6).

4. Détermination des opérations affectées par la latence : méthodologie

i. Détermination des opérations invisibles exécutées au cours médian

Nous déterminons les opérations invisibles exécutées au cours médian en fonction des critères suivants. Pour la période historique, nous incluons les opérations dont le cours est un multiple impair d'un demi-cent¹. Cette méthode permet d'éliminer les opérations invisibles non exécutées au cours médian, au prix toutefois de l'omission de certaines opérations exécutées au cours médian dont le cours est un multiple

¹ Le format général d'un multiple impair d'un demi-cent est x,xx5 \$ (par exemple 1,235 \$).

pair d'un demi-cent. Le nombre d'opérations déterminées à l'aide de cette méthode représente la limite inférieure du nombre réel d'opérations exécutées au cours médian. Pour le sous-échantillon d'opérations effectuées en juillet et août, nous déterminons les opérations exécutées au cours médian sur certains marchés en utilisant un champ du message de négociation (décrit à la section 3). Cette méthode permet de déterminer plus précisément les opérations exécutées au cours médian car elle tient compte des opérations dont le cours est un multiple pair d'un demi-cent. Ces opérations forment le groupe d'opérations invisibles exécutées au cours médian que nous utilisons pour quantifier la latence.

ii. Détermination des opérations affectées par la latence

Parmi les opérations au cours médian déterminées ci-dessus, nous recherchons ensuite les opérations affectées par la latence. Les étapes logiques de cet exercice sont illustrées dans le graphique 1 et sont les suivantes :

Étape 1. Détermination des opérations exécutées à un cours périmé

En analysant toutes les opérations au cours médian déterminées précédemment, nous nous attendrions à ce que les côtés actif et passif de ces opérations « traversent » la moitié de l'écart. Nous désignons les opérations qui sont exécutées à des cours qui ne correspondent pas au point médian du meilleur cours acheteur et vendeur le plus récent comme des opérations exécutées à un cours périmé. Pour plus de détails, veuillez consulter l'annexe A. Nous classons les opérations exécutées à un cours périmé comme suit :

- Opérations avantageant le négociateur actif : Dans ce cas, le cours d'exécution périmé avantage le participant actif car le cours médian actuel a évolué en défaveur du négociateur actif. Celui-ci obtient un meilleur cours d'exécution que le cours médian actuel (l'acheteur achète à un prix inférieur au cours médian actuel, le vendeur vend à un prix supérieur au cours médian actuel).
- Opérations avantageant le négociateur passif : Scénario analogue au précédent, excepté que le cours d'exécution périmé avantage le participant passif car le cours médian actuel a évolué en défaveur du négociateur passif.
- Non concluantes : Nous ne trouvons aucune cotation qui explique le cours médian.
- Limitations liées à la granularité de l'horodatage : Lorsque des événements qui surviennent sur plusieurs marchés sont horodatés au cours du même intervalle d'une milliseconde (la plus petite tranche de temps pour laquelle des données réglementaires sont transmises à l'OCRCVM), nous considérons que cet intervalle est sujet à des limitations liées à la granularité de l'horodatage. Dans ces conditions, les événements qui surviennent sur plusieurs marchés ne peuvent faire l'objet d'un tri fiable.

Étape 2. Recherche de la cotation de référence correspondant au cours des opérations exécutées à un cours périmé

Nous recherchons un cours médian égal au cours de l'opération à l'intérieur de la plage de 0 à 100 millisecondes qui précède l'opération exécutée au moment $t_{\text{opération}}$. Nous désignons la plus

récente cotation dont le point médian est égal au cours de l'opération comme cotation de référence (au moment t_{ref}) (se reporter à l'illustration figurant dans le graphique 2). Par la suite :

- Si nous ne trouvons aucune cotation → nous mettons fin à la recherche et désignons l'opération comme non concluante.
- Si nous trouvons une cotation de référence et que $t_{opération} - t_{ref} < 1 \text{ ms}$ → nous désignons l'opération comme sujette à des limitations liées à la granularité de l'horodatage.
- Si nous trouvons une cotation de référence et que $t_{opération} - t_{ref} \geq 1 \text{ ms}$ → nous passons à l'étape 3.

Étape 3. Calcul de la latence

Nous recherchons dans le futur le premier cours médian ($t_{miseàjour}$) qui n'est pas égal à la cotation de référence établie à l'étape précédente (se reporter à l'illustration figurant dans le graphique 2). Plus précisément :

$$t_{ref} \leq t_{miseàjour} \leq t_{opération}$$

Par la suite,

- Si $t_{opération} - t_{miseàjour} < 1 \text{ ms}$ → nous désignons l'opération comme sujette à des limitations liées à la granularité de l'horodatage.
- Si $t_{opération} - t_{miseàjour} \geq 1 \text{ ms}$ → nous désignons l'opération comme opération exécutée à un cours périmé sans limitations liées à la granularité de l'horodatage, et calculons la latence comme suit :

$$Latence = t_{opération} - t_{miseàjour}^2 \quad (1)$$

Nous pouvons définir la latence réelle ($Latence_{réelle}$) comme le délai écoulé entre l'établissement du nouveau meilleur cours acheteur et vendeur national et son observation par le marché opaque (se reporter à l'illustration figurant dans le graphique 2). Nous ne pouvons toutefois observer cette latence réelle parce que nous ignorons combien de temps le marché opaque a continué d'utiliser le cours périmé après l'opération observée. Dans la présente étude, la *Latence* correspond la limite inférieure de la latence réelle. La relation entre les deux valeurs est $Latence_{réelle} \geq Latence$.

iii. Définition des coûts de latence

Le cours médian sert souvent de référence pour l'évaluation des coûts. Dans ce scénario, nous considérons que l'exécution d'une opération au cours médian entraîne un coût nul pour les deux participants. Pour les besoins de notre analyse, nous calculons le coût pour chaque participant par rapport au cours acheteur (vendeur) pour l'acheteur (le vendeur). Dans ce cas, l'exécution d'une

² Plusieurs mises à jour peuvent survenir après la cotation de référence. La première cotation qui met à jour la cotation de référence est appelée première mise à jour de cotation et la dernière cotation précédant l'opération est appelée dernière mise à jour de cotation. L'expression $t_{miseàjour}$ employée ici désigne le moment de la première mise à jour de cotation.

opération au cours médian entraîne un coût égal au demi-écart pour les deux participants. Le coût total d'une opération exécutée au cours médian pour l'acheteur est calculé comme suit :

$$Coût_{total_{it}} = (C_{it} - acheteur_{it}) \times volume_{it} \quad (2)$$

C_{it} est le cours de la t^e opération sur l'action i (qui peut se baser sur un cours médian de référence périmé). $acheteur_{it}$ est le cours acheteur actuel de la t^e opération sur l'action i . $volume_{it}$ est le volume de négociation de la t^e opération sur l'action i . Dans les situations normales (non affectées par la latence), le coût total d'une opération exécutée au cours médian est égal à la moitié de l'écart acheteur-vendeur. Pour calculer le coût total pour le vendeur, il faut soustraire le cours du cours vendeur actuel.

Le coût total peut être décomposé en deux parties : le coût prévu et le coût de latence. La relation entre ces deux composantes est la suivante :

$$Coût_{total_{it}} = Coût_{prévu_{it}} + Coût_{delatence_{it}} \quad (3)$$

Le coût prévu d'une opération exécutée au cours médian pour l'acheteur est calculé comme suit :

$$Coût_{prévu_{it}} = (M_{it} - acheteur_{it}) \times volume_{it} \quad (4)$$

M_{it} est le cours médian actuel de la t^e opération sur l'action i . Ce coût est toujours égal à la moitié de l'écart actuel. Pour calculer le coût prévu pour le vendeur, il faut soustraire le cours médian actuel du cours vendeur actuel.

Le coût de latence pour un acheteur est calculé comme suit :

$$Coût_{delatence_{it}} = (C_{it} - M_{it}) \times volume_{it} \quad (5)$$

Pour calculer le coût de latence pour le vendeur, il faut soustraire le cours actuel du cours médian actuel. Dans des situations normales (non affectées par la latence), C_{it} est égal à M_{it} , ce qui se traduit par un coût de latence nul. Un coût de latence négatif est interprété comme un avantage. Lorsque l'acheteur supporte un coût, le vendeur obtient un avantage, et vice-versa.

Nous nous intéressons au pourcentage des coûts de latence par rapport aux coûts prévus liés à l'ensemble des opérations exécutées au cours médian. Ce pourcentage est calculé comme suit :

$$Coût_{delatence_{relatif}} = \frac{\sum Coût_{delatence_{it}}}{\sum Coût_{prévu_{it}}} \quad (6)$$

Le $Coût_{delatence_{relatif}}$ peut être interprété comme le coût supplémentaire que les négociateurs affectés par la latence paient par suite des cours médians périmés, par rapport au coût qu'ils étaient disposés à payer ou censés payer pour l'ensemble de leurs opérations au cours médian.

5. Résultats

Dans la présente section, nous étudions l'ampleur et l'incidence de la latence et les sources de latence pour le sous-échantillon d'opérations effectuées en juillet et août 2015. À la fin de la section, nous proposons un historique des coûts de latence du marché fondé sur l'échantillon tout entier.

Le classement par segment de négociateurs utilisé dans l'étude est décrit dans les travaux précédents de l'OCRCVM (2014).

iv. Ampleur et impact de la latence et sources de latence

Dans la section intitulée « Méthodologie », nous avons défini la valeur économique des opérations affectées par la latence en termes tant absolus (en dollars) que relatifs (en pourcentage du coût prévu de l'opération). Dans la discussion qui suit, nous étudions l'ampleur et l'impact de la latence et les sources de latence pour le sous-échantillon d'opérations effectuées en juillet et août 2015. Ce jeu de données permet une détermination plus précise des opérations exécutées au cours médian, comme expliqué dans la section II.

Ampleur

Le tableau 2A présente des statistiques sommaires qui permettent de mieux comprendre l'ampleur des opérations exécutées à un cours périmé au sein du sous-échantillon d'opérations effectuées en juillet et août. Les opérations exécutées à un cours périmé représentent 4,0 % des opérations invisibles exécutées au cours médian et avantagent les négociateurs actifs dans 88 % des cas. Les statistiques sommaires selon la valeur moyenne négociée (tableau 2B) donnent des résultats semblables.

L'observation selon laquelle les négociateurs actifs profitent plus fréquemment des opérations affectées par la latence confirme largement l'idée voulant que les latences sous-jacentes sont suffisamment observables et prévisibles pour que les participants sensibles à la latence en tirent profit. Dans le cas contraire, nous nous attendrions à ce que le coût de la latence soit réparti équitablement entre les participants actifs et passifs. Nous examinons cette idée plus en détail ci-dessous.

Le flux asymétrique des coûts et avantages

Répartition par segment de négociateurs

Nous avons vu ci-dessus que les opérations affectées par la latence se divisent en deux catégories, soit un groupe important d'opérations avantageant le négociateur actif et un groupe plus petit d'opérations avantageant le négociateur passif. Dans les opérations avantageant le négociateur actif, le côté passif paie le coût et le côté actif obtient l'avantage de la latence. À l'inverse, dans les opérations avantageant le négociateur passif, l'avantage échoit au côté passif et le coût est payé par le côté actif. Comme expliqué dans la section III, cet avantage ou ce coût s'ajoute au demi-écart que le participant s'attend à payer.

Le tableau 3 et le graphique 3 illustrent les avantages obtenus par le côté actif des opérations avantageant le négociateur actif et le côté passif des opérations avantageant le négociateur passif, par

négoceur et par contrepartie³. Bien que les négoceurs à haute vitesse soient les négoceurs actifs dans seulement 32 % des opérations invisibles exécutées au cours médian, ils obtiennent souvent un cours périmé avantageux. Parmi les négoceurs segmentés qui obtiennent un avantage en négociant activement à un cours périmé, les négoceurs à haute vitesse obtiennent 83 % de l'avantage, la plupart du temps au détriment de négoceurs ordinaires (97 %) plutôt que d'autres négoceurs à haute vitesse (3 %). En revanche, les négoceurs à haute vitesse actifs négocient rarement à un cours désavantageux et ne paient que 11 % des coûts totaux lorsque les négoceurs passifs obtiennent un avantage en négociant à un cours périmé. Lorsque les négoceurs ordinaires négocient activement, les coûts qu'ils paient et les avantages qu'ils obtiennent sont d'un ordre de grandeur semblable.

Les négoceurs à haute vitesse obtiennent l'avantage le plus important sur le plan du coût de latence relatif observé dans les opérations invisibles exécutées au cours médian. Lorsque les négoceurs à haute vitesse négocient activement contre des négoceurs ordinaires, les négoceurs ordinaires passifs paient 13,1 % de plus que le coût prévu pour négocier au point médian. En revanche, lorsque les négoceurs ordinaires négocient avec d'autres négoceurs ordinaires, le coût de latence relatif représente 1 % de plus que le coût prévu.

Il faut bien préciser que ces coûts et avantages sont dus à l'inefficacité des marchés pour ce qui est de traiter et d'indiquer la cotation la plus à jour. Nos résultats montrent que les négoceurs à haute vitesse peuvent exploiter de façon fiable cette inefficacité des marchés à établir les cours. Le tableau 4 montre que lorsque les négoceurs à haute vitesse négocient activement et obtiennent un avantage, ils sont en mesure de l'exploiter dans un délai moyen de 2,9 millisecondes.

L'avantage global obtenu par les négoceurs actifs qui négocient à un cours périmé avantageux est en moyenne de 2 969 \$ par jour pour les titres composant l'indice S&P/TSX 60⁴, soit 62 349 \$ par mois sur la période récente de deux mois, ce qui équivaut à un coût annuel estimatif de 748 188 \$. Si nous pouvions élargir le groupe d'opérations au cours médian déterminé précédemment (par exemple en y incluant les opérations invisibles donnant lieu à une amélioration du cours qui ne représente pas une fraction de cent ou les opérations qui ne sont pas entièrement attribuables à la latence en raison de limitations liées à la granularité de l'horodatage), nous verrions probablement l'avantage total en dollars augmenter. La section consacrée à l'historique de la latence précise la limite inférieure des avantages tout au long de la période historique.

Répartition par marché

L'avantage varie d'un marché à l'autre; se reporter au tableau 5 et au graphique 4, qui présentent les résultats des opérations avantageant le négoceur actif par marché. Les négoceurs passifs paient des coûts en dollars totaux élevés (et les négoceurs actifs obtiennent des avantages équivalents) sur deux marchés. Au total, quatre marchés affichent des coûts de latence relatifs élevés (allant de 2,2 % à

³ Nous classons les opérations faisant intervenir le groupe d'utilisateurs non segmenté dans la catégorie « Autres ». Soixante-dix pour cent de la valeur mensuelle négociée durant la période de l'étude est attribuable aux segments d'utilisateurs « à haute vitesse » et « ordinaires ».

⁴ Le calcul des coûts quotidiens moyens totaux tient compte des opérations exécutées par le groupe d'utilisateurs non segmenté.

4,2 % du coût prévu du demi-écart). Cette différence de coûts d'un marché à l'autre peut être corrélée à plusieurs facteurs comme la valeur totale des opérations, le nombre d'opérations, la volatilité et les stratégies de négociation dépendant de la conjoncture.

La différence de coûts dépend aussi de la durée de la latence. Lorsque la latence ne dure pas suffisamment longtemps pour qu'un négociateur puisse y réagir, nous pouvons nous attendre à observer une distribution symétrique des opérations avantageant le négociateur actif et le négociateur passif. Lorsque la latence est suffisamment importante pour que les négociateurs à haute vitesse puissent l'exploiter systématiquement, nous observerons un flux d'avantage asymétrique (ce dont fait état notre étude). Par conséquent, nous considérons la durée de la latence réelle comme le facteur sous-jacent clé qui contribue à l'ampleur des opérations affectées par la latence. Le tableau 6 illustre nos différentes observations de la durée de la latence d'un marché à l'autre. L'examen des facteurs corrélés à la latence fait l'objet de la discussion qui suit.

Examen des sources de latence dans les opérations invisibles exécutées au cours médian

Nous utilisons une analyse de régression (appliquée tant aux opérations avantageant le négociateur passif qu'aux opérations affectées par la latence avantageant le négociateur actif) pour examiner la relation entre la durée observée de la latence et un certain nombre de facteurs, dont les suivants :

- 1) le degré d'activité du marché concerné durant une brève période précédant l'opération exécutée à un cours périmé⁵;
- 2) le recours par les marchés à la co-implantation dans les locaux de la TSX⁶.

Le modèle de régression est le suivant :

$$Latence_{it,h,m} = \alpha_i + VNCO_m + \beta N_{it,h} + \sum_{j=1}^3 \gamma_j Contr\hat{o}le_{j,it} + \varepsilon_{it} \quad (7)$$

Ici, $Latence_{it,h,m}$ est la latence observée (définie dans la section III) mesurée en millisecondes pour l'opération h sur l'action i à la date t sur le marché m . $N_{it,h}$ représente l'activité du marché global au moment de l'opération exécutée à un cours périmé⁷. Cette expression mesure le nombre de messages d'ordres (sur tous les titres de capitaux propres canadiens) immédiatement avant la h^e opération sur l'action i au moment t sur tous les marchés. Cet indicateur est calculé pour la période $t_{mise\grave{a}jour} - 20$ ms jusqu'à $t_{mise\grave{a}jour}$. $VNCO_m$ est la variable nominale Co-implantation qui est égale à 0 pour les marchés qui ne sont pas co-implantés avec la bourse principale et à 1 pour ceux qui le sont. α_i représente un ensemble d'effets fixes liés aux actions. Les variables de contrôle comprennent la valeur quotidienne des opérations sur tous les marchés, le cours du titre et l'heure de la journée correspondant au moment de l'opération, afin de tenir compte des tendances tout au long de la journée.

⁵ Représenté par le nombre d'ordres, de modifications d'ordres et d'annulations d'ordres pour l'ensemble des titres de capitaux propres canadiens.

⁶ La majorité de l'information sur l'exécution des opérations et les cotations provient de la TSX.

⁷ Dans le cadre d'une vérification de la robustesse, nous avons calculé (mais non présenté) $N_{it,h}$ sur différentes périodes. Les deux méthodes produisent des résultats semblables.

Le tableau 7 présente les résultats de cette régression. Nous constatons que le recours à la co-implantation est corrélé négativement avec la latence. Cela peut indiquer que les mises à jour de cotations provenant des bourses importantes peuvent prendre davantage de temps à parvenir aux marchés qui ne sont pas co-implantés avec la bourse principale. La latence et le degré d'activité du marché (mesuré par le nombre d'ordres saisis immédiatement avant les opérations affectées par la latence) sont corrélés positivement avec une signification statistique élevée. Ce résultat indique qu'un nombre élevé de messages se traduit par des latences mesurables. Nous pressentons que des retards importants sont plus susceptibles de se produire lorsque le nombre de messages qui parvient aux marchés par unité de temps dépasse la capacité technologique de ces derniers.

v. Latence historique

Afin d'estimer la latence historique, nous utilisons les données de la période d'échantillonnage tout entière, qui comprennent toutes les opérations affectées par la latence (tant celles qui avantagent le négociateur actif que celles qui avantagent le négociateur passif). Le graphique 5 montre qu'au cours de cette période, le pourcentage d'opérations invisibles n'a cessé de croître. Le graphique 6 illustre l'évolution de la latence observée dans le cadre des opérations exécutées au cours médian et les impacts économiques qui en résultent depuis octobre 2012. La latence moyenne au cours de cette période de 35 mois est de 4,3 millisecondes, ce qui se traduit par un coût (ou un avantage) d'environ 42 500 \$ par mois. Comme l'indique la section III, cette estimation représente la limite inférieure du coût réel en raison des limites de nos capacités d'observation (granularité égale à la milliseconde) et de la présence des annotations nécessaires dans les données sur les opérations.

Nous observons également une grande variation dans la latence et son impact économique sur la période d'observation de trois ans, certaines périodes étant marquées par une détérioration importante. Les mois de décembre 2012 et janvier 2013, en particulier, révèlent des latences moyennes importantes sur tous les marchés. Cette période, qui figure en rouge dans le graphique 3, correspond sans doute aux changements dans la technologie ou la co-implantation sur certains marchés. Les mois écoulés entre octobre 2012 et avril 2013 révèlent également des coûts économiques élevés par rapport à la plupart des autres périodes. Les coûts mensuels totaux enregistrés au cours de cette période atteignent un sommet d'environ 100 000 \$.

6. Conclusion

Notre objectif était d'étudier la latence dans le calcul des cours de référence sur les marchés opaques canadiens. Les résultats montrent que même si la latence existe, les coûts économiques, en proportion des coûts de négociation globaux, ne sont pas élevés. Nous constatons que l'avantage lié à l'établissement de cours périmés, bien que faible, échoit de façon disproportionnée aux négociateurs à haute vitesse actifs.

Nous montrons également que la co-implantation et la capacité de gérer des volumes de messages importants sont corrélés avec la latence. Les marchés pourraient résoudre ce problème en investissant dans la technologie afin de veiller à ce que les cours de référence demeurent à jour quelle que soit la

conjoncture. On peut se demander si l'avantage qui en résulterait pour les participants aux marchés opaques contrebalancerait le coût des investissements en technologie. Il s'agit selon nous d'un débat important, que nos travaux contribueront à guider.

7. Tableaux

Tableau 1 : Marchés canadiens et types d'ordres

Marché	Ordres visibles/invisibles	Ordres invisibles au point médian/fixes
La Neo Bourse Aequitas - Registre transparent	Les deux	
La Neo Bourse Aequitas - Registre Neo Book	Les deux	
Alpha Exchange Inc. (Alpha)	Les deux	✓
Chi-X Canada ATS Ltd. (Chi-X Canada)	Les deux	✓
CX2 Canada ATS (CX2)	Les deux	✓
CNSX Markets Inc. – Canadian Securities Exchange (CSE)	Ordres visibles	
Instinet Canada Cross Ltd. (ICX)	Ordres invisibles	
Liquidnet Canada	Ordres invisibles	
Lynx ATS	Ordres visibles	
MATCH Now	Ordres invisibles	✓
Omega ATS	Les deux	✓
Bourse de Toronto (TSX)	Les deux	✓
Bourse de croissance TSX (BC-TSX)	Les deux	✓
TMX Select	Ordres visibles	

Tableau 2A : Statistiques sur le volume quotidien moyen d'opérations

Catégorie d'opérations	Opérations avantageant le négociateur passif	Opérations avantageant le négociateur actif	Total des opérations
Opérations exécutées à un cours périmé	253	1 828	2 081
Opérations sujettes à des limitations liées à la granularité de l'horodatage	173	993	1 166
Opérations non concluentes	417	184	601
Opérations exécutées à un cours approprié			48 141
Opérations invisibles exécutées au cours médian			51,989

Tableau 2B : Statistiques sur la valeur quotidienne moyenne des opérations (M\$)

Catégorie d'opérations	Volume d'opérations avantageant le négociateur passif	Volume d'opérations avantageant le négociateur actif	Volume total d'opérations
Opérations exécutées à un cours périmé	2,5	14,7	17,2
Opérations sujettes à des limitations liées à la granularité de l'horodatage	1,6	7,6	9,2
Opérations non concluentes	6,5	1,9	8,4
Opérations exécutées à un cours approprié			314,1
Opérations invisibles exécutées au cours médian			348,9

Tableau 3 : Avantage quotidien moyen lié à la latence, par contrepartie

Négociateur actif	Négociateur passif	Opérations avantageant le négociateur actif		Opérations avantageant le négociateur passif		Totaux	
		Avantage obtenu par le négociateur actif (\$)	Valeur des opérations (M\$)	Avantage obtenu par le négociateur passif (\$)	Valeur des opérations (M\$)	Coût prévu (\$)	Valeur des opérations (M\$)
à haute vitesse	ordinaire	1 353 (13,1 %)	6,6	17 (0,2 %)	0,1	10 264	50,1
ordinaire	ordinaire	261 (1,0 %)	1,4	117 (0,5 %)	0,4	25 268	85,9
à haute vitesse	à haute vitesse	39 (5,3 %)	0,19	5,0 (0,6 %)	0,03	763	3,2
ordinaire	à haute vitesse	30 (0,4 %)	0,15	59,7 (0,7 %)	0,3	8 974	26,8
	Autres	1 286 (2,6 %)	6,4	332 (0,01 %)	1,7	49 293	182,8
	Ensemble	2 969 (3,1 %)	14,7	531 (0,6 %)	2,5	94 562	348,8

Les valeurs entre parenthèses représentent le coût de latence relatif en pourcentage.

Tableau 4 : Durée de la latence observée (par opération), par contrepartie - Opérations avantageant le négociateur actif (ms)

Côté actif	Latence moyenne pondérée en fonction de la valeur	Moyenne	Écart-type	Médiane	3 ^e quantile	95 ^e centile
Négociateur à haute vitesse	3,65	2,93	7,00	1	2	10
Négociateur ordinaire	6,90	6,59	12,59	1	5	34
Autres	6,59	4,73	10,26	1	3	23
Ensemble	4,75	3,72	8,65	1	2	16

Tableau 5 : Avantage quotidien moyen généré par les opérations avantageant le négociateur actif, par marché

	Opérations avantageant le négociateur actif		Opérations avantageant le négociateur passif	
	Avantage obtenu par le négociateur actif (\$)	Coût de latence relatif (%)	Avantage obtenu par le négociateur passif (\$)	Coût de latence relatif (%)
M1	1 214	4,2 %	397	1,4 %
M2	1 420	3,9 %	15	0,04 %
M3	179	3,5 %	97	1,9 %
M4	14	2,2 %	13	1,0 %
M5	150	0,8 %	13	0,1 %
M6	4	0.1%	7	0,2 %
Ensemble	2 969	3,1 %	531	0,6 %

Tableau 6 : Durée de la latence observée (par opération), par marché (ms)

	Latence moyenne pondérée en fonction de la valeur	Moyenne	Écart-type	Médiane	3 ^e quantile	95 ^e centile
M1	8,56	7,28	13,27	2	6	35
M2	2,50	2,16	5,20	1	2	5
M3						
M4	7,76	6,46	13,08	1	5	33
M5						
M6						
Ensemble	5,98	4,80	10,61	1	3	24

Tableau 7 : Analyse de régression de la latence

Latence (ms)	
<i>Variable nominale Co-implantation</i>	-2,34 (-15,86***)
N_{ordres}	0,00949 (18,04***)
<i>Valeur quotidienne des opérations (M\$)</i>	0,115 (3,75***)
<i>Cours</i>	-0,0261 (-1,43)
<i>Heure de la journée</i>	0,0629 (3,62***)
Observations	66 445
R au carré ajusté	0,4067
Statistique F	701,8
Effet fixe	Actions

Les erreurs-types sont doublement groupées par action et par date.

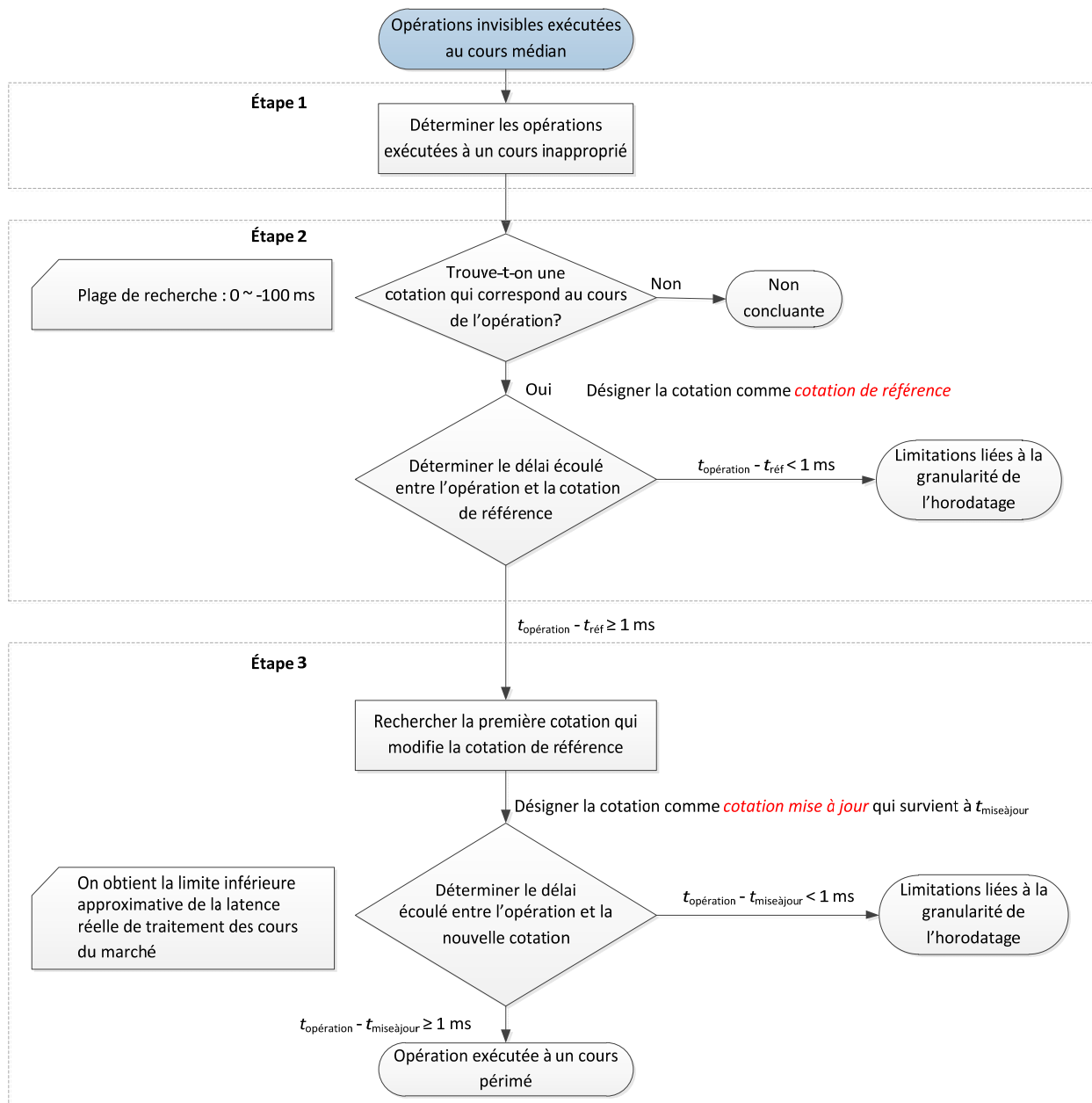
Les données utilisées pour la régression comprennent à la fois les opérations avantageant le côté actif et les opérations avantageant le côté passif.

Au niveau de la latence sur le marché, les valeurs aberrantes correspondant à des $N_{it,h}$ très élevés sont éliminées au seuil de 99 %.

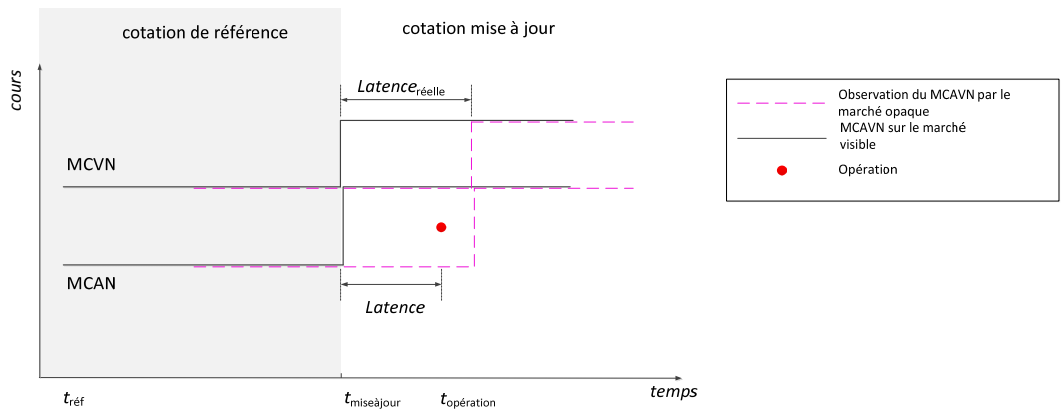
Le calcul des variables de contrôle est le suivant :

1) le cours se base sur l'écart moyen pondéré en fonction du temps; 2) l'heure de la journée est exprimée en fractions d'heures de la journée; p. ex., 09:30:00 est représenté par 9,5.

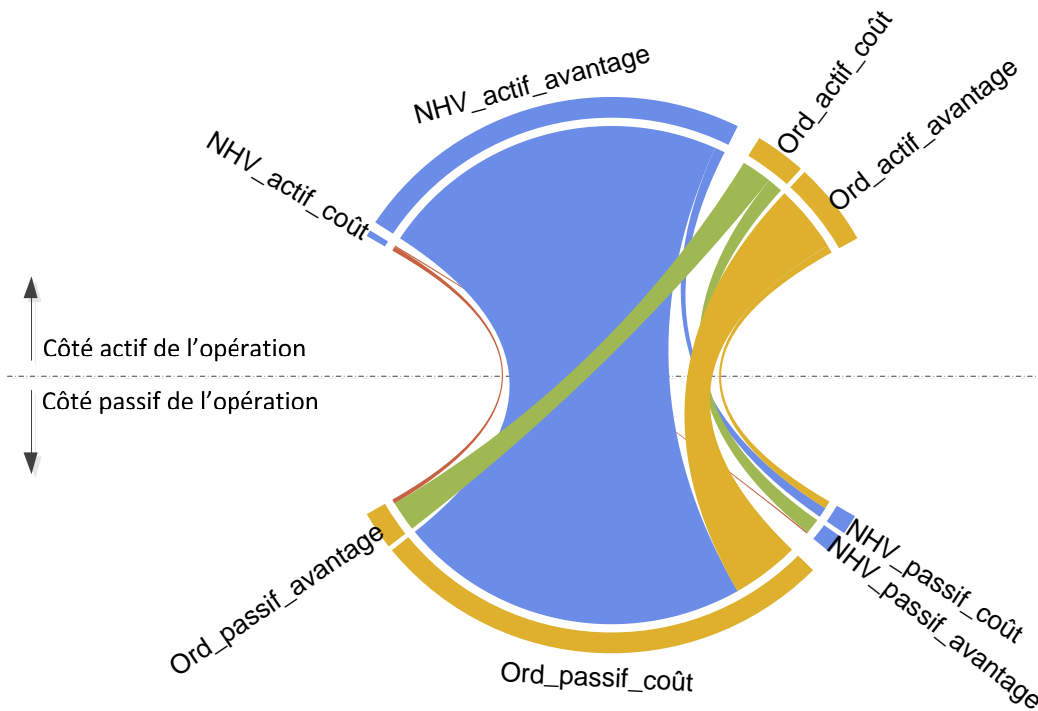
8. Graphiques



Graphique 1 : Étapes logiques de la détermination des opérations exécutées à un cours périmé

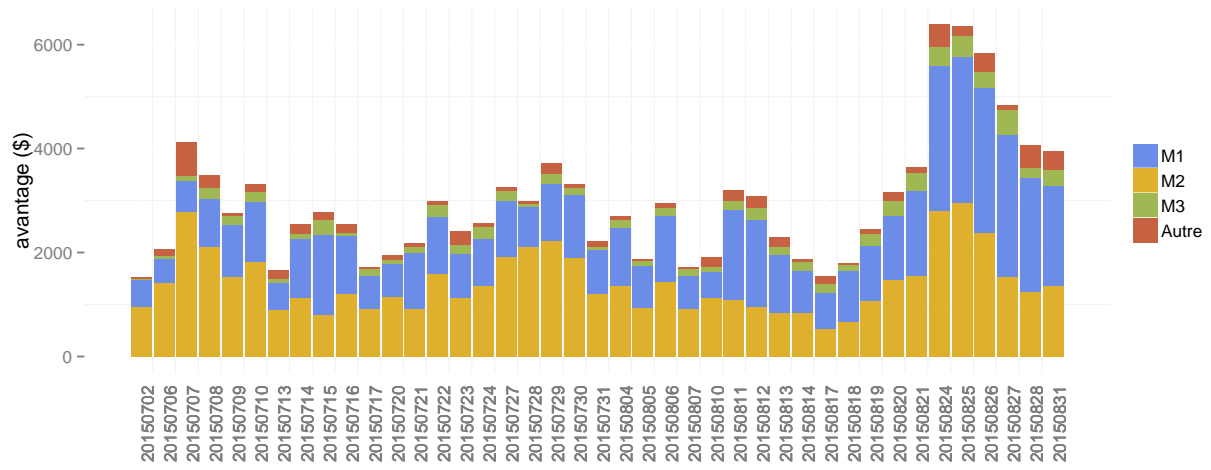


Graphique 2 : Illustration de la relation entre la latence observée et la latence réelle

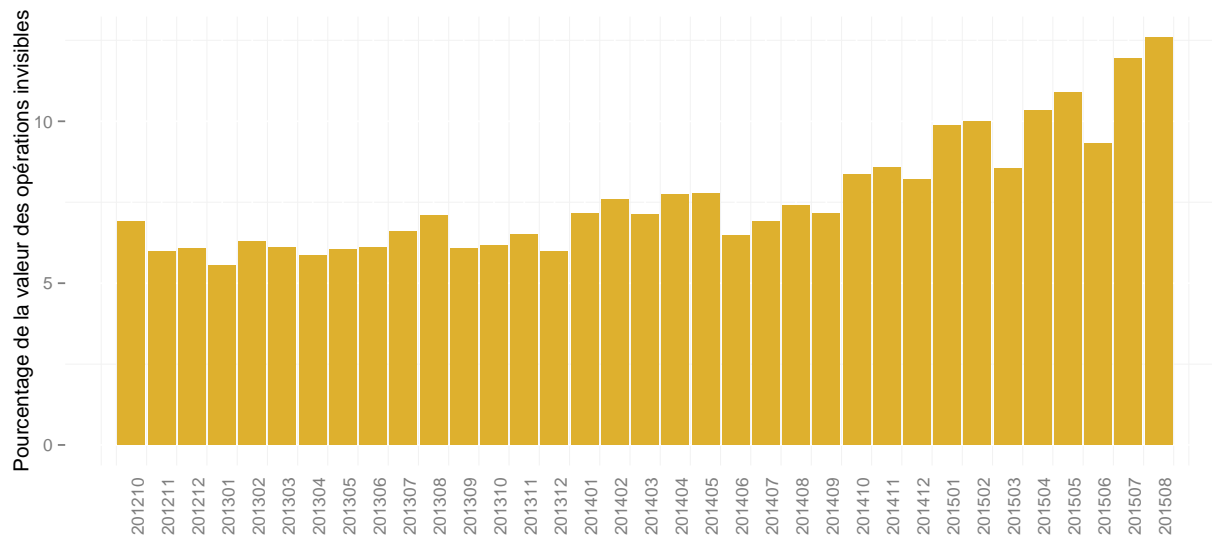


Graphique 3 : Analyse avantages/coûts des opérations affectées par la latence avantageant les négociateurs actifs et les négociateurs passifs, par segment de négociateurs et par contrepartie

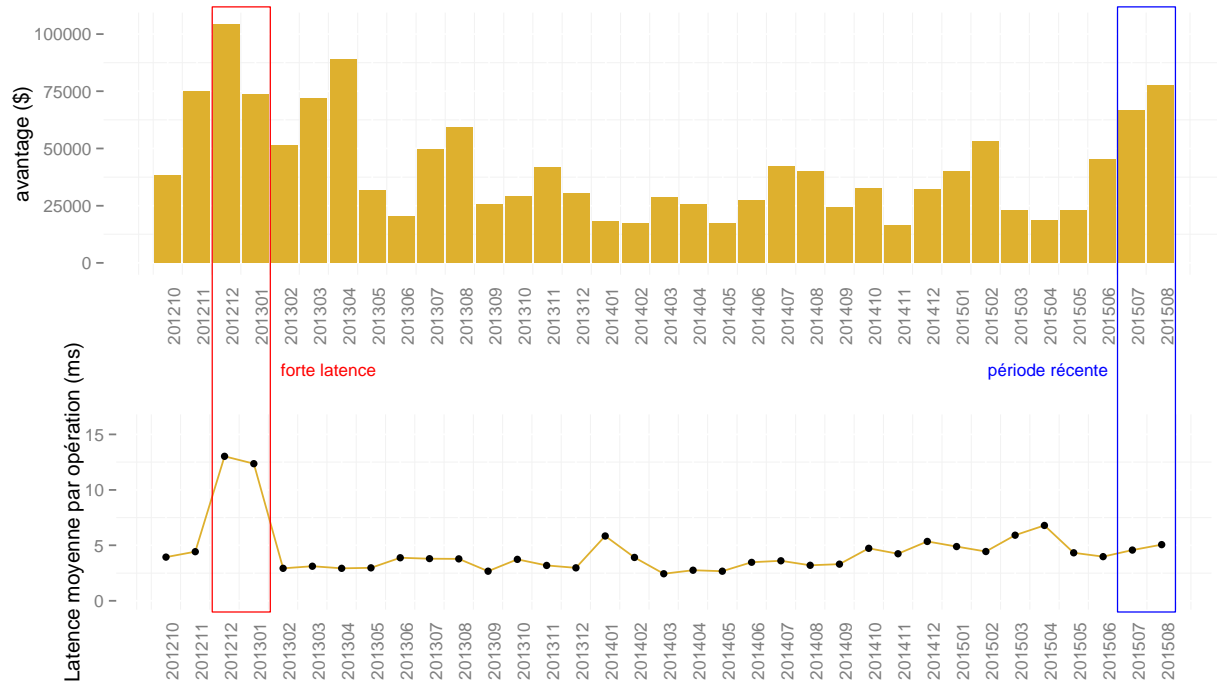
Pour chaque opération, le côté actif obtient un avantage ou paie un coût. La contrepartie passive à l'opération paie un coût lorsque le côté actif obtient un avantage, et obtient un avantage lorsque le côté actif paie un coût. Le graphique 3 est une représentation visuelle des données présentées aux quatre premières lignes du tableau 3, et illustre la proportion des coûts payés et des avantages obtenus par les segments des négociateurs à haute vitesse et des négociateurs ordinaires, du côté actif et du côté passif de l'opération, en montrant la contrepartie associée à chaque opération.



Graphique 4 : Avantage obtenu par le négociateur actif, par marché



Graphique 5 : Pourcentage de la valeur des opérations invisibles



Graphique 6 : Latence historique – Durée et avantage économique

Bibliographie

Alexander, Jeff, Linda Giordano et David Brooks. *Dark Pool Execution Quality: A Quantitative View*, Tabb Forum, 2015.

Bessembinder, Hendrik et Kumar Venkataraman. « Bid-Ask Spreads », dans *Encyclopedia of Quantitative Finance*, John Wiley & Sons, Ltd., 2010.

Biais, Bruno, Thierry Foucault et Sophie Moinas. « Equilibrium Fast Trading », *AFA 2013 San Diego Meetings Paper*, 2014.

Ding, Shengwei, John Hanna et Terrence Hendershott. « How Slow is the NBBO? A Comparison with Direct Exchange Feeds », *Financial Review*, vol. 49, n^o 2, mai 2014, p. 313 à 332.

Easley, David, Terrence Hendershott et Tarun Ramadorai. « Levelling the Trading Field », *Journal of Financial Markets*, mars 2013.

Foucault, Thierry, Johan Hombert et Ioanid Rosu. « News Trading and Speed », *Journal of Finance*, à paraître.

Hoffmann, Peter. « A Dynamic Limit Order Market with Fast and Slow Traders », *Journal of Financial Economics*, 2014, p. 156 à 169.

OCRCVM. « Classification des groupes de négociation – Méthodologie et résultats », 2014.

Annexe A

L'écart effectif en tant qu'indicateur d'opérations potentiellement exécutées à un cours périmé

L'écart effectif (EE) est une estimation du coût d'exécution payé par le consommateur de liquidité et du revenu gagné par le fournisseur de liquidité. Le semi-écart effectif est défini comme suit (Bessembinder *et al.*, 2010) :

$$EE = D \times \frac{C_{it} - M_{it}}{M_{it}}$$

où D est égal à 1 pour les ordres saisis par l'acheteur et à -1 pour les ordres saisis par le vendeur; C_{it} est le cours du titre i au moment t ; et M_{it} est la représentation observable de la valeur sous-jacente réelle du titre i au moment t . Ici, M_{it} correspond à la plus récente cotation avant l'opération.

Pour les opérations exécutées au cours médian, quel que soit le côté (acheteur ou vendeur) qui effectue l'opération, P_{it} est, en théorie, toujours égal à M_{it} et

$$EE = D \times \frac{M_{it} - M_{it}}{M_{it}} = 0$$

Cependant, l'écart effectif peut être négatif lorsqu'une cotation change en faveur du négociateur passif mais que le marché continue d'observer une « ancienne » cotation inférieure (soit la cotation de référence définie à l'étape 2 de l'alinéa 4.iv) et exécute l'opération en fonction de l'ancien cours médian. De la même façon, l'écart effectif peut être positif lorsque la cotation change en faveur du négociateur actif mais que le marché continue d'observer l'ancienne cotation et d'exécuter l'opération en fonction de celle-ci.

Supposons par exemple qu'à 11:30:00.000, le meilleur cours acheteur et vendeur national soit 10,01 x 10,05. À 11:30:00.003, le marché A transmet une nouvelle cotation qui modifie le meilleur cours acheteur et vendeur national, qui devient 10,03 x 10,05. Cependant, à 11:30:00.004, le marché opaque B observera peut-être encore la cotation 10,01 x 10,05 en raison du retard avec lequel il reçoit et/ou traite l'information transmise par les autres marchés. Si un ordre d'achat actif est saisi avant que le marché B puisse observer la nouvelle cotation, il sera exécuté au cours médian 10,01 x 10,05. L'acheteur actif, qui devait acheter à $(10,03 + 10,05)/2 = 10,04$, achètera plutôt à $(10,01 + 10,05)/2 = 10,03$.

Dans cet exemple, l'écart effectif pour l'acheteur actif est négatif et est calculé comme suit :

$$EE = D \times \frac{C_{it} - M_{it}}{M_{it}} = 1 \times \frac{(10,01 + 10,05)/2 - (10,03 + 10,05)/2}{(10,03 + 10,05)/2} < 0$$

Dans notre étude, nous désignons les opérations qui présentent un écart effectif non nul comme des opérations exécutées à un cours périmé.