

Résumé de la thèse

Four Essays in Financial Econometrics

Georgiana-Denisa BANULESCU¹

Ma thèse de doctorat se concentre principalement sur des mesures du risque financier et la modélisation de la volatilité. L'objectif général est : (i) de proposer de nouvelles techniques pour mesurer à la fois le risque systémique et le risque à haute fréquence, et (ii) d'appliquer et d'améliorer les outils économétriques de modélisation et de prévision de la volatilité. Ce travail comporte quatre chapitres (papiers de recherche).

La première partie de la thèse traite des questions liées à la modélisation et la prévision des mesures du risque à haute fréquence et du risque systémique. Plus précisément, le deuxième chapitre propose une nouvelle mesure du risque systémique utilisée pour identifier les institutions financières d'importance systémique (SIFIs). Basée sur une approche spécifique, cette mesure originale permet de décomposer le risque global du système financier tout en tenant compte des caractéristiques de l'entreprise. Le troisième chapitre propose des mesures du risque de marché intra-journalier dans le contexte particulier des données à haute fréquence irrégulièrement espacées dans le temps (*tick-by-tick*).

La deuxième partie de la thèse est consacrée aux méthodes d'estimation et de prévision de la volatilité incluant directement des données à haute fréquence ou des mesures réalisées de volatilité. Ainsi, dans le quatrième chapitre, nous cherchons à déterminer, dans le contexte des modèles de mélange des fréquences d'échantillonnage (MIDAS), si des régresseurs à haute fréquence améliorent les prévisions de la volatilité à basse fréquence. Une question liée est de savoir s'il existe une fréquence d'échantillonnage optimale en termes de *prévision*, et non de *mesure* de la volatilité. Le cinquième chapitre propose une version robuste aux jumps du modèle Realized GARCH. L'application porte sur la crise financière récente, et vise à identifier et analyser les événements ayant induit ces sauts pendant la crise.

Mots clés : Risque Systémique, Component Expected Shortfall, Mesures de Risque à Haute Fréquence, Backtesting, Value-at-Risk, Time-at-Risk, MIDAS, Volatilité

¹ European University Institute et LEO, email : georgiana.banulescu@eui.eu

La recherche en économétrie financière a connu un profond renouvellement ces dernières années suite à deux mutations majeures des *données de marché* collectées et mises à disposition dans la sphère académique : (i) la plus grande disponibilité des données à très haute fréquence, *e.g.*, des informations précises sur les transactions sur le marché des actions, obligations, devises, options etc., et (ii) l'accès à des ensembles de données de très grande dimension : par exemple des rendements historiques portant sur un très grand nombre d'actifs financiers et des portefeuilles de grande dimension.

I. Données financières à haute fréquence

La disponibilité des données à haute fréquence a impacté principalement trois domaines de recherche : (i) la mesure de la volatilité, (ii) l'analyse empirique du bruit de microstructure et (iii) la prévision de la volatilité.

Mesure de la volatilité

La volatilité a été, et restera, l'un des sujets de recherche les plus actifs dans le domaine de l'économétrie financière et de la prévision économique. La caractéristique principale du processus de la volatilité est le fait qu'il n'est jamais observé. Les premiers travaux sur la volatilité, initiés par Engle (1982) et généralisés par Bollerslev (1986), ont été conçus pour fournir des estimations *ex-ante* de la volatilité conditionnelle. Or, la disponibilité des données financières échantillonnées à très haute fréquence, a permis d'initier un courant de recherche sur la mesure *ex-post* de la volatilité. L'idée principale est d'utiliser des données à haute fréquence pour construire *ex-post* des mesures réalisées de la volatilité à une fréquence plus basse. L'économétrie de la volatilité réalisée a été initiée notamment par Andersen et Bollerslev (1998a), Andersen et al. (2001a, 2003), Barndorff-Nielsen et Shephard (2002), Dacorogna et al. (2001), *inter alios*.

L'intuition de la volatilité réalisée s'appuie sur le processus de volatilité en temps continu qui est évidemment lié à l'évolution de la volatilité instantanée. Néanmoins, la volatilité est mesurée généralement à des intervalles discrets dans le temps, même si le processus sous-jacent évolue en temps continu. En intégrant la volatilité instantanée sur un horizon quotidien, par exemple, on obtient une mesure *ex-post* de la variance intégrée quotidienne (IV : *Integrated Volatility*). En présence de sauts, la volatilité réalisée est un estimateur convergent de la variation quadratique (QV : *Quadratic Variation*) qui étend la définition de la volatilité intégrée en incluant également les sauts au carré.

Etant donné que la volatilité est un processus latent, il est en effet nécessaire de construire un estimateur de l'IV ou de la QV. La première mesure réalisée de volatilité a été introduite par Andersen et Bollerslev (1998a). Leur estimateur est connu sous le nom de variance réalisée (RV : *Realized Variance*). La RV est définie comme la somme des carrés des rendements intra-journaliers. Andersen et al. (2003, 2001b) et Barndorff-Nielsen et Shephard (2002) montrent que la variance réalisée est un estimateur sans biais de la variance intégrée et qu'elle apporte une information beaucoup plus précise sur le niveau actuel de volatilité que le rendement journalier au carré. En effet, la variance réalisée fournit une estimation précise de la volatilité car les prix sont observés en continu et sans erreur de mesure (Merton, 1980). Toutefois, lorsque la fréquence d'échantillonnage des rendements intrajournaliers augmente, la variance réalisée souffre d'un problème de biais qui a été étudié notamment par Andreou et Ghysels (2002), Griffin et Oomen (2008), Bai et al. (2004) et Ait-Sahalia et al. (2005). Ces études montrent que les données à très haute fréquence sont généralement polluées par le bruit de microstructure qui induit de l'autocorrélation dans les rendements intra-journaliers et rend les estimateurs réalisés de la volatilité peu fiables. Ainsi, en utilisant des méthodes d'évaluation spécifiques (*e.g.*, *signature plot*), Andersen et al. (2000) montrent que la variance réalisée peut être sujette à un biais de microstructure. Cette source de biais liée au bruit de microstructure (principalement représentée par des sauts importants dans la série de rendements) a ainsi motivé le développement d'estimateurs robustes de la volatilité. Mentionnons ici les estimateurs introduits par Barndorff-Nielsen et Shephard (2004b) (*power* et *bipower variation*), les estimateurs à deux- et multi-échelles proposés par Zhang (2004) et Zhang (2006), l'estimateur *realized kernel* de Barndorff-Nielsen et al. (2008), ainsi que les estimateurs proposés par Christensen et al. (2008). Voir aussi Andersen et al. (2008), Hansen et Horel (2009a) pour plus de détails.

Bruit de microstructure

Ait-Sahalia et Yu (2009) identifient trois sources principales de bruit de microstructure : (i) les frictions inhérentes au processus de transaction telles que la différence entre l'offre et la demande, les changements discrets de prix, et les arrondis ; (ii) les effets informationnels tels que la latence, l'asymétrie d'information, la réponse séquentielle sur les prix des transactions en bloc, les effets des inventaires sur les stocks ; et (iii) la mauvaise mesure ou les erreurs d'enregistrement des données. Dans ce contexte, il est plus réaliste de décomposer le logarithme du prix observé de l'actif dans la somme du logarithme du prix, que l'on appelle efficace, et une composante de bruit de microstructure.

Les discontinuités importantes présentes dans les séries de prix à très haute fréquence sont principalement associées à ce que nous appelons sauts ou *jumps*. Ces sauts ne sont pas réguliers et apparaissent la plupart du temps en raison de nouvelles surprises. Comme nous l'avons mentionné précédemment, ils peuvent induire un biais dans les estimations de la volatilité. Pour étudier l'apparition de ces sauts, ainsi que l'information de marché associée, plusieurs tests formels ont été proposés afin d'identifier la présence de ces sauts et de leur appliquer un traitement spécifique. Ces tests ont été tout d'abord conçus pour identifier les sauts dans des données à basse fréquence (voir Ait-Sahalia 2002; Carr et Wu, 2003; Johannes, 2004; Wang, 1995, entre autres), mais ont été rapidement étendus à des données à haute fréquence.

Les premières contributions à la détection de sauts dans le contexte des données intra-journalières sont celles de Barndorff-Nielsen et Shephard (2004b), Barndorff-Nielsen et al. (2006), Andersen et al. (2007a), Mykland et Zhang (2008) (généralisé par Lee et Hannig, 2010) et Boudt et al. (2012), entre autres. Barndorff-Nielsen et Shephard (2004b), Barndorff-Nielsen et al. (2006) proposent une mesure de la variance intégrée robuste aux sauts appelée *bipower variation*. En comparant une mesure de volatilité sensible aux sauts (par exemple, la variance réalisée) à la *bipower variation*, les auteurs identifient les observations affectées et mesurent la contribution de ces sauts à la variation du rendement total.

Andersen et al. (2007a) et Mykland et Zhang (2008) ont étendu l'identification des sauts au niveau intrajournalier en développant un test non paramétrique conçu pour tester directement si un rendement intra-journalier donné est affecté d'une composante de saut. Les avantages de ce test sont principalement liés à l'identification de la date exacte des sauts et à la possibilité d'étudier la distribution des sauts ainsi que leur intensité stochastique. Boudt et al. (2012) améliorent les méthodes existantes de détection de sauts en prenant également en compte la dynamique spécifique des rendements à haute fréquence, à savoir la périodicité intra-journalière. La courbe en U utilisée pour mettre en évidence la périodicité de la volatilité intra-journalière a ses racines dans les études de Wood et al. (1985) et Harris (1986), mais elle a été également intégrée dans des études plus récentes de Andersen et Bollerslev (1997, 1998b) et Hecq et al. (2012). Boudt et al. (2012) prouvent que la prise en compte de la périodicité intra-journalière dans le cadre de ces tests améliore considérablement leur performance. La nouvelle spécification des tests détecte donc des sauts relativement petits qui se produisent pendant des périodes de faible volatilité et réduit le nombre de "faux" sauts détectés lorsque la volatilité est élevée.

Une autre branche de cette littérature est consacrée à l'étude des sauts contemporains (*cojumps*). Gilder et al. (2014) définissent les *cojumps* comme des sauts qui apparaissent de façon simultanée dans les prix de plusieurs actifs. Gilder et al. (2014) envisagent deux hypothèses pour justifier la présence de *cojumps*. La première hypothèse repose sur une association entre les sauts dans le rendement des portefeuilles et les sauts dans les actifs constituants ou les actifs sous-jacents dans le cas de portefeuille de dérivés. Les *cojumps* apparaissant dans un portefeuille de marché sont également appelés *cojumps* systématiques, car ils font référence à des sauts non diversifiables. Bollerslev et al. (2008) ont été les premiers à examiner cette hypothèse en proposant des tests non paramétriques. L'identification des grands sauts présents à la fois dans le portefeuille de marché et les actions individuelles est également abordée dans Todorov et Bollerslev (2010) et Bollerslev et al. (2013). La seconde hypothèse examine si les *cojumps* systématiques sont associés à des annonces macroéconomiques, sachant que ce type de nouvelle possède une influence considérable au niveau des marchés. Cette question a également été documentée sur d'autres marchés, tels

que les marchés obligataires (Dungey et al., 2009; Dungey et Hvozdík, 2012), les marchés actions (Bajgrowicz et Scaillet, 2011), obligations et taux de change (Lahaye et al., 2011), le marché à terme (Evans, 2011).

L'influence des sauts présente un intérêt particulier en période de crise, caractérisée par les phénomènes de contagion. Ait-Sahalia et al. (2013) montrent qu'un saut qui se produit dans une certaine région du monde augmente l'intensité des sauts à la fois dans la même région (auto-excitation) ainsi que dans d'autres régions (excitation mutuelle). Ce qui importe n'est pas le saut initial, mais l'amplification ultérieure induite sur la volatilité des différents marchés dans les prochaines heures ou prochains jours. Les auteurs montrent l'existence de clusters de volatilité, à la fois dans le temps et dans l'espace. Ils proposent un modèle capable de reproduire la propagation des chocs dans le temps et dans l'espace, ainsi que des méthodes d'estimation et des tests appropriés à ce type de modèle.

Prévision de volatilité

La disponibilité des données à haute fréquence a deux implications distinctes sur la littérature de la volatilité. La première implication, discutée précédemment, concerne la *mesure* du processus latent de la volatilité. Cette dimension ne nécessite pas de modèle explicite. La seconde implication concerne la modélisation de la dynamique de la volatilité quotidienne grâce à des techniques standard de séries temporelles, tout en exploitant les informations fournies par les données à haute fréquence. Nous distinguons ainsi le contexte de *prévision* de la volatilité de celui de la *mesure* de la volatilité.

Comme nous l'avons dit précédemment, les mesures réalisées de volatilité sont de plus en plus utilisées car elles fournissent des informations beaucoup plus précises sur le niveau actuel de la volatilité quotidienne que les rendements au carré. Par conséquent, ces mesures réalisées ont été utilisées dans le cadre des modèles de volatilité conditionnelle (GARCH). Engle (2002b) a été le premier à examiner ce type d'approche, en proposant le modèle GARCH-X qui inclut la variance réalisée dans l'équation de GARCH. Barndorff-Nielsen et Shephard (2007) ont étendu ce modèle en utilisant à la fois la variance réalisée et la bipower variation. Cependant, le modèle GARCH-X est considéré comme incomplet (partiel), car il ne s'intéresse pas du tout à la dynamique des mesures réalisées. Une approche complète, modélisant de manière jointe les rendements et la mesure réalisée de la volatilité, a été proposée par Engle et Gallo (2006) dans le cadre d'un modèle à erreur multiplicative (MEM : *Multiplicative Error Model*). Une structure MEM simplifiée a été proposée par Shephard et Sheppard (2010), sous le nom de modèle HEAVY. Plus récemment, Hansen et al. (2012) ont introduit le modèle Realized GARCH qui repose sur une approche légèrement différente de la modélisation jointe des rendements et des mesures réalisées de la volatilité. La principale différence avec les modèles précédents réside dans la présence d'une équation de mesure qui détermine la dynamique de la mesure réalisée en relation avec la variance conditionnelle sous-jacente.

Une autre direction de cette littérature se concentre sur les modèles qui incluent des données échantillonnées à des fréquences différentes, à savoir les modèles de régression MIDAS (*MIXed DATA Sampling*). L'idée de cette approche est de considérer des prévisions de volatilité à basse fréquence (variable dépendante) en conservant l'information provenant des données à haute fréquence (variable explicative). Par exemple, la volatilité journalière peut être prédite directement à l'aide de prédicteurs intra-journaliers. Le cadre théorique des modèles MIDAS est proposé dans Ghysels et al. (2004b). L'application au contexte de la prévision de volatilité a été proposée dans Ghysels et al. (2005) Ghysels et al. (2006), Ghysels et Sinko (2011), Ghysels et Valkanov (2012), Chen et Ghysels (2011), Ghysels et al. (2007), Ghysels et Sinko (2011), Alper et al. (2008), Arago et Salvador (2010), Chen et al. (2011), Forsberg et Ghysels (2007), entre autres. Une des caractéristiques essentielles des modèles MIDAS est qu'ils fournissent une spécification parcimonieuse. Sachant que l'inclusion des données à haute fréquence pourrait impliquer une augmentation significative du nombre de retards des variables explicatives, une solution consiste à introduire une fonction de poids paramétrique pour contourner ce problème de prolifération des paramètres.

Plus récemment, Corsi (2009) a proposé le modèle HAR-RV (*Heterogeneous AutoRegressive Realized Volatility*) qui est lié au modèle de régression MIDAS. En fait, HAR-RV peut être situé à la frontière entre les processus de volatilité à mémoire longue et ceux à mémoire courte car il est conçu comme une spécification additive (en cascade) de la volatilité réalisée agrégée à différents horizons temporels. Andersen et al. (2007b) montrent que ce modèle peut être considéré comme une régression MIDAS avec une fonction de poids en escaliers. Entre temps, différentes extensions du modèle HAR univarié ont été proposées. Ces extensions visent à prendre en compte l'influence des jumps dans la prévision (Andersen et al. 2007b), l'effet de levier, ainsi que les ruptures structurelles (Bauwens et al. 2012).

II. Modélisation en grande dimension

Les défis associés aux bases de données actuelles ne sont pas uniquement liés à la fréquence d'échantillonnage des observations, à des intervalles de temps irréguliers ou à la dynamique intrajournalière de l'activité financière. Ils sont aussi liés à l'adaptation des modélisations économétriques à la problématique des bases de données de très grande dimension (portefeuilles de grande dimension, grande diversité d'actifs, etc.). On distingue généralement deux problèmes : le fait de travailler avec des échantillons contenant un grand nombre d'actifs, et le fait de travailler avec des échantillons de grande dimension temporelle. Nous nous concentrerons sur le premier problème qui appelle des traitements spécifiques. Les principales contributions sur ce sujet portent sur (i) l'estimation des matrices de variance-covariance de grande dimension, (ii) la mise en oeuvre de nouvelles techniques d'estimation dans le cadre de modèle de grande dimension, et (iii) de nouveaux enjeux relatifs à la théorie asymptotique.

Estimation de matrices de variance-covariance de grande dimension

La théorie du portefeuille optimal implique de prédire au mieux le rendement attendu (en excès) de tous les actifs ainsi que leur niveau de risque et de dépendance, mesurés dans un contexte gaussien par la matrice de variance-covariance. Plusieurs études ont montré qu'il est plus difficile de prédire l'espérance des rendements que la matrice de variance-covariance (voir Merton, 1980; Nelson, 1992; Chopra et Ziemba, 1993). C'est pourquoi il est souvent suggéré dans la littérature de fonder la décision d'investissement exclusivement sur les estimations de la matrice de variance-covariance et d'investir donc dans le portefeuille à variance globale minimale (voir Haugen et Baker, 1991; Jagannathan et Ma, 2003). Dans ce cadre, se pose le problème du risque d'estimation de la matrice de variance-covariance. Ce problème se pose avec d'autant plus d'acuité dès lors que l'on considère des portefeuilles de grande dimension incluant un très grand nombre d'actifs par rapport au nombre d'observations historiques disponibles pour les rendements. La dimension transversale, N , de la matrice de variance-covariance peut être plus importante que la dimension temporelle, T , de l'échantillon qui sert à estimer cette matrice. L'estimateur usuel de la variance-covariance empirique est alors soit non défini, soit peu performant (Ledoit et Wolf, 2004). Ainsi, l'estimateur défini par la matrice de variance-covariance empirique devient donc inadapté pour de nombreuses applications, y compris l'optimisation de portefeuille et la minimisation de l'erreur de suivi.

Comme la structure de la matrice de variance-covariance est inconnue, Ledoit et al. (2012) montrent que l'approche la plus efficace pour l'estimer consiste à utiliser une technique dite de rétrécissement. Sur la base d'un raisonnement asymptotique où N tend vers l'infini avec T , Ledoit et Wolf (2004) dérivent une formule de rétrécissement linéaire optimale.² Cette approche consiste à trouver un bon compromis entre l'erreur d'échantillonnage et l'erreur de spécification. La procédure de rétrécissement de la matrice de variance-covariance est alors fondée sur une cible visant un niveau faible du risque d'estimation. Cependant, la qualité de cette approximation est mitigée parce les estimateurs à rétrécissement du portefeuille à variance minimale ne sont plus optimaux pour des petites valeurs de T . Candelon et al. (2012) montrent qu'une méthode de rétrécissement double pourrait améliorer l'estimateur de la matrice de variance-covariance dans le contexte de petits échantillons d'estimation. En outre, Ledoit et al. (2012) abordent également la nécessité d'un estimateur qui se démarque à la fois dans des échantillons à distance finie et

² La théorie asymptotique standard suppose que N est fixe, tandis que T tend vers l'infini.

asymptotiquement. A cet effet, ils proposent une méthode de rétrécissement non linéaire. La différence par rapport à l'estimateur linéaire consiste à appliquer une intensité de rétrécissement spécifique à chaque valeur propre de l'échantillon (Ledoit et Pécché, 2011). Les auteurs montrent qu'à distance finie l'estimateur non linéaire est considérablement plus performant que l'estimateur linéaire.

Modèles économétriques à grande dimension

Une autre implication de la grande profusion des données individuelles et/ou financières, est la nécessité de concevoir des techniques d'estimation appropriées pour des modèles économétriques incluant un grand nombre de variables explicatives. Deux principales questions sont à considérer dans ce cas : la spécification la plus adaptée pour construire un modèle de prévision et le développement de techniques d'inférence adaptées.

Les modèles associés à des bases de données de large dimension sont principalement caractérisés par une grande variété de variables de contrôle/prédicteurs potentiels, sachant que seulement un petit nombre de variables sont réellement importantes pour prédire la variable d'intérêt. Dans ce cas, lorsque le nombre initial de variables est élevé, le chercheur a besoin de réduire leur nombre à une taille gérable. Hastie et al. (2009) proposent une revue des outils statistiques disponibles pour la sélection d'un sous-ensemble de variables de contrôle qui pourraient fournir de bonnes prévisions de la variable d'intérêt. Les auteurs distinguent deux classes de méthodes : les méthodes de sélection de sous-ensembles et les méthodes de rétrécissement. Le premier procédé de sélection consiste à retenir uniquement un sous-ensemble de variables et d'éliminer le reste. Cette technique est présentée dans la littérature comme une forme de "seuil-dur" (*hard-thresholding*). Trois méthodes sont distinguées dans ce cas : la sélection du meilleur sous-ensemble, la régression "forward-stage-wise" et les procédures "forward", "backward" et "stepwise". Les méthodes de rétrécissement, *i.e.*, les régressions "ridge" et les méthodes LASSO (*Least Absolute Shrinkage and Selection Operator*), viennent compléter les procédures de sélection précédentes en surmontant certains de leurs inconvénients. Ces dernières méthodes sont plus continues et leurs résultats ne sont pas soumis à une variabilité élevée suivant le choix de seuils arbitraires, etc. L'idée principale de la régression "ridge" est de diminuer le nombre des coefficients de la régression en imposant une pénalité sur leurs valeurs. La méthode LASSO (Tibshirani, 1996) est conçue comme une méthode de sélection des variables dans le cadre d'une régression linéaire, mais le seuil imposé est plus "doux" que celui utilisé par d'autres méthodes de sélection (*soft thresholding*). LASSO appartient à la famille des modèles de régression à pénalité et fonctionne en pénalisant l'objectif de l'optimisation par un terme qui implique la norme l_1 des coefficients. Toutes ces méthodes constituent des alternatives à l'analyse en composantes principales.

De nombreux auteurs montrent que les méthodes à pénalité ont de bonnes propriétés d'estimation ; voir Candès et Tao (2007), Bickel et al. (2009), Meinshausen et Yu (2009), Huang et al. (2010), Belloni et al. (2011). En outre, Bickel et al. (2009) et Belloni et al. (2012a) contribuent à cette littérature par l'extension de ces méthodes à des cas non paramétriques et non gaussiens.

Théorie asymptotique

Les méthodes mentionnées ci-dessus sont connues pour bien prédire la variable d'intérêt, mais ils conduisent souvent à des conclusions erronées en ce qui concerne l'inférence sur les paramètres du modèle (Leeb et Pötscher, 2008). Belloni et al. (2012b) soulignent que dans le cas d'un modèle de sélection imparfait, la littérature sur l'inférence est relativement pauvre. En outre, la plupart des méthodes standard de sélection des variables ne sont pas robustes et peuvent donner de mauvais résultats en matière d'inférence. La plupart du temps, afin d'évaluer la robustesse de leur inférence, les chercheurs considèrent des analyses de sensibilité *ad hoc* qui consistent à faire changer tout simplement l'ensemble des variables de contrôle. Pour améliorer cette analyse de sensibilité, Belloni et al. (2012b) appliquent des méthodes modernes de sélection des variables menant à une inférence robuste et valide. Les auteurs présentent des résultats théoriques et prouvent que l'inférence associée à leur méthode est généralement valable sur des classes vastes de modèles.

Par ailleurs, dans le cadre des applications où la dimension N est comparable ou supérieure à T , le théorème centrale limite classique ne s'applique plus. En ce sens, Chernozhukov et al. (2013) dérivent un théorème centrale limite pour le maximum d'une somme des vecteurs aléatoires de grande dimension. Ils indiquent les conditions sous lesquelles la distribution du maximum est approximée par le maximum d'une somme de vecteurs aléatoires gaussiens ayant les mêmes matrices de variance-covariance que les vecteurs d'origine. Cette approximation est utile dans le cadre de l'estimation des matrices de grande dimension et dans la construction des tests d'hypothèses multiples, entre autres.

L'utilisation de bases de données de grande dimension a également inspiré et motivé des progrès dans la théorie asymptotique en panel. Il existe ici trois principales directions de recherche : (i) la convergence séquentielle lorsque T puis N (ou vice et versa) tendent vers l'infini (ii) la convergence diagonale lorsque T et N tendent vers l'infini mais le ratio T/N est supposé constant ; (iii) la convergence libre où T et N tendent vers l'infini sans aucune restriction. Gagliardini et al. (2013) proposent une méthodologie économétrique afin d'étudier l'évolution des primes de risque à partir d'un panel non-cylindré d'environ 10 000 actions. L'originalité de leur travail consiste à dériver la convergence et la normalité asymptotique de leurs estimateurs en laissant la dimension temporelle, T , et la dimension transversale, N , croître à l'infini simultanément et non successivement, tout en prenant en compte les dépendances (transversales) potentielles dans la série des rendements.

III. Mesures de risque

Au-delà des progrès de la théorie économétrique induits par l'utilisation des données à haute fréquence et des bases de grande dimension, d'autres facteurs ont conduit à un renouvellement de la recherche académique récente sur les mesures de risque financier, à savoir : (i) le trading à haute fréquence (HFT : *High-Frequency Trading*), et (ii) le risque systémique.

Au cours des dernières années, le trading à haute fréquence a reçu une attention particulière de la part du public, des régulateurs et des chercheurs, du fait de son importance dans l'activité et le fonctionnement actuel des marchés financiers. Ainsi, en 2010, le HFT représentait 56% du total des transactions sur les actions aux Etats-Unis et 38% en Europe (Grant, 2010). La recherche académique a mis en évidence à la fois les aspects positifs et négatifs de ce type d'activité. Menkveld (2013) montre l'importance des traders à haute fréquence sur les marchés. Ils jouent en effet un rôle moderne de teneurs de marchés en fournissant un flux presque continu de liquidité. Brogaard (2010) soutient que le HFT réduit les coûts de transaction, favorise la révélation des prix et l'efficacité des marchés en interconnectant différentes plateformes financières. D'autre part, de nombreuses études empiriques et théoriques mettent en évidence les aspects négatifs de l'activité de HFT. Un des principaux reproches est que certaines stratégies de HFT peuvent conduire à une manipulation des cours des titres à court terme (Wah et Wellman, 2013). En outre, ce type d'activité peut générer une volatilité importante au niveau du marché, affectant par conséquent sa stabilité. Cela peut déstabiliser la structure du marché ou conduire à une réaction en chaîne générant du risque systémique (Zhang, 2010; Hanson, 2012). Kirilenko et al. (2011) associent le HFT à des périodes de forte illiquidité menant à des chutes financières brutales de type *flash crash*. L'exemple le plus célèbre est le flash crash du 6 mai 2010, lorsque l'indice Dow Jones a perdu plus de 600 points, puis est revenu à son niveau initial, le tout en quelques minutes. Ces enjeux liés au HFT plaident en faveur du développement de mesures intra-journalières de risque de marché, spécialement conçues pour ce type d'activité. Ces mesures pourraient être utilisées par les gestionnaires de portefeuilles à haute fréquence dans le but de fixer des limites de trading, mais aussi par les autorités réglementaires afin d'évaluer les impacts négatifs potentiels du HFT. Les règlements actuels incluent déjà l'interdiction de certaines stratégies de trading, ainsi que l'application d'une taxe sur chaque opération effectuée.

Parallèlement, suite à la crise financière récente, le débat sur la régulation financière a connu un profond renouvellement avec l'émergence d'une approche macroprudentielle et la notion de risque systémique. De très nombreux travaux académiques ont été consacrés à cette notion de risque systémique, que ce soit concernant sa définition, sa mesure ou les aspects de régulation (voir Bisias et al. 2012; Benoit et al. 2013,

pour une présentation générale). Un intérêt particulier repose sur l'identification des institutions financières qui contribuent le plus au risque global du système financier, les soi-disant institutions financières d'importance systémique (SIFI : *Systemically Important Financial Institution*) ou les banques d'importance systémique globale (G-SIB : *Global Systemically Important Banks*). Puisque ces institutions représentent une menace majeure pour le système, les régulateurs du monde entier leur ont imposé une surveillance plus étroite, des exigences de fonds propres supplémentaires, et des réserves de liquidité.

Il y a deux façons de mesurer la contribution d'une institution financière donnée au risque global du système. La première consiste à utiliser des données propriétaires qui se trouvent dans les bilans des banques et des institutions financières (Gouriéroux et al. 2012, Basel Committee on Banking Supervision, 2011, etc.) La seconde approche repose sur des données publiques de marché, telles que les rendements des actions, les prix des options, ou les spreads de CDS. Ces données sont censées refléter toute l'information disponible sur les entreprises cotées en bourse. Les mesures de risque systémique basées sur des données de marché les plus connues sont la MES (*Marginal Expected Shortfall*) proposée par Acharya et al. (2010) et la SRISK introduite par Acharya et al. (2012), Brownlees et Engle (2012) et Jondeau et Rockinger (2013).³ La MES correspond à la perte en capitaux propres lorsque le rendement du marché tombe en dessous d'un certain seuil sur un horizon temporel donné. Les institutions financières avec les MES (en valeur absolue) les plus élevées contribuent le plus à la baisse du marché ; ces institutions sont donc, d'après cette mesure, les institutions les plus systématiquement risquées. La SRISK mesure le besoin espéré en capital, mesuré conditionnellement à une situation de crise financière. La banque avec le plus grand besoin de capital enregistré précisément au cours de la crise, doit être considérée comme la plus risquée systématiquement.

L'un des principaux avantages de ces mesures est qu'elles peuvent être estimées en temps réel (voir par exemple les sites Internet de la V-Lab, NYU Stern et le Centre pour la gestion des risques, HEC Lausanne). En outre, le classement des entreprises issu de la comparaison des SRISKS estimées est très similaire au classement fourni par la BIS (*Bank for International Settlements*) (Brownlees et Engle, 2012). Toutefois, ces mesures ont aussi des limites. En particulier, l'approche marginale utilisée pour la MES tend à identifier comme SIFIs les entreprises qui sont les plus interconnectées (Benoit et al., 2013). Cette mesure ne tient pas compte de la taille des institutions financières et, par conséquent, elle favorise la logique *Too Interconnected To Fail* au détriment de la logique *Too Big To Fail*. En plus, la somme des MES associées aux banques dans le système n'est pas égale à la perte globale du système, généralement mesurée par une espérance de perte (ES : *Expected Shortfall*), ce qui rend son interprétation plutôt difficile. Par contre, la SRISK tient compte de la taille de l'entreprise via la capitalisation boursière et la valeur comptable de la dette. Cependant, la construction de la SRISK conduit à mélanger des valeurs comptables (le montant de la dette totale) avec des valeurs de marché dans la même mesure du risque systémique.

IV. Contributions

Dans ce contexte, notre recherche se concentre sur les mesures du risque financier et la modélisation de la volatilité. L'objectif général de cette thèse est : (i) de proposer de nouvelles techniques pour mesurer à la fois le risque systémique et le risque à haute fréquence, et (ii) d'appliquer et d'améliorer les outils économétriques de modélisation et de prévision de la volatilité. Ce travail se compose de quatre chapitres (articles) qui peuvent être étudiés indépendamment l'un de l'autre.

Les deux premiers chapitres traitent des questions liées à la modélisation et la prévision des mesures du risque à haute fréquence et du risque systémique. Le deuxième chapitre propose une nouvelle mesure du risque systémique utilisée pour identifier les institutions financières d'importance systémique (SIFIs). Basée sur une approche spécifique, cette mesure originale permet de décomposer le risque global du

³ On peut également citer (entre autres) la ΔCoVaR de Adrian et Brunnermeier (2011), la SES introduite par Acharya et al. (2010) et les mesures économétriques de connectivité basées sur l'analyse en composantes principales et la causalité de Granger en réseau proposées par Billio et al. (2012).

système financier tout en tenant compte des caractéristiques de l'entreprise. Le troisième chapitre propose des mesures du risque de marché intra-journalier dans le contexte particulier des données à haute fréquence irrégulièrement espacées dans le temps (*tick-by-tick*).

Les deux autres chapitres sont consacrés aux méthodes d'estimation et de prévision de la volatilité incluant directement des données à haute fréquence ou des mesures réalisées de volatilité. Dans le quatrième chapitre, nous cherchons à déterminer, dans le contexte des modèles de mélange de fréquence (MIDAS), si des régresseurs à haute fréquence améliorent les prévisions de la volatilité à basse fréquence. Une question liée est de savoir s'il existe une fréquence d'échantillonnage optimale en termes de *prévision*, et non de *mesure* de la volatilité. Le cinquième chapitre propose une version robuste aux jumps du modèle Realized GARCH. L'application porte sur la crise financière récente, et vise à identifier et analyser les événements ayant induit ces sauts pendant la crise.

Dans la suite, je détaille le contenu de chaque chapitre.

Chapitre 2 : Which Are the SIFIs? A Component Expected Shortfall Approach to Systemic Risk

Le deuxième chapitre, intitulé "*Which Are the SIFIs? A Component Expected Shortfall Approach to Systemic Risk*", propose une méthode originale et parcimonieuse permettant d'identifier les institutions financières d'importance systémique.⁴ Développée par analogie avec l'analyse du risque de portefeuille et le concept de *Component Value-at-Risk* (Jorion, 2007), notre mesure de risque systémique, appelée CES (*Component Expected Shortfall*), quantifie la contribution de chaque entreprise au risque global du système mesuré par l'ES.

La mesure CES présente plusieurs avantages. Tout d'abord, cette mesure permet d'identifier les SIFIs en classant directement les institutions en fonction de leur degré de risque. Plus sa valeur est élevée, plus la contribution de l'institution au risque systémique est importante. En plus, la CES est normalisée par la valeur totale du risque du système de manière à refléter la contribution en pourcentage d'une entreprise au risque global (comme cela se fait généralement dans l'analyse du risque du portefeuille). Cette caractéristique de la CES simplifie significativement son interprétation. Deuxièmement, de façon similaire à la SRISK, la mesure CES tient compte de la taille relative de chaque institution dans le système financier. Elle peut être ainsi considérée comme une mesure hybride du risque systémique qui combine les deux logiques *Too Interconnected To Fail* et *Too Big To Fail*. Troisièmement, contrairement à la SRISK, la CES est uniquement basée sur des données de marché, et encore plus important, elle ne dépend pas de valeurs comptables telles que la dette. Notre mesure repose sur des données publiques, accessibles à haute fréquence (quotidienne, voir au-delà), qui incorporent des informations utiles pour prévoir le risque systémique. Enfin, la CES peut être utilisée non seulement pour évaluer la contribution d'une entreprise au risque global du système à une date précise, mais aussi pour prévoir sa contribution au cours d'une période donnée. En termes de politique réglementaire, cette mesure est bien adaptée pour des décisions orientées vers la prévention de l'accumulation du risque systémique.

La capacité de la mesure CES à identifier les institutions financières les plus risquées est évaluée dans une application empirique basée sur des données similaires à celles employées par Acharya et al. (2010), Brownlees et Engle (2012) et Acharya et al. (2012). Nous montrons que notre mesure classe de manière assez précise les institutions financières en fonction de leur degré de risque. En effet, le classement met en avant les entreprises qui ont une contribution significative au risque total du système financier, et ce classement est comparable à ceux proposés par Brownlees et Engle (2012) ou publiés par le FSB (*Financial Stability Board*) en 2011. Encore plus important, il y a deux caractéristiques que nous trouvons assez

⁴ Ce chapitre est issu de l'article "*Which Are the SIFIs? A Component Expected Shortfall Approach to Systemic Risk*", Banulescu et Dumitrescu, *Journal of Banking and Finance*, 2015.

importantes du point de vue d'un régulateur financier. Tout d'abord, le risque du système financier apparaît très concentré, ce qui indique que sa stabilité peut être améliorée en imposant un contrôle plus serré à un petit nombre d'établissements. Deuxièmement, les concentrations des institutions les plus risquées sont stables dans le temps, voir presque les mêmes. Cette stabilité est particulièrement importante car les règles imposées aux institutions financières (systémiques ou non systémiques) ne doivent pas être révisées trop fréquemment. Une analyse de robustesse montre que la CES est non seulement en mesure d'identifier les SIFIs, mais aussi de mesurer avec précision le niveau du risque conditionnellement à l'état de l'économie (crise/calme).

Chapitre 3 : High-Frequency Risk Measures

Le troisième chapitre, intitulé "*High-Frequency Risk Measures*", propose des mesures du risque de marché à haute fréquence dans le cas des données échantillonnées de manière irrégulière dans le temps (*tick-by-tick*).⁵ Ces mesures sont spécialement conçues pour la gestion du risque dans le contexte spécifique du trading à haute fréquence (HFT).

Tel que discuté précédemment, compte tenu de l'importance croissante du HFT, le risque de gestion lié à ce type d'activité est essentiel à la stabilité des marchés financiers. Cependant, la gestion des risques à haute fréquence est fondamentalement différente de la gestion des risques proposée dans le cadre des réglementations de Bâle ou des pratiques de la gestion du risque généralement adoptées par les institutions financières. Les différences sont liées principalement aux caractéristiques du HFT (Gomber et Haferkorn, 2013) : (i) des ordres rapides d'annulation, (ii) des positions de détention très courtes, (iii) des marges par transaction très faibles et (iv) pas de positions significatives à la fin de la journée. En particulier, les mesures journalières de risque ne sont pas pertinentes dans le contexte HFT. Par exemple, le risque engendré par un algorithme d'arbitrage statistique ou une stratégie de momentum à court terme peut être élevé pendant une courte période de temps durant la journée, alors qu'aucun risque n'est reporté à la fin de la journée. Le même problème se pose si l'on retient une évaluation du risque intra-quotidien, mais sur une fréquence régulière (par exemple toutes les heures, toutes les 10 minutes, etc.).

Pour ces raisons, dans ce chapitre, nous proposons des mesures du risque de marché spécialement conçues pour des données à haute fréquence irrégulièrement espacées dans le temps (des données "tick-by-tick"). On distingue trois concepts de la Value-at-Risk (VaR) : la VaR totale, la VaR marginale (ou par unité de temps) et la VaR instantanée. La VaR totale correspond à la perte maximale prévue pour l'horizon temporel avant la prochaine transaction, tandis que la VaR marginale fait référence à l'horizon de la prochaine unité de temps (généralement définie en secondes). Le troisième concept correspond à la perte maximale prévue pour l'intervalle de temps entre t_i et $t_{i+\Delta}$, défini par une augmentation infinitésimale et déterministe du temps, Δ , où t_i représente le temps de la dernière transaction.

En considérant des données tick-by-tick, le risque de marché est évidemment lié à la durée entre deux opérations consécutives (voir Diamond et Verrecchia, 1987; Easley et O'Hara, 1992). En conséquence, nos mesures du risque de marché sont complétées par une mesure de risque de durée, à savoir la TAR (*Time-at-Risk*). Cette mesure a été initialement introduite par Ghysels et al. (2004a) et elle fournit une évaluation du risque associé à la durée entre la transaction actuelle et la suivante.

Un des avantages de ces mesures est qu'elles peuvent être étendues à tout type d'événement de microstructure du marché en considérant un sous-ensemble de transactions ayant des caractéristiques spécifiques. Par exemple, nous pouvons définir la VaR et la TaR pour des transactions associées à des changements significatifs des prix (*price event*) ou un volume minimum (*volume event*). Cette caractéristique différencie notre mesure de la VaR intra-journalière proposée par Giot (2005) qui n'est valable que pour des changements significatifs des prix.

⁵ Ce chapitre est issu de l'article "High-Frequency Risk Measures", Banulescu, Colletaz, Hurlin, Tokpavi, 2014.

Sachant que la VaR et la TaR sont définis comme les quantiles de deux distributions conditionnelles (c'est-à-dire, la distribution des rendements intra-journaliers et la distribution des durées), nous modélisons leur dynamique jointe. Pour ce faire, nous considérons un modèle EACD (Engle et Russell, 1998) pour le processus conditionnel de la durée, et un modèle GARCH avec des paramètres variant dans le temps (Ghysels et Jasiak, 1998) pour la volatilité conditionnelle. Nous proposons un algorithme de prévision de la TaR et la VaR intra-journalières en utilisant une approche semi-paramétrique similaire à celle examinée par Manganeli et Engle (2001). Aucune hypothèse distributionnelle (sauf celles requises par la méthode d'estimation du quasi-maximum de vraisemblance) n'est faite en ce qui concerne les distributions conditionnelles des séries de durées et de rendements.

Nos mesures de risque à haute-fréquence sont calculées et évaluées à l'horizon de la prochaine transaction, et contrairement à Dionne et al. (2009), la durée actuelle et le rendement sont (à la fois) considérés comme stochastiques. La validité des prévisions de la VaR et la TaR est examinée à l'aide d'un ensemble de procédures de backtesting compatibles avec les données irrégulièrement espacées, à savoir les tests proposés par Christoffersen (1998), Christoffersen et Pelletier (2004) et Candelon et al. (2011). Contrairement aux études précédentes de Giot (2005), Dionne et al. (2009) et Dionne et al. (2014), les prévisions de la VaR ne sont pas réajustées à des intervalles de temps fixes pour l'objet de backtesting.

Les performances de nos mesures sont vérifiées empiriquement sur la base de deux actions (Bank of America et Microsoft) et d'un ETF (*Exchange-Traded Fund*) basé sur l'indice S&P 500. Les résultats montrent que la VaR et la TaR intra-journalières capturent avec précision la dynamique de la volatilité et de la durée de ces trois actifs. En outre, les résultats de backtesting sont valides et ils restent robustes au choix de la date qui sépare les parties de l'échantillon utilisées pour l'estimation et la prévision (Hansen et Timmermann, 2012).

Chapitre 4 : Do We Need Intradaily Data to Forecast Daily Volatility?

Le quatrième chapitre, intitulé "*Do We Need Intradaily Data to Forecast Daily Volatility?*", considère des modèles de régression MIDAS (Ghysels et al., 2006) pour examiner l'influence de la fréquence d'échantillonnage de prédicteurs sur la qualité des prévisions de la volatilité quotidienne.⁶ L'objectif principal est de vérifier si l'information incorporée par les prédicteurs à haute fréquence améliore la qualité des prévisions de volatilité, et si oui, s'il existe une fréquence d'échantillonnage optimale de ces prédicteurs.

Cette problématique doit être distinguée de la discussion bien documentée sur la fréquence d'échantillonnage optimale des rendements utilisés pour calculer les estimateurs réalisés de la volatilité quotidienne (voir Hansen et Lunde, 2004; Ait-Sahalia et Mancini, 2008; Garcia et Meddahi, 2006; Ghysels et al, 2006, entre autres). Dans ce chapitre, nous nous concentrons sur la fréquence d'échantillonnage optimale à des fins de *prévision* de la volatilité, et non pour la *mesure* de la volatilité.

Pour répondre à ces questions, nous proposons tout d'abord une étude basée sur des simulations Monte Carlo. Nous générons, à partir d'un processus de diffusion sans bruit, des séries de rendements à différentes fréquences d'échantillonnage et des mesures réalisées de la volatilité quotidienne et cela en utilisant le même ensemble de paramètres structurels en temps continu. Ensuite, nous appliquons des spécifications MIDAS simples dans lesquelles la volatilité quotidienne réalisée est prédite par des valeurs passées des rendements intra-journaliers au carré. Ces rendements sont échantillonnés à des fréquences différentes allant d'une minute à trois heures. Les principales conclusions de l'exercice de simulation sont en faveur de l'utilisation des prédicteurs à la fréquence la plus élevée possible. Encore plus important, les gains sont statistiquement significatifs, puisque le test MCS (Model Confidence Set, Hansen et al., 2011) choisit presque toujours le modèle MIDAS basé sur la fréquence la plus élevée comme étant le plus performant en termes de capacité de prévision.

⁶ Ce chapitre est issu de l'article, "Do We Need Ultra-High Frequency Data to Forecast Variances?", Banulescu, Candelon, Hurlin, Laurent, 2014.

Cependant, cette première analyse néglige l'existence de la périodicité intra-journalière ainsi que les potentielles sources de bruit de microstructure des marchés. Pour cette raison, nous utilisons la même approche méthodologique aussi sur deux actifs très liquides (S&P 500 et Microsoft) et nous comparons leurs prévisions de volatilité. Selon le test MCS, les prévisions de volatilité sont statistiquement équivalentes pour une fréquence d'échantillonnage des prédicteurs allant de cinq minutes à une heure. En effet, nous identifions une sorte de "mur" de haute fréquence à cinq minutes au-delà duquel les prévisions de volatilité se détériorent. Ce résultat est lié à la présence de deux caractéristiques importantes dans les données à haute fréquence, à savoir la périodicité intra-journalière et les sauts. Un compromis devrait donc être fait entre l'information contenue dans les données à très haute fréquence et la présence du bruit de microstructure qui pollue ce type de données menant à une détérioration de la qualité des prévisions de volatilité.

Lorsque le modèle MIDAS est appliqué à des prédicteurs corrigés de la saisonnalité et des sauts (Hecq et al. 2012; Lee et Mykland, 2008; Boudt et al. 2011), la conclusion en faveur de l'utilisation de la fréquence la plus élevée reste valable. En outre, les régressions MIDAS fondées sur des variables explicatives intra-journalières ajustées de la périodicité ou des sauts, permettent d'obtenir des prévisions de volatilité comparables, voire meilleures, que celles d'autres modèles concurrents.

Chapitre 5 : Volatility During the Financial Crisis Through the Lens of High Frequency Data : A Realized GARCH Approach

Le cinquième chapitre, intitulé "*Volatility During the Financial Crisis Through the Lens of High Frequency Data : A Realized GARCH Approach*", fournit une analyse des chocs majeurs de volatilité apparus au cours de la crise financière récente.⁷

La contribution économétrique du chapitre repose sur le développement d'une version (i) asymétrique et (ii) robuste aux sauts, du modèle Realized GARCH initialement introduit par Hansen et al. (2012). Premièrement, le modèle proposé prend en compte l'asymétrie des effets des chocs, à la manière d'un modèle de type EGARCH. Deuxièmement, notre innovation principale par rapport au modèle Realized GARCH standard est similaire à l'intuition de Harvey (2013) et des modèles GAS (*Generalized Autoregressive Score*). Elle consiste à introduire une fonction score d'une distribution à queues plus épaisses dans le modèle GARCH classique afin de limiter les effets de chocs extrêmes (jumps) sur la dynamique de la volatilité conditionnelle et de la volatilité réalisée. Les résultats empiriques montrent que le nouveau modèle se comporte mieux que les approches classiques en termes de réponse aux chocs extrêmes sur la volatilité et réduit la persistance de leurs effets dans le temps.

Les séries de la volatilité journalière ainsi que les chocs sur la volatilité sont ensuite estimées par le modèle Realized GARCH robuste. Les plus grands chocs sur la volatilité sont utilisés pour identifier des événements majeurs durant les périodes de crise, ainsi que leurs causes. Nous nous concentrons en particulier sur la crise financière récente (2007-2009) et identifions les principaux événements économiques/financiers/sociaux au cours de cette période ainsi que leur impact direct sur la volatilité. Cette analyse est effectuée en utilisant une technique économétrique qui permet d'identifier la datation précise des chocs à partir des données à haute fréquence.

Le plus grand choc sur la volatilité conditionnelle a été identifié le 27 Février 2007. L'analyse approfondie des rendements intra-journaliers montre que cette hausse anormale de la volatilité est due à un problème informatique dans le système de trading. Néanmoins, les autres jours avec de fortes hausses de volatilité ont été caractérisés dans la plupart par des décisions gouvernementales, des nouvelles économiques et financières ou des événements sociaux.

⁷ Ce chapitre est issu de l'article "*Volatility During the Financial Crisis Through the Lens of High Frequency Data : A Realized GARCH Approach*", Banulescu, Hansen, Huang, Matei, 2014.

Références bibliographiques

- [1] Acharya, V., Engle, R., Richardson, M., 2012. Capital shortfall: A new approach to ranking and regulating systemic risks. *The American Economic Review* 102 (3), 59–64.
- [2] Acharya, V., Pedersen, L., Philippon, T., Richardson, M., 2010. Measuring systemic risk. CEPR Discussion Paper No. DP8824.
- [3] Adrian, T., Brunnermeier, M. K., 2011. Covar. Working paper.
- [4] Aït-Sahalia, Y., 2002. Telling from discrete data whether the underlying continuous-time model is a diffusion. *The Journal of Finance* 57 (5), 2075–2112.
- [5] Aït-Sahalia, Y., Cacho-Diaz, J., Laeven, R. J., 2013. Modeling financial contagion using mutually exciting jump processes. Tech. rep., National Bureau of Economic Research.
- [6] Aït-Sahalia, Y., Mancini, L., 2008. Out of sample forecasts of quadratic variation. *Journal of Econometrics* 147, 17–33.
- [7] Aït-Sahalia, Y., Mykland, P. A., Zhang, L., 2005. How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial studies* 18 (2), 351–416.
- [8] Ait-Sahalia, Y., Yu, J., 2009. High frequency market microstructure noise estimates and liquidity measures. Tech. Rep. 1, Institute of Mathematical Statistics.
- [9] Alper, C. E., Fendoglu, S., Saltoglu, B., 2008. Forecasting stock market volatilities using MIDAS regressions: An application to the emerging markets. Working Paper.
- [10] Andersen, T., Bollerslev, T., Diebold, F. X., 2007a. Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility. *Review of Economics and Statistics* 89, 701–720.
- [11] Andersen, T., Dobrev, D., Schaumburg, E., 2008. Duration-based volatility estimation. Working paper.
- [12] Andersen, T. G., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance* 4, 115–158.
- [13] Andersen, T. G., Bollerslev, T., 1998a. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39 (4), 885–905.
- [14] Andersen, T. G., Bollerslev, T., 1998b. Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance* 53 (1), 219–265.
- [15] Andersen, T. G., Bollerslev, T., Das, A., 1998. Testing for market microstructure effects in intraday volatility: A reassessment of the Tokyo FX experiment. Working paper, NBER.
- [16] Andersen, T. G., Bollerslev, T., Diebold, F. X., 2007b. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics* 89 (4), 701–720.
- [17] Andersen, T. G., Bollerslev, T., Diebold, F. X., Ebens, H., 2001a. The distribution of realized stock return volatility. *Journal of Financial Economics* 61 (1), 43–76.
- [18] Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2000. Exchange rate return standardized by realized volatility are (nearly) Gaussian. *Multinational Finance Journal* 4 (3&4), 159–179.
- [19] Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2001b. The distribution of exchange rate volatility. *Journal of the American Statistical Association* 96 (453), 42–55, correction published in 2003, volume 98, page 501.
- [20] Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2003. Modeling and forecasting realized volatility. *Econometrica* 71 (2), 579–625.
- [21] Andreou, E., Ghysels, E., 2002. Rolling-sample volatility estimators: Some new theoretical, simulation, and empirical results. *Journal of Business & Economic Statistics* 20 (3), 363–376.
- [22] Aragón, V., Salvador, E., 2010. Re-examining the risk-return relationship: The influence of financial crisis (2007-2009). Working Paper.
- [23] Bai, X., Russell, J. R., Tiao, G. C., 2004. Effects of non-normality and dependence on the precision of variance estimates using high-frequency financial data. Working paper, University of Chicago,

GSB.

- [24] Bajgrowicz, P., Scaillet, O., 2011. Jumps in high-frequency data: Spurious detections, dynamics, and news. Working Paper.
- [25] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N., 2008. Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* 76, 1481–536.
- [26] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N., 2009. Realized kernels in practice: Trades and quotes. *The Econometrics Journal* 12 (3), C1–C32.
- [27] Barndorff-Nielsen, O. E., Shephard, N., 2002. Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society B* 64, 253–280.
- [28] Barndorff-Nielsen, O. E., Shephard, N., 2004b. Power and bipower variation with stochastic volatility and jumps (with discussion). *Journal of Financial Econometrics* 2, 1–48.
- [29] Barndorff-Nielsen, O. E., Shephard, N., 2006. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of financial Econometrics* 4 (1), 1–30.
- [30] Barndorff-Nielsen, O. E., Shephard, N., 2007. Variation, jumps and high frequency data in financial econometrics. *Advanced in Economics and Econometrics. Theory and Applications*, 328–372.
- [31] Bauwens, L., Hafner, C. M., Laurent, S., 2012. *Handbook of volatility models and their applications*. Vol. 3. John Wiley & Sons.
- [32] Belloni, A., Chen, D., Chernozhukov, V., Hansen, C., 2012a. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* 80 (6), 2369–2429.
- [33] Belloni, A., Chernozhukov, V., Hansen, C., 2012b. Inference on treatment effects after selection amongst high-dimensional controls. Working Paper.
- [34] Belloni, A., Chernozhukov, V., et al., 2011. 11-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics* 39 (1), 82–130.
- [35] Benoit, S., Colletaz, G., Hurlin, C., Pérignon, C., 2013. A theoretical and empirical comparison of systemic risk measures. Working Paper.
- [36] Bickel, P. J., Ritov, Y., Tsybakov, A. B., 2009. Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, 1705–1732.
- [37] Billio, M., Getmansky, M., Lo, A. W., Pelizzon, L., 2012. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics* 104 (3), 535–559.
- [38] Bisias, D., Flood, M., Lo, A. W., Valavanis, S., 2012. A survey of systemic risk analytics. *Annu. Rev. Financ. Econ.* 4 (1), 255–296.
- [39] Bollerslev, T., 1986. Generalized autoregressive heteroskedasticity. *Journal of Econometrics* 31, 307–327.
- [40] Bollerslev, T., Law, T. H., Tauchen, G., 2008. Risk, jumps, and diversification. *Journal of Econometrics* 144 (1), 234–256.
- [41] Bollerslev, T., Todorov, V., Li, S. Z., 2013. Jump tails, extreme dependencies, and the distribution of stock returns. *Journal of Econometrics* 172 (2), 307–324.
- [42] Boudt, K., Cornelissen, J., Croux, C., Laurent, S., 2012. Nonparametric tests for intraday jumps: Impact of periodicity and microstructure noise. *Handbook of Volatility Models and Their Applications*, 447–463.
- [43] Boudt, K., Croux, C., Laurent, S., 2011. Robust estimation of intraweek periodicity in volatility and jump detection. *Journal of Empirical Finance* 18 (2), 353–367.
- [44] Brogaard, J., 2010. High frequency trading and its impact on market quality. Working Paper.
- [45] Brownlees, C. T., Engle, R., 2012. Volatility, correlation and tails for systemic risk measurement. Working Paper, NYU-Stern.
- [46] Candelon, B., Colletaz, G., Hurlin, C., Tokpavi, S., 2011. Backtesting value-at-risk: a GMM duration-based test. *Journal of Financial Econometrics* 9 (2), 314–343.
- [47] Candelon, B., Hurlin, C., Tokpavi, S., 2012. Sampling error and double shrinkage estimation of minimum variance portfolios. *Journal of Empirical Finance* 19 (4), 511–527.

- [48] Candes, E., Tao, T., 2007. The Dantzig selector: Statistical estimation when p is much larger than n . *The Annals of Statistics*, 2313–2351.
- [49] Carr, P., Wu, L., 2003. What type of process underlies options? A simple robust test. *The Journal of Finance* 58 (6), 2581–2610.
- [50] Chen, X., Ghysels, E., 2011. News-good or bad-and its impact on volatility predictions over multiple horizons. *Review of Financial Studies* 24 (1), 46–81.
- [51] Chen, X., Ghysels, E., Wang, F., 2011. HYBRID-GARCH a generic class of models for volatility predictions using mixed frequency data. Working paper.
- [52] Chernozhukov, V., Chetverikov, D., Kato, K., 2013. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics* 41 (6), 2786–2819.
- [53] Chopra, V. K., Ziemba, W. T., 1993. The effect of errors in means, variances, and covariances on optimal portfolio choice. *The journal of portfolio management* 19 (2), 6–11.
- [54] Christoffersen, P., Pelletier, D., 2004. Backtesting value-at-risk: A duration-based approach. *Journal of Financial Econometrics* 2, 84–108.
- [55] Christoffersen, P. F., 1998. Evaluating interval forecasts. *International economic review*, 841–862.
- [56] Corsi, F., 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7 (2), 174–196.
- [57] Dacorogna, M. M., Gencay, R., Müller, U., Olsen, R. B., Pictet, O. V., 2001. *An introduction to high-frequency finance*. Academic Press, London.
- [58] Diamond, D. W., Verrecchia, R. E., 1987. Constraints on short-selling and asset price adjustments to private information. *Journal of Financial Economics* 18, 277–311.
- [59] Dionne, G., Duchesne, P., Pacurar, M., 2009. Intraday value at risk (IVaR) using tick-by-tick data with application to the toronto stock exchange. *Journal of Empirical Finance* 16 (5), 777–792.
- [60] Dionne, G., Pacurar, M., Zhou, X., 2014. Liquidity-adjusted intraday value at risk modeling and risk management: an application to data from deutsche börse. Working Paper.
- [61] Dungey, M., Hvozydk, L., 2012. Cojumping: Evidence from the us treasury bond and futures markets. *Journal of Banking & Finance* 36 (5), 1563–1575.
- [62] Dungey, M., McKenzie, M., Smith, L. V., 2009. Empirical evidence on jumps in the term structure of the us treasury market. *Journal of Empirical Finance* 16 (3), 430–445.
- [63] Easley, D., O’Hara, M., 1992. Time and the process of security price adjustment. *The Journal of Finance* 47 (2), 577–605.
- [64] Engle, R. F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* 45, 987–1007.
- [65] Engle, R. F., 2002a. Dynamic conditional correlation - a simple class of multivariate GARCH models. *Journal of Business & Economic Statistics* 20 (3), 339–350.
- [66] Engle, R. F., Gallo, G., 2006. A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics* 131, 3–27.
- [67] Engle, R. F., Russell, J. R., 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* 66 (5), 1127–1163.
- [68] Evans, K. P., 2011. Intraday jumps and us macroeconomic news announcements. *Journal of Banking & Finance* 35 (10), 2511–2527.
- [69] Financial Stability Board, 2010. Reducing the moral hazard posed by systemically important financial institutions. Basel October 20.
- [70] Financial Stability Board, 2011. Policy measures to address systemically important financial institutions. On-line paper: http://www.financialstabilityboard.org/publications/r_111104bb.pdf.
- [71] Forsberg, L., Ghysels, E., 2007. Why do absolute returns predict volatility so well? *Journal of Financial Econometrics* 5, 31–67.
- [72] Gagliardini, P., Ossola, E., Scaillet, O., 2013. Time-varying risk premium in large cross-sectional equity datasets. Working Paper.

- [73] Garcia, R., Meddahi, N., 2006. Comment on realized variance and market microstructure noise. *Journal of Business & Economic Statistics* 24, 184–191.
- [74] Ghysels, E., Gouriéroux, C., Jasiak, J., 2004a. Stochastic volatility duration models. *Journal of Econometrics* 119 (2), 413–433.
- [75] Ghysels, E., Jasiak, J., 1998. GARCH for irregularly spaced financial data: The ACD-GARCH model. *Studies in Nonlinear Dynamics and Econometrics* 2, 133–149.
- [76] Ghysels, E., Santa-Clara, P., Valkanov, R., 2004b. The MIDAS touch: Mixed data sampling regression models. Working paper.
- [77] Ghysels, E., Santa-Clara, P., Valkanov, R., 2005. There is a risk-return trade-off after all. *Journal of Financial Economics* 76 (3), 509–548.
- [78] Ghysels, E., Santa-Clara, P., Valkanov, R., 2006. Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics* 131 (1), 59–95.
- [79] Ghysels, E., Sinko, A., 2011. Volatility forecasting and microstructure noise. *Journal of Econometrics* 160 (1), 257–271.
- [80] Ghysels, E., Sinko, A., Valkanov, R., 2007. MIDAS regressions: Further results and new directions. *Econometric Reviews* 26 (1), 53–90.
- [81] Ghysels, E., Valkanov, R., 2012. Forecasting volatility with MIDAS. In L. Bauwens, C. Hafner, and S. Laurent (Eds.), *Handbook of Volatility Models and Their Applications*, 383–401.
- [82] Gilder, D., Shackleton, M. B., Taylor, S. J., 2014. Cojumps in stock prices: Empirical evidence. *Journal of Banking & Finance* 40, 443–459.
- [83] Giot, P., 2005. Market risk models for intraday data. *The European Journal of Finance* 11 (4), 309–324.
- [84] Gomber, P., Haferkorn, D. W.-I. M., 2013. High-frequency-trading. *Business & Information Systems Engineering* 5 (2), 97–99.
- [85] Gouriéroux, C., Héam, J.-C., Monfort, A., 2012. Bilateral exposures and systemic solvency risk. *Canadian Journal of Economics/Revue canadienne d'économie* 45 (4), 1273–1309.
- [86] Grant, J., 2010. High-frequency trading: Up against a bandsaw. *Financial Times*, available at September 10.
- [87] Griffin, J. E., Oomen, R. C. A., 2008. Sampling returns for realized variance calculations: Tick time or transaction time? *Econometric Reviews* 27, 220–253.
- [88] Hansen, P. R., Horel, G., 2009a. Quadratic variation by Markov chains. Working paper.
- [89] Hansen, P. R., Horel, G., 2009b. Quadratic variation by markov chains. Working paper.
- [90] Hansen, P. R., Huang, Z., 2012. Exponential GARCH modeling with realized measures of volatility. Working paper.
- [91] Hansen, P. R., Huang, Z., Shek, H., 2012. Realized GARCH: A joint model of returns and realized measures of volatility. *Journal of Applied Econometrics* 27, 877–906.
- [92] Hansen, P. R., Lunde, A., 2004. Technical appendix: An unbiased measure of realized variance. Brown University, 2004.
- [93] Hansen, P. R., Lunde, A., Nason, J. M., 2011. The model confidence set. *Econometrica* 79, 456–497.
- [94] Hansen, P. R., Timmermann, A., 2012. Choice of sample split in out-of-sample forecast evaluation. Working paper.
- [95] Hanson, T. A., 2012. The effects of high frequency traders in a simulated market. Working Paper.
- [96] Harris, L., 1986. A transaction data study of weekly and intradaily patterns in stock returns. *Journal of Financial Economics* 16, 99–117.
- [97] Harvey, A. C., 2013. *Dynamic Models for Volatility and Heavy Tails*. Cambridge University Press.
- [98] Hastie, T., Tibshirani, R., Friedman, J., 2009. *The elements of statistical learning*. Vol. 2. Springer.
- [99] Haugen, R. A., Baker, N. L., 1991. The efficient market inefficiency of capitalization-weighted stock portfolios. *The Journal of Portfolio Management* 17 (3), 35–40.
- [100] Hecq, A., Laurent, S., Palm, F. C., 2012. Common intraday periodicity. *Journal of Financial Econometrics* 10 (2), 325–353.

- [101] Huang, J., Horowitz, J. L., Wei, F., 2010. Variable selection in nonparametric additive models. *Annals of statistics* 38 (4), 2282.
- [102] Jagannathan, R., Ma, T., 2003. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance* 58 (4), 1651–1684.
- [103] Johannes, M., 2004. The statistical and economic role of jumps in continuous-time interest rate models. *The Journal of Finance* 59 (1), 227–260.
- [104] Jondeau, E., Rockinger, M., 2013. Systemic risk in europe. *Global Credit Review* 3 (01), 1–6.
- [105] Jorion, P., 2007. Value at risk: the new benchmark for managing financial risk. Vol. 2. McGraw-Hill New York.
- [106] Kirilenko, A., Kyle, A. S., Samadi, M., Tuzun, T., 2011. The flash crash: The impact of high frequency trading on an electronic market. Working Paper.
- [107] Lahaye, J., Laurent, S., Neely, C. J., 2011. Jumps, cojumps and macro announcements. *Journal of Applied Econometrics* 26 (6), 893–921.
- [108] Ledoit, O., Péché, S., 2011. Eigenvectors of some large sample covariance matrix ensembles. *Probability Theory and Related Fields* 151 (1-2), 233–264.
- [109] Ledoit, O., Wolf, M., 2004. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis* 88 (2), 365–411.
- [110] Ledoit, O., Wolf, M., et al., 2012. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics* 40 (2), 1024–1060.
- [111] Lee, S. S., Hannig, J., 2010. Detecting jumps from Lévy jump diffusion processes. *Journal of Financial Economics* 96 (2), 271–290.
- [112] Lee, S. S., Mykland, P. A., 2008. Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial studies* 21 (6), 2535–2563.
- [113] Leeb, H., Pötscher, B. M., 2008. Can one estimate the unconditional distribution of post-model-selection estimators? *Econometric Theory* 24 (02), 338–376.
- [114] Manganelli, S., Engle, R. F., 2001. Value at risk models in finance. ECB Working Paper No. 75.
- [115] Meinshausen, N., Yu, B., 2009. Lasso-type recovery of sparse representations for high-dimensional data. *The Annals of Statistics*, 246–270.
- [116] Menkveld, A. J., 2013. High frequency trading and the new market makers. *Journal of Financial Markets* 16 (4), 712–740.
- [117] Merton, R. C., 1980. On estimating the expected return on the market: An exploratory investigation. *Journal of financial economics* 8 (4), 323–361.
- [118] Mykland, P. A., Zhang, L., 2008. Inference for continuous semimartingales observed at high frequency: A general approach. Working paper.
- [119] Nelson, D. B., 1992. Filtering and forecasting with misspecified ARCH models I: Getting the right variance with the wrong model. *Journal of Econometrics* 52 (1), 61–90.
- [120] Shephard, N., Sheppard, K., 2010. Realising the future: forecasting with high-frequency-based volatility (heavy) models. *Journal of Applied Econometrics* 25 (2), 197–231.
- [121] Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B* 58, 267–288.
- [122] Todorov, V., Bollerslev, T., 2010. Jumps and betas: A new framework for disentangling and estimating systematic risks. *Journal of Econometrics* 157 (2), 220–235.
- [123] Wah, E., Wellman, M. P., 2013. Latency arbitrage, market fragmentation, and efficiency: a two-market model. In: *Proceedings of the fourteenth ACM conference on Electronic commerce*. ACM, pp. 855–872.
- [124] Wang, Y., 1995. Jump and sharp cusp detection by wavelets. *Biometrika* 82 (2), 385–397.
- [125] Wood, R. A., McInish, T. H., Ord, J. K., 1985. An investigation of transactions data for NYSE stocks. *The Journal of Finance* 40 (3), 723–739.
- [126] Zhang, F., 2010. High-frequency trading, stock volatility, and price discovery. Working Paper.
- [127] Zhang, L., 2004. Efficient estimation of stochastic volatility using noisy observations: A multi-scale

approach. Working Paper.

- [128] Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli* 12 (6), 1019–1043.