

L'ampleur de la crise financière de 2008-2009 a poussé les économistes, pris au dépourvu par celle-ci, à réorienter leurs efforts de recherche vers la prise en compte du risque comme variable fondamentale de la modélisation macroéconomique. Un champ d'études empruntant aux outils de la finance quantitative, comme l'utilisation du temps continu et des mouvements browniens comme source de risque, et avec pour objet l'explication jointe des variables macroéconomiques et financières, s'est développé (voir Brunnermeier et Sannikov, 2016, pour une revue de la littérature). Celui-ci a permis de mettre en lumière l'importance d'inclure de façon explicite le secteur financier et de tenir compte de l'existence de non-linéarités afin de comprendre les dynamiques de prix des actifs et d'investissement durant une crise financière.

Le principe premier de cette approche est de considérer le risque lui-même comme une variable endogène qui est déterminée dans le modèle par les actions des agents économiques et les frictions auxquelles ils sont confrontés. Le niveau de risque que les agents sont prêts à prendre dépend de la stabilité perçue à un moment donné et peut mener à l'accumulation de fragilités financières risquant de se matérialiser dans le futur. En outre, la relation entre le risque endogène et les prix des actifs peut présenter de fortes non-linéarités par son interaction avec le secteur financier. Par exemple, dans le modèle de Brunnermeier et Sannikov (2014), lorsque les banques sont bien capitalisées, elles jouent leur rôle dans l'absorption des chocs sur la valeur fondamentale des actifs. Cependant, après avoir atteint certains seuils, les banques ne peuvent plus absorber de nouveaux chocs et les prix des actifs doivent chuter fortement afin d'équilibrer offre et demande pour le capital risqué. Cette baisse se traduit ensuite par une volatilité endogène plus élevée et conduit à un fort mécanisme d'amplification. Cette thèse s'inscrit dans une même lignée en adoptant la méthodologie de la finance quantitative pour étudier les questions macroéconomiques liées à la dynamique, à la distribution et au prix des risques macroéconomiques.

Chapitre 1 Dans le premier chapitre, intitulé "politiques monétaires non conventionnelles et risques de liquidité de financement", nous étudions l'efficacité de différentes politiques monétaires ayant pour objectif la stabilisation du prix des actifs

lors d'une crise de liquidités. À cette fin, nous proposons un modèle de macro-finance avec des banques hétérogènes dans lequel ces dernières sont soumises à un risque de liquidité de financement.

La première contribution de ce chapitre est de fournir un modèle reliant le risque de refinancement au prix des actifs via le bilan des intermédiaires financiers. Dans notre modèle, les intermédiaires s'engagent dans une activité de transformation de liquidité en détenant des actifs moins liquides que leurs passifs. Après la réalisation d'un choc de financement négatif, un intermédiaire doit combler un déficit de financement (la différence entre les actifs illiquides et le besoin de financement après choc) soit en faisant acquisition de nouveau prêts sur les marchés monétaires (à un coût négligeable), soit en vendant des titres à un prix de vente au rabais (à un coût élevé).

En raison d'une asymétrie d'information, les prêteurs sur le marché monétaire exigent de leur contrepartie qu'elle dépose un montant suffisant de titres en garantie pour sécuriser l'opération. Cette hypothèse crée deux régimes endogènes dans l'économie. En temps normal, les banques peuvent utiliser efficacement les marchés monétaires pour éviter une vente au rabais, ou « fire-sale », coûteuse de leurs actifs. Le risque de liquidité de financement est donc faible dans ce régime et n'apparaît pas dans le prix d'équilibre des actifs à maturité longue. En période de crise, cependant, la volatilité peut forcer les appels de marge (une augmentation de la quantité de titres nécessaire pour la garantie des échanges) à devenir si élevés que le montant global des garanties disponibles est inférieur aux exigences d'accès aux marchés monétaires (un mécanisme semblable à la spirale de décote de Brunnermeier et Pedersen, 2009).

Étant donné que les intermédiaires financiers tiennent compte de leur structure de financement lorsqu'ils évaluent la valeur des actifs dans l'économie, une augmentation de ce risque de liquidité de financement a une incidence négative sur les prix des actifs via un mécanisme d'équilibre général.

Nous utilisons le modèle pour étudier l'efficacité des différentes politiques monétaires dans divers régimes de liquidités (selon un bon et mauvais fonctionnement des marchés monétaires) et sous différentes structures financières (taille du secteur ban-

caire parallèle). Suivant la littérature sur la mise en place opérationnelle des politiques monétaires, nous faisons l'hypothèse que la monnaie émise par la banque centrale (les réserves) est utilisée pour les règlements interbancaires. En détenant des réserves, les banques peuvent, par conséquent, réduire leur exposition au risque de refinancement. Nous démontrons comment cette utilité non pécuniaire à détenir des réserves rompt le résultat de neutralité monétaire de Wallace (1981) et permet aux politiques monétaires d'influencer les prix des actifs et les variables macroéconomiques en réduisant le niveau global du risque de liquidité du financement.

Ce résultat s'applique à la fois aux injections de liquidités, à la politique de prêteur en dernier ressort ainsi qu'à la politique d'achat d'actifs par la banque centrale. L'injection de réserves et la réduction du coût de l'emprunt auprès du guichet d'es-compte en urgence contribuent à atténuer le risque de liquidités dans le secteur bancaire traditionnel, mais ne parviennent pas à atteindre le secteur bancaire parallèle. Dans cette situation, la banque centrale peut acheter et détenir directement des actifs illiquides dans son bilan. Elle est ainsi capable de diminuer les quantités d'équilibre de risques de financement dans l'économie. Ceci est rendu possible par le fait qu'elle n'est pas exposée au risque de liquidités en raison de sa capacité à émettre des réserves qui sont toujours acceptées par les agents de l'économie comme instrument de règlement. Cette dernière forme de politique a l'avantage d'opérer à travers un canal d'équilibre général avec, par conséquent, une portée plus large.

Notre analyse conclut qu'en présence d'un secteur bancaire parallèle important et des marchés monétaires perturbés, l'injection de liquidités et les politiques de prêteur en dernier ressort peuvent ne pas être suffisants pour atténuer les tensions financières. Ce résultat est visible dans la partie supérieure du graphique 1 représentant respectivement le prix des titres (à gauche) et le risque de liquidité agrégé (à droite) comme fonction d'une mesure de la quantité de réserve dans l'économie. Lorsque l'ensemble du secteur bancaire possède un accès aux opérations de liquidité de la banque centrale (ligne rouge), il est possible pour cette dernière de fournir suffisamment de liquidité aux banques afin de ramener le prix des actifs au niveau de leur valeur fondamentale (ligne noire). Ce résultat n'est cependant plus atteignable lorsque le secteur bancaire

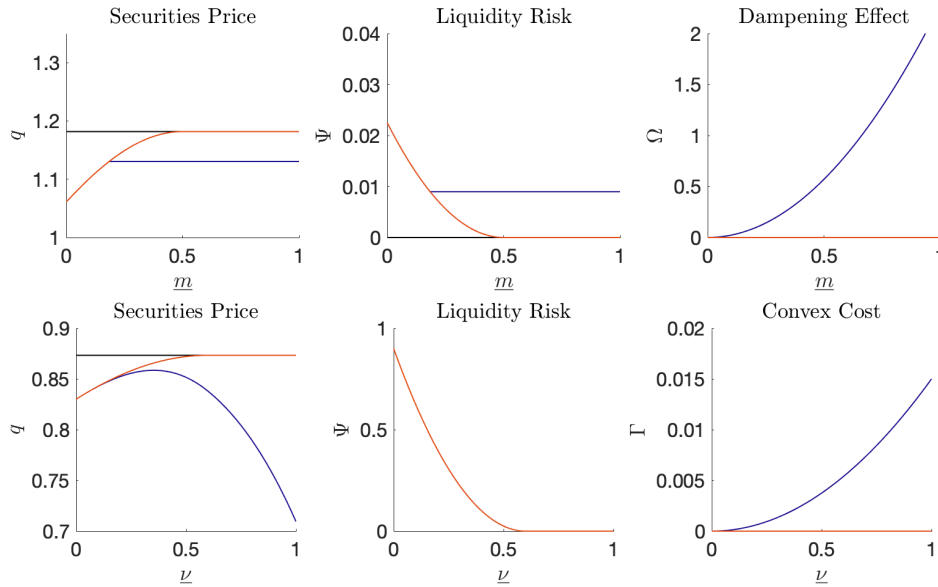


FIGURE 1 – La figure est reproduite à partir du texte principal de la thèse à titre d’illustration de nos résultats. Nous renvoyons le lecteur intéressé à celle-ci pour une compréhension plus approfondie de sa signification.

parallèle est large (ligne bleue). L’existence de ce dernier, ne bénéficiant pas d’un accès aux opérations de politiques monétaires, a pour conséquence de créer une limite sur l’effet des politiques de liquidités.

Il est dès lors possible qu’il soit nécessaire pour la banque centrale d’étendre les facilités de prêt aux institutions bancaires parallèles ainsi que d’appliquer une politique d’achat d’actifs si celle-ci souhaite aller plus loin dans sa politique de stabilisation financière. La partie inférieure du graphique 2 illustre ce phénomène en représentant l’évolution du prix des actifs en fonction de la quantité des actifs détenus par la banque centrale (en proportion de leur stock total) sous deux scénarios différents.

Dans un premier cas nous supposons que l’effet de la politique n’a pas d’effet de distorsion sur l’économie. Selon ce scénario, il est possible pour la banque centrale de ramener le prix des actifs à leur niveau fondamental, même lorsque le secteur bancaire parallèle est large (ligne rouge). Dans un deuxième temps, nous étudions le cas où un programme d’achat d’actif de la banque centrale est coûteux pour l’économie.¹

1. Ce coût peut être justifié de façons diverses. Il est généralement admis par la littérature qu’il existe un bénéfice à une détention privée des actifs risqués afin d’obtenir une meilleure allocation du capital

Dans ce cas, si ces coûts sont croissants, il est possible que l'effet négatif sur la production finisse par dominer l'effet de réduction du risque de liquidités (ligne bleue). L'utilisation de cet outil de politique monétaire non conventionnelle doit donc se juger à l'aune de l'équilibre entre ces deux effets et sera d'autant plus efficace et nécessaire que le secteur bancaire parallèle est large et les coûts macroéconomiques engendrés sont faibles.

Ce chapitre fournit donc une formalisation de l'argument selon lequel la crise a poussé les banques centrales à prendre leurs responsabilités en tant que pourvoyeur de liquidités en dernier ressort pour le secteur bancaire parallèle qui s'est développé dans les années précédant la crise hors de leur portée, avec des avantages potentiels pour la stabilité financière (Mehrling, 2010).

Chapitre 2 Notre deuxième chapitre intitulé "stagnation de prime risque élevée" pose la question suivante. Comment la croissance de la productivité interagit-elle avec les cycles financiers ? Dans les années qui ont suivi la grande récession, la diffusion d'innovations nouvelles a été particulièrement faible, ce qui a entraîné une faible croissance de la productivité. Ce chapitre montre que ce schéma peut être rationalisé dans un modèle macrofinancier standard lorsque nous supposons que les agents sont hétérogènes quant à leur aversion pour le risque et que la mise en pratique de ces innovations est une entreprise risquée.

Dans le modèle, les agents détiennent un portefeuille diversifié d'actions d'entreprises privées et ont la possibilité d'activer de nouvelles technologies en réaffectant une partie du capital physique à de nouveaux projets. En particulier, nous faisons l'hypothèse de l'existence d'un continuum de projets présentant différents profils de risques. Les agents évaluent en continu le potentiel de nouveaux projets et choisissent un portefeuille de nouvelles entreprises afin d'obtenir un profil de rendement-risque efficace qui correspond à leur préférences.

En temps normal, les intermédiaires financiers profitent d'un arbitrage entre une variance moyenne plus élevée dans la technologie de diffusion de l'innovation et une

aversion au risque plus faible. En raison de l'incomplétude du marché, ils le font en s'appuyant sur l'effet de levier, ce qui fragilise le système économique. Après une série de chocs négatifs, le système entre dans un régime de crise dans lequel les intermédiaires vendent leurs actifs aux ménages au prix du marché. Une spirale de liquidité est déclenchée lorsqu'une augmentation du risque endogène se répercute sur la valeur nette des intermédiaires par une baisse des prix des actifs.

Le modèle met en évidence trois canaux reliant les crises financières au ralentissement de la croissance de la productivité. Premièrement, pendant le régime de crise, l'augmentation du risque endogène crée un effet d'éviction qui réduit l'investissement dans la diffusion de l'innovation. Parce que les agents se soucient uniquement de leur exposition globale au risque, lorsque le risque endogène augmente, ils réagissent en réduisant leur exposition au risque d'innovation. En outre, les crises financières redistribuent la richesse des intermédiaires financiers, aux ménages peu enclins à prendre des risques et ayant une expertise financière moindre. Par conséquent, l'aversion au risque de l'investisseur moyen dans l'économie augmente, ce qui diminue le risque global pris dans la diffusion des technologies.

Enfin, parce que les ménages font face à un arbitrage défavorable entre le risque et le rendement lorsqu'ils investissent dans de nouveaux projets, la croissance de la productivité est plus faible que lorsque cet investissement se fait par l'intermédiaire du secteur financier.

La figure 2 nous permet d'illustrer ces résultats. Celle-ci représente différentes variables d'intérêts à l'équilibre dans le modèle en fonction de la capitalisation du secteur la moins averse au risque. Le panneau supérieur droit nous montre comment l'aversion au risque agrégé, c'est-à-dire qui est en vigueur sur les marchés financiers, est décroissante dans la richesse relative du secteur avec une aversion au risque faible. Les trois panneaux inférieurs nous montrent comment cette capitalisation relative du secteur financier influence le choix de risque de technologie, le taux de diffusion des technologies et le ratio de Sharpe, représentant le coût du risque sur les marchés financiers, pour des distances à la frontière technologique différentes. Nous pouvons y

voir que le risque optimal dans l'économie ainsi que la diffusion de la technologie est croissante dans la capitalisation des intermédiaires alors que le coût du risque est lui décroissant. Par conséquent, lorsque le secteur bancaire est sous capitalisé à la suite d'une crise financière, la croissance de la productivité se fait à un rythme plus lent.

Ce phénomène est renforcé lorsque l'économie opère à un niveau rapproché de la frontière technologique (ligne bleue par comparaison à la ligne rouge). Lorsqu'une série de chocs négatifs frappe, l'économie se déplace dans des états du monde dans lesquels les primes de risque sont élevées, où la prise de risque est faible et où l'innovation est lente. L'économie s'éloigne donc de la frontière technologique. Deux régularités intéressantes se dégagent du modèle. Tout d'abord, en période d'essor technologique, la prise de risque est plus élevée et, par conséquent, la probabilité d'une crise est plus élevée si celui-ci est suivi d'une série de chocs négatifs. En ce sens, le modèle correspond au modèle de la bulle technologique sans avoir à s'écarter de la valeur fondamentale ni de la rationalité. Deuxièmement, la dynamique de la croissance de la productivité s'inverse à long terme. À mesure que l'économie subit une série de chocs négatifs qui nuisent à la capacité agrégée à la prise de risque, elle s'éloigne de la frontière technologique et augmente la capacité d'innovation à implémenter dans le futur. Par conséquent, une fois que l'économie aura atteint la tolérance au risque d'avant la crise, elle croîtra à un rythme plus rapide. Il est intéressant de noter que cette tendance est conforme aux travaux de Field (2004), selon lesquels la majeure partie de l'augmentation de la productivité dans la période d'après-guerre a été rendue possible par les découvertes à la frontière technologique qui ont eu lieu dans les années 1930.

Chapitre 3 Dans le troisième chapitre "une méthode de résolution pour modèles de temps continu", nous proposons un algorithme capable de résoudre une classe générale de modèles de manière rapide et standardisée. La plupart des modèles de valorisation des actifs avec des agents hétérogènes possèdent une structure mathématique très similaire. En particulier, ces derniers sont composés d'une équation Hamilton-Jacobi-Bellman (HJB) pour chaque agent, couplée à un système d'équations algébriques

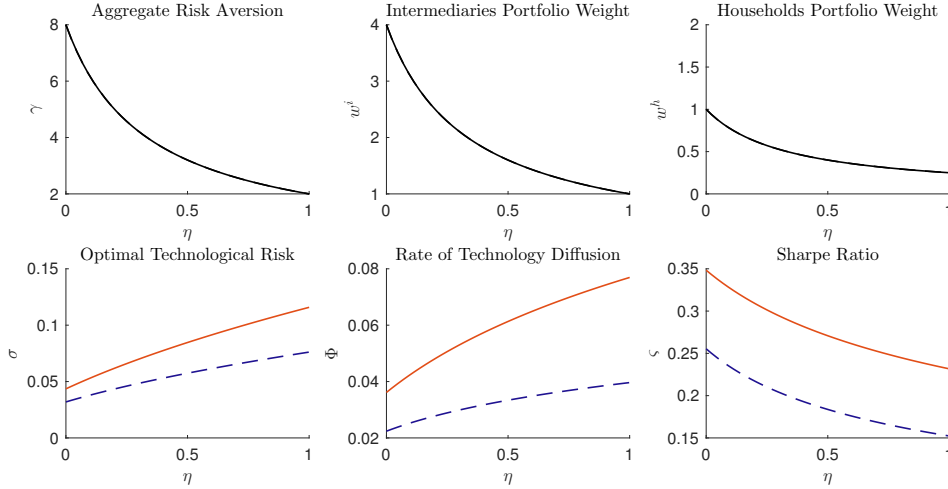


FIGURE 2 – La figure est également reproduite à partir du texte principal de la thèse à titre d’illustration de nos résultats. Nous renvoyons le lecteur intéressé à celle-ci pour une compréhension plus approfondie de sa signification.

issues des conditions d’équilibre du marché et de contraintes financières limitant les distributions de risques possibles.

Du fait que les modèles de macro-finance sont exprimés sous formes récursives, les variables formant cet équilibre dépendent exclusivement de la position d’un nombre restreint de variables dites d’états et les équations HJB sont donc indépendantes de la variable temps. Mathématiquement, ceci se traduit par le fait que celles-ci sont des équations aux dérivées partielles (EDP) elliptiques, dégénérées et non linéaires. La résolution d’un tel système d’EDP est, a priori, un problème fastidieux car ces équations sont connues pour être très instables : les erreurs d’approximation ont tendance à s’amplifier pour créer une dynamique explosive. L’algorithme proposé parvient à surmonter ces problèmes en combinant des éléments de différentes parties de la littérature sur les résolutions numériques.

Tout d’abord, comme il est d’usage dans la littérature physique, nous ajoutons une dimension temporelle fictive avec pour objet de résoudre le système dans le temps jusqu’à la convergence vers l’équilibre. Plus précisément, nous suivons Brunnermeier et Sannikov (2016) et résolvons la partie algébrique du système de façon statique tout en résolvant les fonctions de valeur des différents agents en remontant le temps.

La partie statique du système d'équations est résolue entre chaque itération en utilisant une méthode simple de Newton-Raphson, en utilisant la solution sans contrainte comme estimation initiale. La résolution de la fonction de valeur exige une attention particulière, car l'équation HJB hérite de l'instabilité inhérente à l'équation d'advection dont elle est composée. Exprimé de façon informelle, il est nécessaire d'effectuer l'approximation des dérivées partielles dans la direction spatiale qui préserve la monotonie de l'opérateur elliptique. Avec une variable d'état (ou plusieurs variables d'état avec des lois de mouvement non corrélées), nous pouvons cependant appliquer le traditionnel « en amont », en prenant l'approximation des différences finies selon le signe de la dérivée de la loi de mouvement de la variable d'état correspondant.

Le concept clé à cet égard est que l'approximation préserve la *monotonie* de la solution à chaque itération temporelle, c'est-à-dire qu'elle n'ajoute pas un nouveau maximum local. Nous utilisons un schéma de différence finie en amont de type *implicite*. Une méthode implicite, bien que plus complexe à programmer et nécessitant plus d'effort de calcul à chaque étape de la solution, est plus stable et permet des pas de temps importants. Les méthodes explicites calculent l'état d'un système à un moment ultérieur à partir de l'état du système au moment actuel, tandis que les méthodes implicites trouvent une solution en résolvant une équation impliquant à la fois l'état actuel du système et le dernier.

Lorsque nous avons au moins deux variables d'état corrélées, le problème est plus complexe car la direction préservant la monotonie peut se trouver à l'intérieur de l'espace d'état mais pas nécessairement sur un point de la grille discrète. Dans ce cas, nous utilisons la méthode développée par Bonnans, Ottenwaelter, et Zidani (2004) consistant à utiliser un degré de liberté disponible dans le problème d'interpolation afin de créer une rotation dans l'espace d'état dans un temps de calcul réduit. Cette méthode de décomposition du stencil permet de préserver la stabilité numérique de notre algorithme tout en nous permettant d'utiliser des étapes d'itération de tailles larges afin de réduire le temps de convergence.

La décomposition du stencil proposée est cependant problématique pour les

points de la grille qui sont trop proches de la limite. Si la taille du stencil décomposé est de taille $P > 1$, il se peut que la décomposition du stencil nécessite d'utiliser un point qui est potentiellement au-delà de la grille. Par conséquent, pour les points trop proches de la frontière de sorte que, pour l'une quelconque des dimensions, on impose que la dérivée du second ordre soit égale à celle du point le plus proche de la grille qui ne soit pas trop proche de la frontière.

Enfin, nous devons également traiter les non-linéarités découlant de la partie régulée de l'équation HJB. Nous suivons ici la suggestion de Candler (1999) de traiter le problème comme s'il était linéaire et de relâcher la partie non linéaire à chaque itération. Nous résolvons ensuite le système dans la dimension temporelle à l'aide d'un algorithme d'Euler entièrement implicite jusqu'à la convergence. Ceci nous permet de résoudre une gamme large de modèles dans un temps record et avec un taux d'échec très faible ; en claire amélioration par rapport aux techniques communément utilisées.