



Document de travail

Séries de pièces et billets socialement optimales : les coûts de production comptent-ils réellement ?

Yassine BOUHDAOUI

Les opinions exprimées dans ce Document de Travail sont celles des auteurs et ne reflètent pas nécessairement la position de Bank Al-Maghrib. Afin de garantir une meilleure qualité et rigueur scientifique, les documents de travail publiés sont évalués par des arbitres externes, universitaires et chercheurs de banques centrales modernes.

Aucune reproduction ou traduction de la présente publication ne peut être faite sans l'autorisation des auteurs.

L'objet de la publication du présent Document de Travail est de susciter les débats et d'appeler commentaires et critiques.

Si vous avez des commentaires sur ce Document de Travail, veuillez les faire parvenir par e-mail : dr@bkam.ma

Ou par courrier à l'adresse suivante :

Bank Al-Maghrib, Département de la Recherche
277, Avenue Mohammed V - B.P 445 Rabat

Ce document peut être téléchargé sans frais par voie électronique sur : www.bkam.ma

ISSN (en ligne) : 2509-0658

Séries de pièces et billets socialement optimales: les coûts de production comptent-ils réellement ?¹

BOUHDAOUI Yassine¹

RESUME

En ajoutant de nouvelles valeurs faciales aux séries de pièces et billets en circulation, les banques centrales visent à faciliter les paiements en espèces. Cependant, en pratique ces dernières optent pour des séries de dénominations monétaires de taille limitée. La littérature explique ce choix par les coûts de production supplémentaires associés à l'introduction de nouvelles valeurs faciales. Bouhdaoui et Van Hove (2017) teste cette hypothèse en introduisant un coût fixe par valeur faciale dans le modèle d'appariement de Lee et al. (2005) sur la base des données de production des dénominations monétaires aux Etats-Unis. Le présent papier réalise cette analyse dans le contexte marocain. Les résultats obtenus confirment que la taille des séries de dénominations peut être augmentée au-delà des seuils observés en pratique sans que les économies, résultant des gains obtenus en efficacité des paiements, ne soient dominées par les coûts de production additionnels. Ainsi, l'étude suggère que la limitation du nombre de dénominations proviendrait de coûts calculatoires et d'habituation supportés par le public lors de l'utilisation de nouvelles valeurs faciales. Ces coûts sont anticipés par les banques centrales mais ne sont pas pris en compte dans les modèles existants.

Mots-clés : Divisions monétaires, Billets de banque, Banque centrale, Coût social, modèle de search.
Classification JEL: E40, E42, E47.

ABSTRACT

By adding denominations to their coin and banknote series central banks can increase the efficiency of cash payments. In practice, however, they opt for a denominational structure with a relatively low density. The literature holds that this is because of the production costs involved. Bouhdaoui and Van Hove (2017) test this proposition by introducing a per-denomination fixed cost into the matching model of Lee et al. (2005) and parameterize the model with data on the production of US dollar banknotes. This paper aims to test this assumption using production figures of Moroccan banknotes. The results confirm that central banks could increase the density of their currency systems beyond the observed level without the efficiency gains for transactors being dwarfed by the additional production costs for the central bank itself. Our finding shows that the explanation for the low density rather lies with computational and habituation costs incurred by consumers and merchants - and anticipated by central banks - that are not yet in any of the extant models.

Keywords: Denominational structure, Banknotes, Central bank, Social cost, Search model.
JEL-classification : E40, E42, E47.

¹ Economiste chercheur à Bank Al-Maghrib : y.bouhdaoui@bkam.ma.

1. Introduction

Assurer un système de paiement en espèces robuste et efficace est une des missions principales des banques centrales. Cette mission est d'autant plus importante que plusieurs études font état du coût social² important des systèmes de paiement de détail, dont la monnaie fiduciaire fait partie (Garcia-Swartz et al, 2006; Schmiedel et al., 2012). Au Maroc, la majorité des paiements de détail sont effectués en espèces³. Les politiques permettant de réduire ce coût social des paiements pour le cas du cash sont au centre d'intérêt des banques centrales.

Parmi les actions envisageables, les banques centrales peuvent agir sur leurs coûts privés⁴, dans la mesure où ces actions n'entraînent pas d'externalités indues telles que l'augmentation des coûts supportés par les autres intervenants du circuit du cash. Une première option consiste ainsi à améliorer le *rapport durabilité-coût* des billets de banque. Cela passe par le renforcement du substrat des billets via des procédés tels que le vernissage ou encore l'introduction de substrats en plastique lorsque ces derniers sont adaptés aux usages du public et à l'environnement d'utilisation (Menzies, 2004; Bouhdaoui et al., 2013; Van Hove, 2015)⁵. Une autre option portant sur les petites coupures consiste à remplacer le billet de plus faible valeur faciale par une pièce de monnaie dont le coût de maintenance est plus faible (Lambert et al., 2013). La banque centrale peut également optimiser la gestion du stock, la taille et l'ordonnancement des commandes de billets à la production (Massoud, 2005).

Le présent article se propose d'examiner une autre option considérée par les banques centrales qui consiste en la modification du système de divisions monétaires (SDM), une expression faisant référence aux valeurs faciales des pièces et billets en circulation, avec pour objectif d'améliorer l'*efficacité* des paiements en espèces. Cette option est particulièrement intéressante pour les banques centrales dont le SDM est de faible « densité », c'est-à-dire comportant un faible nombre de divisions⁶. Ces dernières dans cette situation peuvent envisager l'ajout d'une ou plusieurs valeurs faciales à la série en circulation et faire ainsi en sorte de rendre les transactions en espèces plus aisées.

² Le coût social d'un moyen de paiement fait référence à l'ensemble des frais engagés pour sa mise à disposition auprès du grand public et des coûts résultants de son utilisation pour effectuer des transactions. Par exemple, pour le cas du cash, le coût social se compose des coûts de production et de distribution supportés par la banque centrale, des coûts de gestion des banques commerciales relatifs aux dépôts et aux retraits (GAB, agence, etc.), des coûts de gestion de caisse des commerçants prenant en compte les risques de vol et de contrefaçon, et enfin des coûts divers des consommateurs liés à la facilités d'utilisation du cash comparativement aux autres moyens de paiements.

³ Alors que le taux de bancarisation atteignait 69% en 2016, le cash reste prédominant pour les transactions de détail. En 2016, environ 85% des opérations réalisées par carte concernent les retraits dans les GAB.

⁴ Ces derniers comprennent les coûts de conception des divisions monétaires, les coûts de production et de distribution des coupures.

⁵ L'introduction de billets en polymère est aussi motivée par la lutte anti-contrefaçon (Kim et Turton, 2014).

⁶ La densité d'un SDM se mesure par le nombre de divisions par décade ; i.e. dans chaque intervalle $[10^n, 10^{n+1}[$.

En effet, disposer de plus de divisions permet de réaliser des paiements en mobilisant, en moyenne, un nombre plus faible de billets et pièces en circulation.

L'impact sur le coût social d'une telle politique fait l'objet d'un débat dans la littérature académique. En effet, à première vue, cette dernière pourrait satisfaire les intérêts de toutes les parties prenantes impliquées. Cela fût le cas, jusqu'à Van Hove (2001) où le paradigme dominant dans la littérature académique suggérait qu'un SDM permettant aux agents d'économiser sur le nombre de coupures monétaires échangées dans les transactions réduirait le volume de pièces et billets en circulation, et permettrait ainsi d'abaisser les coûts de maintenance et de production encourus par les banques centrales.

De récentes études ont cependant contesté cette affirmation. Bouhdaoui et al. (2011) en particulier ont démontré que, sous certaines conditions, les économies de coûts de production résultant d'un SDM plus efficace sont annulées par les *coûts fixes* engendrés par l'introduction de nouvelles valeurs faciales. Ils montrent que les banques centrales font face à un compromis entre efficacité des paiements et coût.

Cependant, Bouhdaoui et al. (2011) ne modélisent pas explicitement l'utilité et les coûts supportés par les intervenants du circuit du cash en particulier le coût de transport des coupures dans le portefeuille du public. Bouhdaoui et Van Hove (2017) lèvent cette limitation en introduisant un modèle général qui adopte le *bien-être social* comme critère d'évaluation. Ce modèle permet de comparer des SDM en tenant compte des interactions entre les valeurs faciales et la distribution de transactions en espèces observée dans l'économie.

Plus spécifiquement, Bouhdaoui et Van Hove (2017) se basent sur le modèle de Lee et al. (2005), qui endogénéise à la fois la distribution des transactions et les portefeuilles détenus par les agents économiques. Ils étendent ce dernier en introduisant les coûts de production du cash, en particulier un coût fixe par division qui est au cœur du modèle de Bouhdaoui et al. (2011) – ainsi que deux autres composantes de coût de production – associant ainsi deux modèles, l'un se focalisant sur le volet transactionnel et l'autre sur la fonction de coût de la banque centrale.

Cette étude estime les gains espérés de l'introduction de nouvelles valeurs faciales sur la base d'une calibration du modèle à partir de statistiques de l'économie des États-Unis et de données de production des billets de banque américains. Le présent article réalise cette estimation pour le cas du

Maroc. Il confirme les résultats obtenus par Bouhdaoui et Van Hove (2017) dans le cas américain, à savoir, qu'à mesure que la densité du SDM augmente, les gains en efficacité des paiements pour les agents économiques restent supérieurs aux coûts additionnels de production supportés par les banques centrales, jusqu'à quatre valeurs faciales par décade.

Ce résultat semble en contradiction avec la limitation du nombre de valeurs faciales par les banques centrales dans la vie réelle, la densité du SDM plafonnant à trois ou, dans certains cas, même à deux. L'étude suggère l'existence d'autres coûts supportés par le public, consommateurs et commerçants, non pris en compte dans les modèles existants et qui affectent l'adoption des nouvelles valeurs faciales mis à disposition par la banque centrale. Autrement dit, les coûts de production de la banque centrale ne constituent pas le facteur déterminant du nombre optimal de valeurs faciales d'un SDM, mais ce dernier se trouverait plutôt du côté du public.

La suite de l'article est structurée comme suit. Dans la section suivante, nous revenons sur les principales évolutions du SDM marocain sur les deux dernières décennies et abordons plus en détail, et au moyen d'exemples de vie réelle, pourquoi une augmentation de la densité d'un SDM réduit le volume des coupures en circulation et peut donc réduire les coûts de production de la banque centrale. Dans la partie 3, nous décrivons l'état stationnaire du modèle, les différentes composantes de coût de production des espèces considérées, ainsi que la fonction de bien-être social. Dans la partie 4, nous paramétrons notre modèle et présentons les résultats de nos simulations. Dans la partie 5, nous évaluons les implications de nos résultats. Enfin, la partie 6 conclut l'article.

2. Contexte de l'étude

2.1. SDM marocain :

Le SDM marocain actuel est composé de 12 divisions et suit une structure de type 1-2-5 qualifiée de binaire-décimale. Il couvre 5 décades auxquelles il manque les divisions 1 centime, 2 centimes et 500 DH, ce qui correspond à une moyenne de 2,4 divisions par décade, contre 2,2 aux US et 3 dans la zone euro.

Cette structure a connu quelques évolutions durant les deux dernières décennies. En 1995, elle a connu le déplacement de la frontière pièces-billets avec l'introduction de la pièce de 10DH en remplacement du billet de même valeur de la série 1987. Cette transformation a été accompagnée la

même année par l'introduction d'une nouvelle division de valeur intermédiaire, 20DH, qui n'existait pas jusqu'alors, les séries précédentes se contentant des divisions 10DH et 50DH. Par la suite, une nouvelle division 2 DH en format pièce a été introduite en 2002. Contrairement à la division 20 DH, cette pièce n'a pas connu de demande forte de la part du public depuis son introduction.⁷ Enfin, ces exemples illustrent l'évolution continue des séries de valeurs faciales qui concernent la majorité des pays. L'introduction des nouvelles divisions vise théoriquement à réduire la demande adressée aux divisions voisines et à améliorer l'efficacité des paiements en espèces du public. Cependant, dans certains cas les nouvelles divisions ne rencontrent pas un succès de la part du public et ce en dépit de l'avantage qu'elles présentent. Ce constat concerne les formats pièce et billet de banque comme nous l'illustrons par quelques exemples plus loin dans cette section.

2.2. Comparaison de SDM :

La Figure 1 présente des données sur le nombre de pièces et billets en circulation, par coupure dans quatre pays. Les chiffres sont présentés décade par décade, le but du diagramme étant de comparer les positions relatives des coupures adjacentes. Afin de pouvoir comparer les résultats sur plusieurs décades et ensuite par pays, nous ne présentons pas de données brutes mais plutôt recalculées. En particulier, nous présentons pour chaque division le ratio du volume en circulation par rapport à celui de la « valeur faciale 5 » appartenant à la même décade.

Notons aussi que nous ne couvrons pas toutes les décades car nous excluons les grandes coupures servant aussi à des fins de thésaurisation, en l'occurrence pour le cas du Maroc la division 200 DH. Nous avons procédé de la sorte car la demande pour de telles coupures est différente de la demande pour des coupures dites « transactionnelles » ; plus particulièrement, la demande pour ces coupures de thésaurisation n'est pas impactée par l'efficacité des paiements⁸.

La Figure 1(a) montre que la demande relative adressée à la division 1DH dépasse largement celle des divisions 10 centimes et 10DH. En effet, on dénombre 4,2 fois plus de coupures 1DH en circulation comparé à la division 5DH, contre 1,3 fois pour la division 10 DH par rapport à 50 DH.

⁷ Dans des volumes plus modestes, Bank Al-Maghrib a émis en 2012 la division 25DH sous la forme d'une coupure commémorative de la fête de l'indépendance. Cette dernière fut tirée à 4,2 millions d'unité. A la différence des autres divisions de la série 2002 et de la série de billets 2012 imprimés sur un substrat papier, le substrat de la division 25 DH est constitué d'une couche de polymère, insérée en sandwich entre deux couches de papier.

⁸ Pour cette raison, les résultats de nos simulations dans la partie 4 ne doivent pas être utilisés pour déduire des implications de politique publique concernant les coupures de haute valeur. Une hypothèse implicite dans notre modèle est que toutes les coupures sont utilisées à des fins transactionnelles et jamais uniquement comme réserve de valeur.

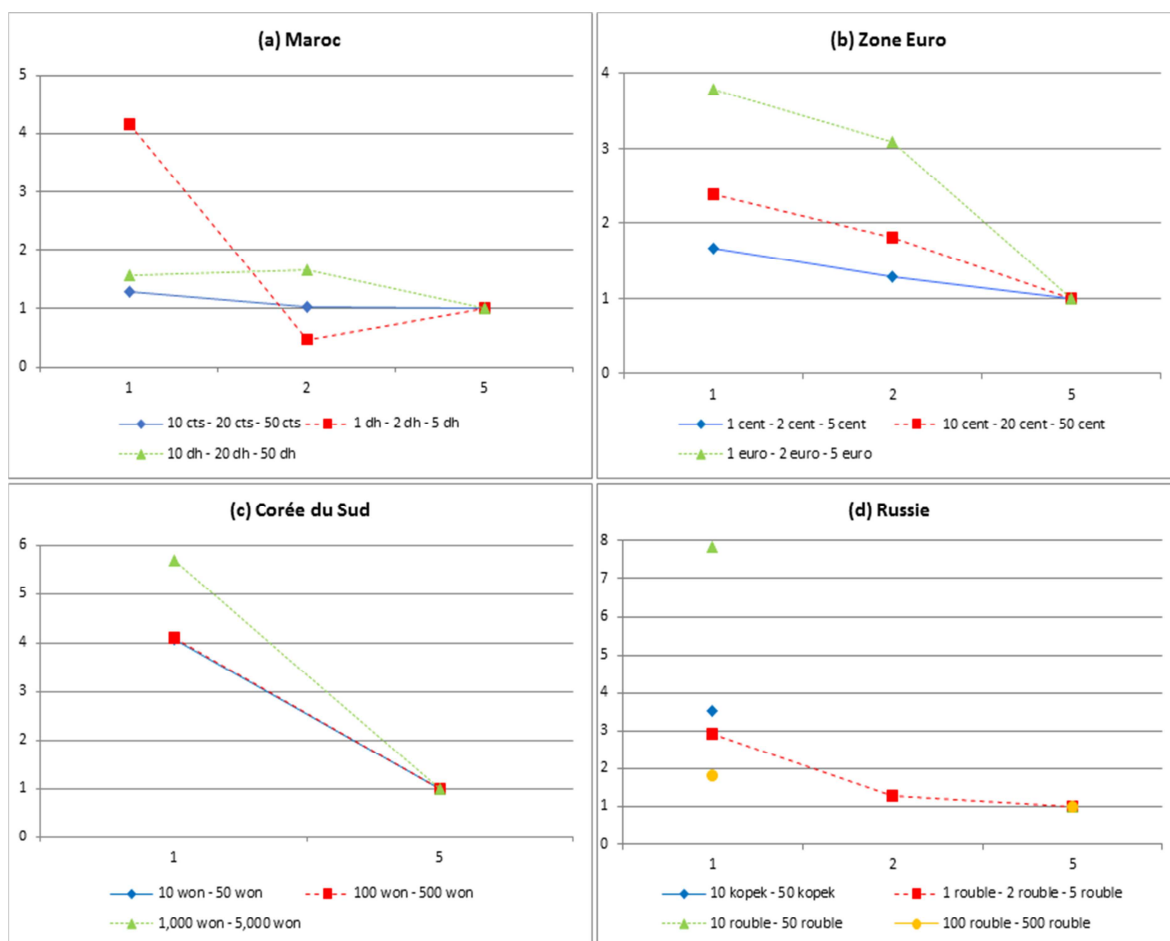


Figure 1 : Circulation de pièces et billets, par coupure et décade, fin-2015

Ces écarts de circulation s'observent aussi dans d'autres pays. Un exemple de référence de banques centrales adoptant aussi une structure binaire décimale est la Banque centrale européenne (BCE) dont le SDM contient 15 valeurs faciales, allant de la pièce de 1 centime jusqu'au billet de 500 EUR. Cependant d'autres banques centrales ont opté pour des structures plus réduites de type 1 – 5 telles que la Corée du Sud.⁹

Il est intéressant de noter que la différence de densité entre les séries – ex. 3 divisions par décade contre 2 - a un impact notable sur la composition de la circulation monétaire. Dans le cas de la BCE et de la Banque de Corée une comparaison des Figures 1(b) et 1(c) montre que la circulation des

⁹ Afin d'identifier de telles banques et compiler les données nécessaires, nous nous en sommes remis au rapport « Red Book » publié par la Banque des règlements internationaux (BRI, 2016). Dans ce rapport, le tableau 4 procure des données par pays sur la composition des monnaies en circulation. Sur tous les pays couverts par le tableau, seule la Corée du Sud possède une structure complète de type 1 – 5.

pièces et billets en euro, est mieux équilibrée que leur contrepartie en won. En Corée du Sud, le rapport de circulation entre les valeurs faciales « 1 » et « 5 » se situe dans la tranche 4,1-5,7 par rapport à 1,7-3,8 pour la zone euro. Ce résultat n'est pas surprenant. La littérature sur l'efficacité des paiements en espèces montre - tout comme notre simulation dans la partie 4.2 - que lorsque la valeur faciale « 2 » manque, cela se traduit, principalement mais pas exclusivement, par une hausse de la circulation de la coupure voisine « 1 ».¹⁰

En inversant la logique, il apparaît que la Banque de Corée (BoK) pourrait, *ceteris paribus*, améliorer l'efficacité des paiements en espèces en introduisant, par exemple un billet de 2.000 KRW, et pourrait par la même occasion réduire ses propres coûts privés. C'est le cas de certains pays comme le Japon et la Russie qui possèdent une structure 1 – 5 mais mettent en circulation, respectivement, les divisions 2000 JPY et 2 RUB. L'impact de la division 2 RUB sur les ratios des volumes en circulation est représenté dans la Figure 1(d). On observe que la valeur faciale intermédiaire fait reculer la circulation de la pièce de 1 RUB : dans la tranche 1 rouble - 2 roubles - 5 roubles, le rapport entre la circulation des pièces de 1 RUB et de 5 RUB est seulement de 2,9, contre 3,5 pour la tranche 10 kopek - 50 kopek et jusqu'à 7,83 pour la tranche 10 rouble - 50 rouble¹¹. Seul le rapport de 1,81 pour la tranche 100 roubles - 500 roubles est plus bas¹². Il est intéressant de noter que la Banque de Russie (BOR) a annoncé en Avril 2016 qu'elle mettrait en circulation deux nouvelles coupures de valeur de 200 RUB et de 2.000 RUB fin 2017.

Les exemples ci-dessus aident à comprendre pourquoi un SDM comportant davantage de divisions réduit non seulement les coûts privés des consommateurs et des commerçants - en raison de la facilité de paiement accrue et du nombre plus faible de coupures dans le portefeuille des consommateurs - mais aussi les coûts privés supportés par la banque centrale.

Cependant, il existe toujours un écart entre, d'une part, les prédictions de la littérature théorique qui considère que le public bénéficie, au même titre que les banques centrales à travers des économies de coûts de production, de disposer d'un choix plus étendu de valeurs faciales, et d'autre

¹⁰ Supposant un comportement de paiement efficace, Cramer (1983, p. 301) montre qu' « en général, la distance relative à la division supérieure limite le nombre de coupures échangés dans les paiements ». L'impact précis d'une division de 2 manquante dépend de la répartition des transactions, mais si l'on considère une distribution uniforme de transactions sur l'intervalle [1, 100] et que l'on supprime les coupures de 2 de la structure 1-2-5, alors, la fréquence moyenne d'utilisation des 1 et 10 augmente, respectivement de 210% et 243% contre 18% et 13% pour les 5 et les 50. Le rapport utilisé dans le texte principal est donc une sous-estimation du déséquilibre de circulation, tout comme le manque du 2 pousse également la circulation du 5 (la référence) à un niveau (un peu) plus élevé.

¹¹ De manière similaire aux calculs effectués pour la zone euro et la Corée du Sud, dans le graphique 1c, nous ne prenons pas en considération la plus haute tranche. De plus, nous ne montrons pas aussi la tranche la plus faible car les pièces de 1 et 5 kopek sont rarement utilisées de nos jours, en raison de leur faible valeur. Nous omettons également la pièce commémorative olympique de 25 roubles.

¹² Cela est probablement dû à l'usage des billets de 500 roubles dans les terminaux GAB.

part, l'observation que ces banques émettent dans la réalité un nombre limité de valeurs faciales¹³. A cela s'ajoute également le fait que certaines divisions sont rejetées par le public en dépit des efforts de communication des banques faisant en sorte que les gains d'efficacité des paiements escomptés ne sont pas pleinement réalisés.

Comme l'illustre la Figure 2, c'est le cas par exemple de la division de 2 DH dont le volume en circulation reste limité. Dans ce cas de figure, la demande s'adresse principalement à la division 1 DH dont le ratio du volume en circulation correspond à 4,2 fois celui de la division 5 DH. Ce même constat s'observe pour le billet de 2 USD pour lequel on observe des ratios quasi-identiques. Enfin, le billet de JPY 2.000 au Japon circule également très peu contrairement à la pièce de 2 euros.

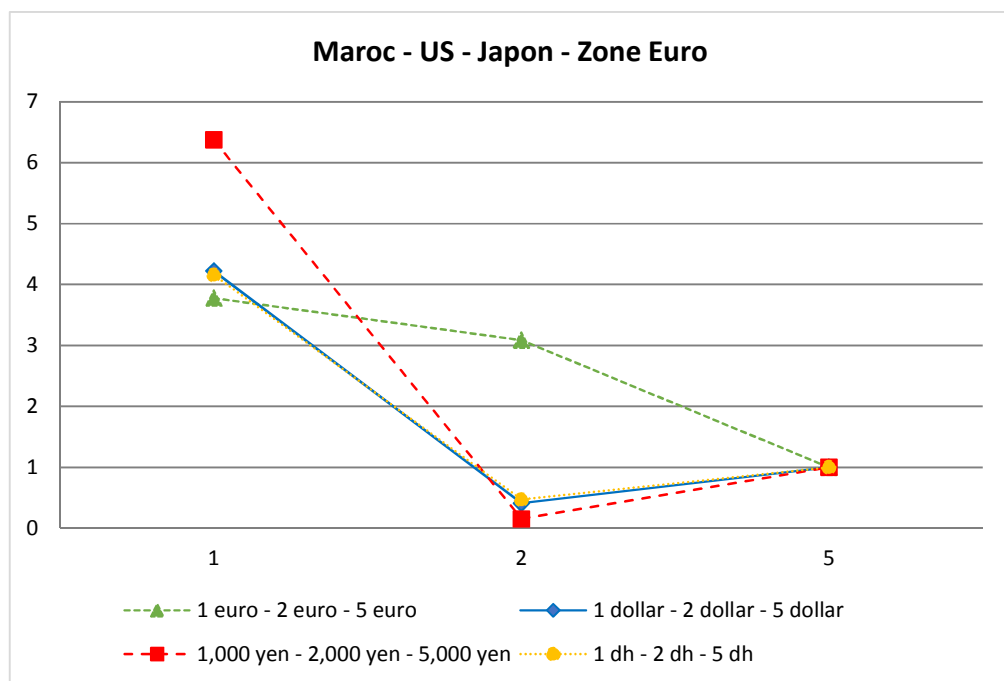


Figure 2 : Comparaison de la circulation de pièces et billets, par coupure sur une décade, fin-2015

La suite de l'article vise, entre autres, à comprendre ce paradoxe dans l'objectif de réconcilier les modèles existants et les critères d'adoption de divisions monétaires par le public qui dépend de plusieurs facteurs. Il s'agit ici, dans la suite de Bouhdaoui et Van Hove (2017), d'évaluer pour le cas du Maroc la densité socialement optimale du SDM sur la base de ce modèle qui incorpore plusieurs éléments : les coûts privés de la banque centrale, le gain d'utilité des consommateurs généré par une

¹³ La densité de SDM maximale observée est 3 valeurs faciales par décade.

plus grande efficacité des paiements en cash¹⁴ et le coût supporté par ces derniers pour le transport du cash qui dépend du *nombre* de coupures monétaires dans le portefeuille.

3. Le modèle

Le modèle Bouhdaoui et Van Hove (2017) est une extension du modèle d'appariement Lee et al. (2005) qui endogénéise le SDM et par là la composition des portefeuilles des agents économiques. Cette caractéristique est importante dans la mesure où des études mettent en évidence des interactions entre la distribution des transactions dans l'économie et la structure du SDM. Par exemple, Knotek (2008) montre que les prix ronds, c'est-à-dire ceux correspondant à des valeurs faciales, ont une rigidité nominale supérieure à la moyenne¹⁵.

D'autre part, le modèle adopté tient compte de la contrainte de disponibilité des coupures monétaires lors de la transaction en rendant endogène la composition et la taille du portefeuille des agents. Ces derniers encourent alors un coût de transport en fonction du nombre de coupures transportées dans le portefeuille, et la réalisation d'une transaction tient compte de la disponibilité des coupures dans le portefeuille de l'acheteur et celui du vendeur.

Enfin, le modèle intègre les coûts privés de la banque centrale à travers trois composantes de coût de production de cash, à savoir un coût fixe par division monétaire, le coût de production du stock initial de coupures en circulation, et les frais de maintenance (voir section 3.3)¹⁶. Par frais de maintenance nous entendons les coûts relatifs au remplacement des billets usés et dont la qualité est inférieure aux standards requis.

A partir de données sur la production de billets de banque marocains nous examinerons comment le gain en utilité du public évolue lorsqu'on augmente la densité du SDM. La question principale est précisément de voir à quel moment la banque centrale devrait cesser d'ajouter de nouvelles coupures parce que les coûts fixes de production l'emportent sur les avantages pour les agents économiques, comme le montrent nos exemples issus de situations réelles. De manière plus générale, le but de l'exercice est de voir si ce compromis explique pourquoi les séries de pièces et billets sont généralement plafonnées à trois divisions par décade.

¹⁴ Un système de paiement de détail efficient profite à l'activité économique. Dans ce cadre, un SDM optimal rend plus efficient les paiements en cash en fluidifiant les transactions.

¹⁵ De plus, Bouhdaoui et al. (2014) montrent que des prix ronds sont plus souvent payés en espèces. Karroubi et Chenavaz (2015) montrent que les vendeurs optent pour des prix qui peuvent facilement être payés en espèces et que cette stratégie pousse les acheteurs à payer effectivement par cash.

¹⁶ Ces deux derniers éléments de coût ont été initialement introduits dans Lee et Wallace (2006) pour des SDM à une seule dénomination.

3.1. Cadre général

Le temps est une variable discrète, il existe N types de biens divisibles et périssables, et pour chaque type il existe un continuum $[0,1]$ d'agents vivant indéfiniment. Un agent de type $n \in \{1,2, \dots, N\}$ produit des biens de type n et consomme uniquement des biens de type $n + 1$ (modulo N), avec $N \geq 3$. Les agents maximisent leur utilité actualisée avec un facteur d'actualisation $\beta \in]0,1[$.

Le SDM est indiquée par $s = (s(1), s(2), \dots, s(J))$, où $s(k)$ fait référence à la valeur nominale de la $k^{\text{ième}}$ division. Nous supposons que $s(k + 1) > s(k)$, que les divisions de plus grande valeur sont des multiples entiers de la division la plus petite, et que cette dernière n'est pas inférieure à l'unité (c.-à-d. $\frac{s(k)}{s(1)}$ et $\frac{1}{s(1)}$ sont des nombres entiers positifs). Nous représentons les portefeuilles détenus par les agents économiques par l'expression $y = (y(1), y(2), \dots, y(J))$, où $y(k) \geq 0$ représente le nombre d'unités monétaires de la $k^{\text{ième}}$ division. La détention moyenne de monnaie par type d'agent est désigné par \bar{m} et un seuil maximal B est appliqué sur les liquidités détenues individuellement. L'ensemble des montants de liquidités détenus s'exprime alors par $\mathbf{B}_s = \{s(1), 2s(1), \dots, B\}$. De la même façon, nous exprimons par $\mathbf{Y}_s = \{y \in \mathbb{Z}_+^J : y \cdot s \leq B\}$ l'ensemble des portefeuilles individuels possibles.

La séquence d'action pour une période donnée est la suivante. Au début de la période, les agents choisissent un portefeuille en accord avec leur contrainte de richesse. Ensuite, les agents sont associés par paires au hasard. Puisqu'il existe $N \geq 3$ types d'agents, les rencontres pertinentes sont celles dites de '*coïncidence simple*' où un agent de type n (le consommateur) rencontre un agent de type $n + 1$ (le producteur) et lui soumet une offre de consommation « à prendre ou à laisser ». L'offre soumise satisfait la contrainte de transfert de richesse entre le portefeuille du consommateur et celui du producteur. Nous considérons que l'utilité réalisée par ces rencontres uniques pour un agent de type n détenant un portefeuille y est $u(q_{n+1}) - q_n - y \cdot c(y)$, où q_n et q_{n+1} sont respectivement, les biens produits de type n et les biens consommés de type $n + 1$, γ représente le coût de transport d'une unité monétaire, et $c(y)$ représente le nombre d'unité monétaire transportées dans un portefeuille ($c(y) = \sum_k y(k)$). La fonction d'utilité $u : \mathbb{R}^+ \rightarrow \mathbb{R}$ est considérée comme étant strictement croissante et concave, dérivable en continue, et satisfaisant les conditions suivantes : $u(0) = 0, u'(\infty)$ et $u'(0)$ suffisamment large.

Finalement, du fait de contraintes computationnelles qui rendent nécessaire de limiter la taille des liquidités détenues (cf. supra), nous adoptons la fonction d'expansion monétaire de Lee (2010). Après les rencontres en paires, chaque agent reçoit une unité monétaire avec la probabilité μ et chaque unité d'argent est détruite avec la probabilité η . Il est à préciser que cette option « d'expansion monétaire » n'augmente pas le stock de monnaie mais reproduit les effets de l'inflation sur l'utilité anticipée¹⁷ des agents. Cette option permet cependant de fournir des transactions moyennes plus larges, permettant ainsi de mieux différencier des structures contenant de multiples coupures. Cette dernière a l'avantage de mettre en évidence de manière plus claire les différences de performance des différents SDM. L'intuition est simple. Dans le cas limite où les transactions réalisées ne dépassent pas l'unité alors tous les SDM performeront de manière équivalente.

3.2. Définition de l'état stationnaire

Nous dénotons la distribution des richesses et l'utilité actualisée espérée avant la sélection du portefeuille respectivement par π et ω . Nous représentons la distribution de portefeuilles par θ après le choix de portefeuille mais avant l'association des agents par paires. Finalement, nous utilisons les symboles $\tilde{\pi}$ et $\tilde{\omega}$ pour respectivement représenter la répartition des richesses et l'utilité actualisée espérée après l'association par paires et avant l'expansion monétaire.

Dans ce qui suit, nous nous concentrons sur une version 'sans-loterie' du modèle afin de renforcer l'arbitrage entre l'utilité espérée des associations et le coût de transport des unités monétaires. Etant donné un portefeuille de consommateur y_c et un portefeuille de producteur y_p , l'ensemble des transferts de richesse possibles de y_c vers y_p s'exprime par :¹⁸

$$T(y_c, y_p) = \left\{ \left((v_c - v_p) \cdot s \in \{0, s(1), \dots, \min\{y_c \cdot s, B - y_p \cdot s\}\} : \right. \right. \quad (1)$$

$$\left. \left. 0 \leq v_i \leq y_i, i \in \{c, p\} \right. \right\}$$

Le problème du consommateur est formulé comme suit :

$$\max_{z \in T(y_c, y_p)} \{u(q) + \tilde{\omega}(y_c \cdot s - z)\} \quad (2)$$

¹⁷ Voir Lee (2010, p. 327-328) pour des explications supplémentaires.

¹⁸ De la même manière, nous définissons l'ensemble des transferts d'unités monétaires comme suit : $T_v(y_c, y_p) = \{(v_c - v_p) : 0 \leq v_i \leq y_i, i \in \{c, p\} \text{ et } (v_c - v_p) \cdot s \in T(y_c, y_p)\}$

sous la contrainte

$$\tilde{\omega}(y_p \cdot s) \leq \tilde{\omega}(y_p \cdot s + z) - q$$

La contrainte est satisfaite à l'égalité puisque c'est le consommateur qui formule une offre « à prendre ou à laisser » et que la fonction u est strictement croissante. Le problème du consommateur peut donc s'écrire comme suit :

$$\max_{z \in T(y_c, y_p)} \left\{ u \left(\tilde{\omega}(y_p \cdot s + z) - \tilde{\omega}(y_p \cdot s) \right) + \tilde{\omega}(y_c \cdot s - z) \right\} \quad (3)$$

La valeur optimale de (3) est exprimée par $f(y_c, y_p, \tilde{\omega})$ et l'ensemble des transferts de richesse optimaux par $optT(y_c, y_p, \tilde{\omega})$. Si (3) donne lieu à des solutions multiples, nous permettons toutes les randomisations possibles sur celles-ci. Nous exprimons par Λ l'ensemble de mesures sur \mathbf{B}_S qui satisfait :

$$\Lambda(y_c, y_p, \tilde{\omega}) = \{ \lambda : \lambda(z, y_c, y_p, \tilde{\omega}) = 0 \text{ if } z \notin optT(y_c, y_p, \tilde{\omega}) \}$$

Le terme $\lambda(z, y_c, y_p, \tilde{\omega})$ peut être interprété comme la probabilité qu'un consommateur détenant un portefeuille y_c effectue un transfert de richesse z après avoir rencontré un producteur détenant un portefeuille y_p . Le transfert de richesse s'effectue au travers l'échange d'unités monétaires entre les deux agents économiques. Lorsque le paiement peut s'effectuer de diverses façons, - c.-à-d., à l'aide de diverses combinaisons d'unités monétaires, une combinaison aléatoire est sélectionnée.¹⁹ L'ensemble des transferts d'unités monétaires optimaux lors une rencontre est exprimé comme suit :

$$optT_v(y_c, y_p, \tilde{\omega}) = \{ v \in T_v(y_c, y_p) : v \cdot s \in optT(y_c, y_p, \tilde{\omega}) \}$$

de même l'ensemble des randomisations Λ_V peut être formalisé comme suit :

$$\Lambda_V(y_c, y_p, \tilde{\omega}) = \{ \lambda_V : \lambda_V(v, y_c, y_p, \tilde{\omega}) = 0 \text{ if } v \notin optT_v(y_c, y_p, \tilde{\omega}) \}$$

¹⁹ Le choix des coupures monétaires transférées du consommateur au producteur (et le retour correspondant du changement, le cas échéant) ne porte pas atteinte au bien-être social des agents à l'état stationnaire. Cependant, il a un impact sur les coûts d'entretien de circulation monétaire, qui sont financés par impôts forfaitaires (voir section 3.3).

Après le choix du portefeuille, l'utilité actualisée espérée associée à la possession d'un portefeuille y est :

$$h(y, \tilde{w}) = -\gamma \cdot c(y) + \frac{N-1}{N} \cdot \tilde{w}(y \cdot s) + \frac{1}{N} \cdot \sum_{y_p \in Y_s} \theta(y_p) \cdot f(y, y_p, \tilde{w}). \quad (6)$$

Au début de la période, les agents économiques disposant d'une richesse initiale z peuvent choisir un portefeuille d'une valeur allant jusqu'à z de sorte à maximiser la fonction suivante :

$$w(z) = \max_{y: y \cdot s \leq z} \{h(y, \tilde{w})\}. \quad (7)$$

Nous dénotons par $optT(z, \tilde{w})$ l'ensemble des choix de maximisations des portefeuilles pour un niveau de richesse z et considérons toutes les randomisations comme appartenant à Φ , un ensemble de mesures sur \mathbf{Y}_s qui satisfait :

$$\Phi(z, \tilde{w}) = \{\phi_z : \phi_z(y, \tilde{w}) = 0 \text{ if } y \notin optY(z, \tilde{w})\}. \quad (8)$$

Le terme $\phi_z(y, \tilde{w})$ fait référence à la probabilité qu'un agent détenant une richesse z choisisse un portefeuille y au début de la période, avant les associations par paires.

Par conséquent, l'ensemble des distributions de portefeuille précédant le couplage par paires s'exprime comme suit :

$$\Theta(\pi, \tilde{w}) = \left\{ \theta : \theta(y) = \sum_{z \in B_s} \pi(z) \cdot \phi_z(y, \tilde{w}) \text{ for } \phi_z \in \Phi(z, \tilde{w}) \right\} \quad (9)$$

De la même manière, l'ensemble des distributions suite au couplage par paires mais avant la création monétaire peut être exprimé comme suit :

$$\tilde{\Pi}(\theta, \tilde{w}) = \left\{ \begin{aligned} &\tilde{\pi} : \tilde{\pi}(z) = \frac{N-2}{N} \cdot \sum_{y: y \cdot s = z} \theta(y) + \frac{1}{N} \cdot \sum_{(y_c, y_p) \in Y_s^2} \theta(y_c) \cdot \theta(y_p) \cdot \\ &[\lambda(y_c \cdot s - z, y_c, y_p, \tilde{w}) + \lambda(z - y_p \cdot s, y_c, y_p, \tilde{w})] \text{ for } \lambda \in \Lambda^{(10)}(y_p, \tilde{w}) \end{aligned} \right\}$$

Le premier terme du côté droit de l'équation correspond à la fraction des agents qui n'étaient pas associés à une rencontre de type coïncidence simple. La valeur de leur portefeuille reste inchangée. Les deuxième et troisième termes se réfèrent respectivement aux fractions des consommateurs et

producteurs qui se retrouvent avec un portefeuille de richesse z suite à une rencontre de type coïncidence simple.

Plus précisément, le terme $\lambda(y_c \cdot s - z, y_c, y_p, \tilde{\omega})$ représente la probabilité qu'un consommateur détenant un portefeuille y_c effectue un transfert $y_c \cdot s - z$, pour se retrouver avec un niveau de richesse z . Le même raisonnement est valable pour le terme $\lambda(z - y_p \cdot s, y_c, y_p, \tilde{\omega})$ dans le cas du producteur. La somme de ces trois termes est la fraction des agents qui détiennent un niveau de richesse z après le couplage par paires.

En dernier lieu, nous présentons la formule de l'expansion monétaire. Nous utilisons μ pour représenter la probabilité qu'un agent reçoive une unité d'argent suite à la rencontre par pair. La matrice de transition de richesse est exprimée par A , lorsque $A_{z,z'}$ fait référence à la probabilité qu'une richesse z devienne z' suite à la création d'argent :

$$A = \begin{cases} a_{i,i} = 1 - \mu \text{ et } a_{i,i+1} = \mu ; i < B \\ a_{i,i} = 1 ; i = B \\ a_{i,j} = 0 ; \text{sinon} \end{cases}$$

Lorsque $0 \leq \mu \leq 1$.

La moyenne de liquidités détenues suite à la création de monnaie est exprimée comme suit :

$$\bar{m}_a = \sum_z \sum_{z'} \tilde{\pi}(z) A_{z,z'} z'$$

En réorganisant les termes, elle peut s'exprimer comme :

$$\bar{m}_a = \bar{m} + \mu \sum_{z=0}^{B-1} \tilde{\pi}(z) \quad (11)$$

Afin de maintenir une moyenne constante des liquidités détenues \bar{m} , nous introduisons une fonction de désintégration proportionnelle de monnaie : chaque unité est désintégrée avec la probabilité η . Nous exprimons par D la matrice de désintégration de monnaie :

$$D_{z,z'} = \begin{cases} \binom{z}{z'} \eta^{z-z'} (1-\eta)^{z'} & \text{if } z \geq z' \\ 0 & \text{sinon,} \end{cases}$$

lorsque $0 \leq \eta \leq 1$.

Le paramètre de la loi binomiale η doit s'assurer d'une détention de monnaie stable en moyenne, \bar{m} , après les phases de création et de désintégration. De la sorte, η remplit la condition :

$$\sum_z \sum_{z'} \sum_{z''} \tilde{\pi}(z) A_{z,z'} D_{z',z''} z'' = \bar{m}. \quad (12)$$

La fonction de valeur suite aux associations par paires et avant l'expansion monétaire, $\tilde{\omega}$, est exprimée comme une fonction de ω comme suit :

$$\tilde{\omega}(z) = \beta \sum_{z'} \sum_{z''} A_{z,z'} D_{z',z''} \omega(z''). \quad (13)$$

Finalement, l'ensemble des distributions de richesse à la fin de la période est exprimé comme suit :

$$\Pi(\theta, \tilde{\omega}) = \left\{ \pi : \pi(z) = \sum_{z''} \sum_{z'} \tilde{\pi}(z'') A_{z'',z'} D_{z',z} \text{ for } \tilde{\pi} \in \tilde{\Pi}((14)) \right\}.$$

Pour un SDM s donné et un paramètre μ d'expansion monétaire, un état stationnaire correspond à un triplet de fonctions (ω, π, θ) vérifiant les équations (3) à (14).

3.3. Le coût de production de l'argent

Dans cette partie, nous introduisons trois composantes de coût de production de monnaie, en particulier un coût de production du stock initial, des coûts de maintenance, et un coût fixe par division.

Nous supposons que ces coûts sont financés par un montant forfaitaire de taxes auquel sont assujettis l'ensemble des agents économiques, puisque ces coûts sont le plus souvent pris en charge

par une autorité publique - la banque centrale. Dans ce qui suit, nous formalisons le coût de production supporté par chaque agent économique.

Avant d'introduire les composantes une par une, nous introduisons d'abord les coûts variables unitaires de production. Nous supposons que ce coût est composé uniquement du facteur travail.²⁰ Les agents dédiés à la production de monnaie reçoivent une compensation pour chaque période de travail. Cette compensation est équivalente à l'utilité moyenne générée par agent à l'état stationnaire, $\frac{\bar{q}}{N}$, où \bar{q} fait référence à la consommation moyenne lors d'une rencontre de type coïncidence simple. La logique qui sous-tend cette idée est de dédommager les agents pour les sacrifices qu'ils font lorsqu'ils renoncent aux associations par paires afin de produire de la monnaie. Le terme \bar{q} est exprimé par la formule suivante :

$$\bar{q} = \sum_{y_c, y_p} \theta(y_c) \cdot \theta(y_p) \cdot q(z, y_p, \tilde{\omega}), \quad (15)$$

avec $z \in \text{opt}T(y_c, y_p, \tilde{\omega})$ et $q(z, y_p, \tilde{\omega}) = \tilde{\omega}(y_p \cdot s + z) - \tilde{\omega}(y_p \cdot s)$ (cf. Equation (2)).

En supposant que chaque agent produisant de la monnaie produit un nombre r d'unités monétaires par période indépendamment de la valeur faciale, le coût de production variable par unité s'écrit :

$$c_u = \frac{\bar{q}}{rN}. \quad (16)$$

Dès lors, le premier composant du coût de production de monnaie - c.-à-d., le coût de production du total du stock initial de coupures monétaires - est exprimé par la formule suivante :

$$C^I = c_u \cdot \bar{n}_p, \quad (17)$$

²⁰ Cette limitation est compensée dans la section de paramétrage où l'on fournit une estimation prenant en compte les autres composants du coût de production, notamment le coût des matières premières (encre, papier-coton et/ou les substrats de polymères, etc.).

où \bar{n}_p représente le nombre moyen d'unités monétaires dans les portefeuilles. Formellement, cela donne :

$$\bar{n}_p = \sum_{y \in Y_s} \theta(y) \cdot c(y). \quad (18)$$

Nous poursuivons avec le coût de maintenance de la monnaie. Ce coût a pour origine le fait que les unités monétaires en circulation doit respecter un niveau de qualité minimal (fixé par les banques centrales). A cette fin, les coupures monétaires en circulation sont soumises à un contrôle continue et les coupures usées ou en dessous du standard de qualité requis sont remplacées. Nous supposons que les unités monétaires se dégradent à chaque fois qu'elles changent de main lors des rencontres par paires.²¹ Par δ nous désignons le nombre moyen de fois où un jeton change de main durant sa durée de vie.²² Le nombre moyen de jetons échangés lors de rencontres de type coïncidence simple, \bar{n}_v , est exprimé par la formule suivante :

$$\bar{n}_v = \sum_{k=0}^J a(k), \quad (19)$$

où $a(k)$ représente la fréquence d'usage de la $k^{\text{ième}}$ division, c.-à-d., le nombre moyen de coupures $s(k)$ échangées par transaction. Ce dernier s'exprime formellement, pour quelque $\lambda_v \in \Lambda_v$ par :

$$a(k) = \sum_{y_c, y_p} \left[\theta(y_c) \cdot \theta(y_p) \cdot \sum_v \lambda_v(v, y_c, y_p, \omega) \cdot |v(k)| \right] \quad (20)$$

Par conséquent, le nombre moyen de jetons remplacé par période et par habitant est $\frac{\bar{n}_v}{\delta N}$ et les coûts espérés de maintenance de monnaie par habitant, C^M , peuvent être formulés comme suit :

²¹ Ceci est particulièrement le cas pour les billets, et moins pour les pièces. A l'inverse, relativement plus de p(21) que billets se perdent. Par conséquent, la circulation des pièces doit également être maintenue ; simplement produire le stock initial ne suffira pas.

²² Par souci de simplification, la valeur de δ est la même pour toutes les coupures.

$$C^M = \frac{1}{1 - \beta} \cdot c_u \cdot \frac{\bar{n}_V}{\delta \cdot N}$$

Enfin, les coûts fixes par division reflètent le coût supporté par la banque centrale, pour la conception et la production d'une nouvelle coupure. Ce coût inclut plusieurs facteurs, principalement la recherche et le développement, le personnel général et administratif, la dépréciation de la prépresse et du matériel de gravure, ainsi que d'autres coûts fixes de production. Nous utilisons q_F pour faire référence au coût fixe par division par période et par habitant exprimé en termes de consommation. Le coût fixe actualisé par habitant pour une série composée d'un nombre J de divisions est exprimé comme suit :

$$C^F = \frac{1}{1 - \beta} \cdot J \cdot q_F. \quad (22)$$

En rassemblant les trois composantes de coût de la monnaie et le bien-être social espéré de l'état stationnaire, nous obtenons la fonction de bien-être social *ex ante* pour un SDM s :

$$W = \pi \cdot \omega - C \quad (23)$$

avec $C = C^I + C^M + C^F$.

Puisque le problème n'admet pas de solution analytique de l'état stationnaire, nous nous appuyons sur des simulations numériques pour comparer différents SDM.²³ Une description des étapes de calcul de la simulation est fournie en annexe.

4. Simulation

Nous entendons par SDM « régulier » ceux qui sont caractérisés par une séquence se répétant à chaque décennie. Si nous exprimons la densité d'une série par m et le nombre de décades couvertes par n_{sup} , les structures régulières ont la forme générale suivante $s = \bigcup_{n=0}^{n_{sup}} \{s(1) \cdot 10^n, \dots, s(m) \cdot 10^n\}$, avec $s(i) < 10$. Les SDM réels, qui ne sont pas toujours réguliers, couvrent typiquement entre 4 et 6 décades et, comme nous l'avons exposé précédemment, se caractérisent par une densité entre 2 et 3.

²³ Pour la preuve d'existence de l'état stationnaire dans la version aléatoire du modèle, voir Lee et al. (2005). Cette preuve est également valable pour la version sans loterie avec expansion monétaire.

Dans l'exercice de simulation, nous calculons les coûts et bénéfices engendrés par une augmentation de la densité du SDM. Nous procédons de la sorte sur la base d'une calibration avec les coûts de production observés au Maroc que nous convertissons à l'échelle temporelle et en termes monétaires du modèle d'appariement.

4.1. Paramétrage

En ce qui concerne le modèle d'appariement, nous conservons le paramétrage adopté par Lee (2010) pour les paramètres de base du modèle qui est résumé dans le tableau 1. En ce qui concerne le nombre de périodes du modèle par année, la calibration $K=4$ cadre bien avec la vitesse de circulation de la monnaie au Maroc en 2016 calculée sur une base trimestrielle des données du PIB nominal et de M3²⁴. Ensuite, en accord avec le théorème de Lee et al. (2005), nous fixons une valeur suffisamment faible pour le coût de transport de monnaie afin d'assurer la convergence vers un état stationnaire non dégénéré²⁵. Nous établissons comme valeur de référence $\gamma = 0,05\%$ et effectuons des tests de robustesse à l'aide d'un paramètre additionnel, $\gamma = 0,01\%$ ²⁶. Le tableau 1 fournit un résumé des paramètres du modèle.

	Paramètre principal	Scénarios alternatifs
$u(q)$	$2 \cdot \log(1 + q)$	-
N	3	-
\bar{m}	10	-
B	30	-
K	4	-
μ	0,2	-
β	0,99	0,97
γ	0,05%	0,01%

Tableau 1 : Résumé du paramétrage du modèle d'appariement

La calibration des paramètres de coût de production r , δ et q_F est effectuée à partir des chiffres fournis par Dar As-Sikkah (DAS), l'institut d'émission du Maroc²⁷.

²⁴ Lee (2010) trouve une calibration similaire dans le cas des Etats-Unis.

Source agrégat M3 : série monétaire Bank Al Maghrib, source PIB nominal trimestriel : Haut-Commissariat au Plan.

²⁵ Un état stationnaire non dégénéré est caractérisé par un support continu de la distribution de richesse π ainsi qu'une utilité actualisée ω strictement croissante et de forme concave.

²⁶ Les tests effectués avec une valeur plus importante, $\gamma = 0.1\%$, n'ont pas convergé vers un état stationnaire pour toutes les structures de coupures dans notre ensemble de données.

²⁷ Dar As-Sikkah conçoit et imprime les billets de banque au Maroc. Elle réalise également la frappe des pièces de monnaie.

Pour le premier paramètre, r , soit le nombre d'unités monétaires produites par période, on calcule la productivité annuelle moyenne des employés de DAS en charge de la production de billets de banque, que nous divisons par K pour convertir en périodes compatibles à notre modèle. À l'aide de l'équation (16), le paramètre r permet de calculer le coût unitaire variable de production, c_u , pour chaque SDM.

Comme nous l'avons mentionné en partie 3.3, ce coût unitaire variable de production reflète uniquement les coûts directs de la main d'œuvre. Afin de rendre notre calibration plus réaliste, nous prenons également en compte le coût des matières premières. Pour cela nous adoptons la même approximation que Bouhdaoui et Van Hove (2017) en supposant le coût unitaire variable de production comme étant six fois supérieur au coût direct de la main d'œuvre²⁸. Nous exprimons cette estimation par c_u^e et l'utilisons dans nos calculs.

Nous calibrons ensuite le paramètre δ , qui fait référence au nombre moyen de fois qu'un billet de banque change de main pendant sa durée de vie. Pour cela, nous utilisons le nombre de cycles de tri d'un billet sur sa durée de vie et faisons une estimation raisonnée du nombre de changements de main au cours d'un seul cycle de tri, n_e . Le cycle de tri comprend toutes les étapes, de la distribution des billets aux consommateurs et commerçants, par le biais de leur circulation réelle, le retour des billets aux banques commerciales - généralement lorsque les commerçants déposent leurs surplus de détentions d'espèces - pour en fin de compte, être transférés à la banque centrale ou centre privés de tri agréés, où l'authenticité et la qualité des billets sont vérifiés. Les billets contrefaits ainsi que ceux qui ne correspondent pas aux normes de qualité sont détruits, tandis que les billets encore de bonne qualité sont remis en circulation.

Le nombre moyen de cycles de tri par billet de banque sur sa durée de vie peut être estimé en divisant le volume total de billets traités par an, N_p , par le volume total détruit par année, N_d (chaque cycle sur la période 2010 - 2015). Les estimations par valeur nominale sont fournies dans le tableau 2. Nous avons établi une estimation prudente, $n_e = 2$, pour le nombre de transferts par cycle de tri. Comme nous l'expliquons dans la section des résultats, cette hypothèse n'a pas d'impact sur nos

²⁸ En plus des coûts de production fiduciaire de la banque centrale, une partie significative du coût global du cash est supportée par les banques commerciales. Le modèle théorique adopté dans cette étude n'intègre pas spécifiquement ce coût. Cependant, comme cela est expliqué dans la section 4.2, la prise en compte de ces derniers ne remet pas en cause les résultats des simulations.

conclusions principales. Le paramètre δ est ensuite calculé comme une moyenne pondérée entre les divisions en fonction de leurs volumes de circulation, N_c , et on obtient $\delta = 15,3$.²⁹

Le dernier paramètre de coût restant à définir est q_F , qui fait référence au coût fixe par division monétaire encouru périodiquement, et qui est exprimé en termes de consommation.

Division	20DH	50DH	100DH	200DH	δ
Volume annuel moyen de billets traités, millions (N_p)	142	95	1074	1014	-
Volume moyen annuel détruit, millions (N_d)	112	57	153	104	-
Nombre de cycle de traitement sur la durée de vie $\frac{N_p}{N_d}$	1,3	1,7	7	9,8	-
Nombre moyen de changement de main sur la durée de vie (hyp. $n_e = 2$)	2,5	3,3	14,0	19,6	15,3
Volume moyen en circulation, millions (N_c)	93,1	55,2	486,5	584,5	-

Tableau 2 : Nombre moyen de changement de main d'un billet de banque sur sa durée de vie sur la base des données de la période 2010-2015.

Nous paramétrons d'abord le coût fixe de production par division par période et par agent, exprimé en unités monétaires du modèle, que l'on dénote par p_F . Le taux de conversion des termes de consommation en termes d'unités monétaires du modèle est donné par le rapport $\frac{\bar{q}}{N\bar{p}}$, où \bar{p} est le PIB par habitant par période du modèle :³⁰

$$q_F = \frac{\bar{q}}{N\bar{p}} \cdot p_F.$$

²⁹ Cette valeur est légèrement plus élevée que celle obtenue pour le cas des Etats-Unis qui est autour de 12.

³⁰ Le rapport \bar{q}/N est la production moyenne par période par habitant.

Afin de déterminer la valeur de p_F , nous nous basons sur le coût fixe annuel moyen par division et par habitant au Maroc, dénoté par p_F^{MA} , que nous convertissons en unités monétaires et à l'échelle temporelle du modèle.³¹ Une estimation de p_F^{MA} est calculée sur la base des coûts fixes par division fournie par Dar As-Sikkah³². Le tableau 3 présente le volume de production annuelle par habitant.

Division	Volume de production annuelle 2013-2015 (Millions)	Volume de prod. annuelle par tête (2013-2015)
20DH	77,1	2,2
50DH	50,05	1,4
100DH	136,43	3,9
200DH	112,37	3,2

Tableau 3 : Production annuelle totale et par habitant de billets de banque marocains, période 2013-2015.³³

Nous convertissons p_F^{MA} dans les termes monétaires du modèle en utilisant l'équivalence de PIB. Plus spécifiquement, nous établissons un taux de change comme rapport du PIB annuel par habitant dans le modèle, c.-à-d. $K \cdot \bar{p}$, et du PIB annuel du Maroc par habitant (PIB^{MA})³⁴. Une fois que nous avons exprimé p_F^{MA} dans les termes monétaires du modèle, nous divisons par K pour obtenir le coût fixe par habitant par période, p_F . Par conséquent, le paramètre K simplifié, nous obtenons l'expression suivante pour p_F comme fonction de p_F^{MA} :

$$p_F = \frac{\bar{p}}{PIB^{MA}} \cdot p_F^{MA} \quad (25)$$

Nous avons calibré p_F^{MA} à la valeur moyenne pondérée des coûts fixes annuels de production par habitant en tenant en compte le volume annuel de production des divisions monétaires.

4.2. Résultats

Dans ce qui suit, nous comparons quatre SDM dont la densité oscille entre 1 et 4. Nous nous limitons à des structures régulières. Nous mettons également l'accent sur les coupures dont les valeurs nominales existent dans la vie réelle ; en particulier les valeurs 1, 2, 5, etc. Dans les deux cas,

³¹ Nous exprimons ce coût par la formule p_F^{MAD} car il est exprimé en MAD et sur une base annuelle.

³² Du fait de leur nature confidentielle, nous ne présentons pas les coûts fixes par division dans le présent document.

³³ La population marocaine totalise 34,5 millions de personnes en 2016 selon le Haut-Commissariat au Plan.

³⁴ Dans nos simulations, nous établissons $PIB^{MA} = 29\,500$ MAD qui correspond à l'estimation du Haut-Commissariat au Plan en 2016.

nous procédons ainsi afin de pouvoir généraliser nos résultats à des séries de billets réelles. Nous commençons par le système $s_1 = (1,10)$, puis ajoutons progressivement les coupures. Puisque nous étudions uniquement des SDM réguliers, à chaque fois que nous ajoutons une division dans la première tranche de 10, nous ajoutons également la division équivalente dans la tranche suivante - à condition que la valeur nominale de cette dernière division ne dépasse pas le maximum imposé sur la trésorerie des agents économiques, B , qui est fixé à 30. Par exemple, lorsque l'on ajoute une division de 2, nous ajoutons également une division de 20. Toutefois, lorsque l'on ajoute une division de 5, nous ne pouvons pas ajouter celle de 50. En bref, en raison des exigences élevées des simulations en capacités computationnelles (et de la limite résultante sur la trésorerie), la série de billets de banque la plus étendue que nous pouvons utiliser dans nos simulations - $s_3 = (1, 2, 5, 10, 20)$, et $s_4 = (1, 2, 5, 7, 10, 20)$ - couvrent environ une décade et demie.³⁵

Le principal résultat de nos simulations est présenté dans la dernière ligne du tableau 4 : lorsque la densité du SDM est augmentée, l'utilité du public augmente. Ce résultat est principalement attribuable à des coûts de transport de portefeuille moins élevés pour les agents économiques car ces derniers contiennent en moyenne un nombre plus faible de coupures monétaires grâce à un SDM plus riche. Comme on peut le voir dans le tableau 5, lorsque la densité augmente, dans l'ensemble, le nombre moyen de coupures monétaires échangées dans les transactions (\bar{n}_V) ainsi que le nombre moyen de coupures par portefeuille (\bar{n}_P) baissent. Cela est dû à la disponibilité d'une nouvelle division qui crée des possibilités de transfert de richesse plus efficaces. Par exemple, le tableau 5 montre que l'introduction d'une division de 5 réduit le nombre d'unités de 1 et de 10 dans le portefeuille moyen, et ainsi réduit sa taille de 5,84 à 4,34 unités. L'ajout de la coupure 2, pour sa part, n'a guère d'impact sur l'incidence des 10 - ce qui n'est pas surprenant - mais diminue de façon spectaculaire l'importance relative du 1, et ainsi rabaisse davantage le nombre moyen global de coupures dans les portefeuilles.³⁶ Enfin, nous soulignons que l'ajout, dans s_4 , d'une valeur nominale

³⁵ Etablir la limite sur la trésorerie à $B = 100$ augmente le temps d'attente d'un facteur de 6 à 8, ce qui est peu pratique lorsque les tests doivent être répétés plusieurs fois - pour différentes séries de billets de banque et avec différents paramétrages. Les tests de robustesse sélective indiquent qu'une limite supérieure ne modifierait pas fondamentalement nos conclusions.

³⁶ Nous pouvons aussi constater que la coupure 20 n'a pratiquement pas d'impact. L'ajout d'un 20 à une série qui avait déjà un plafond supérieur de 10 permet en principe aux agents de garder une valeur de 20 dans leur portefeuille à un coût inférieur à celui des deux unités de 10. Toutefois, compte tenu de la limite supérieure sur avoirs en espèces de 30 et compte tenu du fait que la taille moyenne des transactions est d'environ 3, le 20 est trop grand pour être de quelque utilité. Cela dit, l'absence presque totale des 20 dans les portefeuilles simulés rend caduques les critiques possibles que lorsqu'on ajoute le 2 et le 20 à la série, nous augmentons non seulement la densité de la série, mais aussi le plafond supérieur - avec pour conséquence de mélanger ces deux effets.

de 7 ne diminue que marginalement la taille moyenne des portefeuilles par rapport à s_3 . La nouvelle valeur faciale 7 réduit la part de la valeur faciale 2 et plus particulièrement de la valeur faciale 5.³⁷

Structures de série	s_1	s_2	s_3	s_4
Densité	1	2	3	4
Q moyenne de production par trx \bar{q}	0,689	0,710	0,708	0,711
Taille moyenne des transactions	3,03	3,00	2,97	2,98
Coût de production unitaire d'espèces c_u^e	4.40E-07	4.54E-07	4.52E-07	4.54E-07
Coût fixe par coupure (q_F)	7.21E-06	3.65E-06	1.75E-06	1.54E-06
Bien être ex-ante (W)	10,161	10,531	10,602	10,608

Tableau 4 : Résultats généraux des simulations

Structures de série	s_1	s_2	s_3	s_4
Valeur faciale :				
1	(3,21/5,38)	(1,93/3,25)	(0,52/0,70)	(0,55/0,74)
2	–	–	(0,92/1,39)	(0,98/1,25)
5	–	(0,48/0,83)	(0,47/0,78)	(0,33/0,44)
7	–	–	–	(0,16/0,29)
10	(0,17/0,46)	(0,09/0,26)	(0,08/0,25)	(0,07/0,23)
20	–	–	(0,00/0,01)	(0,00/0,01)
Total	(3,38/5,85)	(2,49/4,34)	(1,99/3,12)	(2,09/2,97)

Tableau 5 : Nombre moyen de jetons échangés / dans les portefeuilles par division.

Les tableaux 6 et 7 montrent clairement comment tout cela impacte les coûts de production de billets de banque et, en fin de compte, le bien-être social. A commencer par les coûts, l'augmentation du total des coûts de production est le résultat de l'interaction entre C_i^I et C_i^M d'une part, et C_i^F de l'autre. Comme on peut le constater, C_i^I , soit le coût de production du stock initial de monnaie, chute de manière constante à mesure que la densité de la série est augmentée. Comme le montre clairement le Tableau 6, cela est dû au fait que le nombre moyen d'unités dans les portefeuilles chute. Les frais d'entretien, C_i^M , baissent aussi - sauf lorsque l'on effectue la transition entre s_3 et s_4 . Cela est dû, comme on peut à nouveau le constater dans le tableau 6, à ce que le nombre de jetons échangés par transaction augmente lorsqu'on ajoute une quatrième division dans la première décade.

³⁷ De manière contre intuitive, le nombre moyen de jetons échangés est plus élevé avec s_4 qu'avec s_3 . Cela est dû au fait qu'avec s_4 , les agents économiques transportent moins d'unités de 2 et par conséquent utilisent davantage d'unités de 1.

L'arbitrage de cette baisse de C_i^I et C_i^M est évident : une densité plus élevée implique un plus grand nombre de divisions. Par conséquent, les coûts fixes de production C_i^F augmentent³⁸ tout comme les coûts de production.

Structures de série	s_1	s_2	s_3	s_4
C_i	7.65E-04	1.18E-03	1.95E-03	2.35E-03
C_i^I	2.57E-06	1.97E-06	1.41E-06	1.35E-06
C_i^M	3.23E-06	2.46E-06	1.95E-06	2.06E-06
C_i^F	7.59E-04	1.17E-03	1.95E-03	2.35E-03

Tableau 6 : Composantes du coût de production de monnaie

Cependant, la première ligne du tableau 7 montre que les bénéfices en termes de bien-être social excèdent les coûts additionnels de production³⁹ : les gains résultant des coûts de transport plus faibles font plus que compenser les coûts fixes de production de divisions supplémentaires (ligne 5). Comme souligné précédemment, il y a aussi des économies dans les coûts de production initiaux (ligne 3) et, pour s_2 et s_3 , des coûts de maintenance (ligne 4), mais ceux-ci restent limités.

Pour résumer, la conclusion principale de nos simulations est que, à mesure que nous augmentons la densité de la série de billets de banque, le bien-être social augmente régulièrement - mais par portions incrémentales de plus en plus faibles - jusqu'à et y compris quatre dénominations par décade⁴⁰.

Densité (λ)	2	3	4
$(W_i - W_{i-1})$	3,70E-01	7,12E-02	5,64E-03
$(C_i - C_{i-1})$	4.14E-04	7.74E-04	3.99E-04
ΔC_i^I	-6.02E-07	-5.56E-07	-6.29E-08
ΔC_i^M	-7.75E-07	-5.00E-07	1.10E-07
ΔC_i^F	4.15E-04	7.75E-04	3.99E-04

Tableau 7 : Variation du bien-être ex-ante et des coûts de production lorsque la densité augmente.

³⁸ Les coûts fixes par coupure q_F ne sont affectés que de façon marginale.

³⁹ Le Tableau 7 montre l'existence d'un facteur multiplicateur supérieur à 10 entre le gain en bien-être social (ΔW) et les coûts additionnels de production fiduciaire (ΔC). Ainsi, la prise en compte des coûts supportés par les autres intervenants du circuit du cash non pris en compte dans le modèle théorique adopté dans l'étude, en particulier les coûts du cash supportés par les banques commerciales, ne remet pas en cause les résultats des simulations.

⁴⁰ Fait intéressant, au-delà de ce niveau, le bénéfice social commence à diminuer. L'intuition est simple. Dans le tableau 5, on peut déjà constater que l'ajout d'une quatrième dénomination diminue légèrement la taille moyenne des portefeuilles. L'ajout d'une cinquième dénomination génère des gains encore plus faibles, le cas échéant, qui sont maîtrisés par l'augmentation des coûts de production.

Contrôles de robustesse

Nous avons effectué les tests de robustesse avec deux paramètres alternatifs, à savoir un taux d'actualisation plus faible $\beta = 0,97$ (au lieu de $\beta = 0,99$), et un coût de transport de cash équivalent à $\gamma = 0,01\%$.⁴¹ Le tableau 8 montre que nos résultats ne sont pas affectés. En effet, les bénéfices en termes de bien-être social d'une densité plus élevée du SDM demeurent supérieurs aux coûts additionnels de production.

Densité (δ)	2	3	4
$\beta = 0,97$			
$(W_i - W_{i-1})$	9,46E-02	2,38E-02	4,07E-03
$(C_i - C_{i-1})$	-1,46E-04	1,72E-03	8,84E-04
$\gamma = 0,01\%$			
$(W_i - W_{i-1})$	3,05E-01	8,61E-03	1,29E-03
$(C_i - C_{i-1})$	7,04E-04	1,52E-03	1,16E-03

Tableau 8 : Variation des bénéfices sociaux ex-ante et coûts de production dans des scénarios alternatifs

5. Implications : le public détermine-t-il le nombre de valeurs faciales ?

Nos simulations dans la section précédente ont montré que lorsque la banque centrale augmente la densité du système de divisions monétaires, les gains d'utilité qui en résultent générés par des portefeuilles plus légers pour les agents économiques l'emportent sur les coûts de production plus élevés pour les banques centrales - jusqu'à et incluant quatre divisions par décade.

Bouhdaoui et Van Hove (2017) montrent que ce résultat peut être généralisé (et reste très pertinent) pour les séries réelles de pièces et de billets. Si l'on confronte cela avec l'observation que le mandat des banques centrales les oblige généralement à favoriser l'efficacité des systèmes de paiement - ce qui implique qu'elles doivent raisonner en termes de coûts sociaux supportés par l'ensemble des utilisateurs plutôt que leurs propres coûts (Van Hove, 2007) - la question clé est de savoir pourquoi les banques centrales n'optent pas pour une série avec plus de trois dénominations par décade (voir section 2). En fait, aucune banque centrale dans les pays de l'OCDE n'a mis en circulation une telle série.

⁴¹ Un facteur d'actualisation plus faible garantit que l'indivisibilité de l'argent n'affecte pas le résultat.

Les résultats obtenus pour le Maroc suggèrent comme pour le cas des Etats-Unis que les raisons sont à rechercher non pas du côté de la banque centrale mais plutôt du côté du public. En particulier, il s'agit de considérer deux types de coûts privés supportés par les consommateurs et les commerçants qui, à notre sens, doivent être pris en considération, à savoir un coût calculatoire et un coût d'habitation, engendrés par l'ajout de nouvelles valeurs faciales.

L'intuition relative au premier coût réside dans le fait que plus le nombre de valeurs faciales que transportent les agents dans leurs portefeuilles est élevé, plus il devient complexe de calculer toutes les combinaisons différentes avec lesquelles un montant donné peut être payé et plus il devient complexe de déterminer la solution la plus efficace (celle qui exige le nombre de coupures monétaires échangé le plus bas). Il y a aussi le problème de la « reconnaissabilité » : plus la variété de valeurs faciales est large, plus il devient difficile pour les agents économiques de reconnaître les différentes pièces et billets⁴² et de retrouver les valeurs faciales spécifiques nécessaires pour effectuer une transaction. Cette complexité accrue - qui n'est pas actuellement prise en compte dans le modèle - pourrait bien, à un certain point, décourager les acteurs à considérer même ces options. Autrement dit, ils pourraient bien décider d'exclure une ou plusieurs valeurs faciales du problème de calcul afin de pouvoir mieux l'appréhender.

Par conséquent, les gains d'efficacité potentiels résultant d'une variété plus large peuvent demeurer non réalisés. Or, comme l'énonce Van Hove (2001, p 1019) « si trop de valeurs faciales sont produites, certaines de ces valeurs vont naturellement devenir impopulaires et tomber hors de l'usage général. Le marché peut donc ainsi signaler les limites de l'utilité de la variété. » Dans nos simulations, les agents ne sont pas confrontés à un coût de calcul et, par conséquent, exploitent pleinement les possibilités offertes par l'utilisation de toutes les valeurs faciales de manière équilibrée (en fonction de la répartition des transactions et de la richesse). En d'autres termes, dans nos simulations, les valeurs faciales ne sont jamais « abandonnées ».

Deuxièmement, l'intuition relative au coût d'habitation réside dans le fait que l'introduction d'une nouvelle valeur faciale nécessite de la part du public de faire évoluer ses habitudes de paiement. C'est le cas par exemple du billet de 2.000 JPY, seule valeur faciale japonaise dont le premier chiffre est 2.

Ces deux coûts fournissent une explication possible de l'utilisation complète et équilibrée de certains SDM, comme c'est le cas dans la zone euro (Franses et Kippers, 2007), tout comme la faible

⁴² Le fait de vérifier si les coupures ne seraient pas contrefaits peut aussi s'avérer une lourde charge.

adoption de certaines valeurs faciales qui sont introduites à un moment où l'utilisation des SDM existants est devenue fermement ancrée, comme pour la pièce de 2DH au Maroc ou encore le billet de 2 USD aux États Unis (voir section 2)⁴³.

6 Conclusion

Ce document calibre, avec des données de production marocaines de billets de banque, le modèle d'évaluation du coût social des SDM introduit dans l'article Bouhdaoui et Van Hove (2017). Ce modèle incorpore les coûts privés de la banque centrale, l'utilité des consommateurs générée par l'efficacité des paiements en cash et le coût supporté par ces derniers pour le transport des coupures monétaires dans le portefeuille.

Les résultats obtenus confirment, comme dans le cas américain, que les économies de coûts de transport du cash continuent de l'emporter sur les coûts de production supplémentaires de nouveaux billets encourus par la banque centrale jusqu'à quatre valeurs faciales par décade. Nous montrons que ce résultat reste valable pour des systèmes de divisions monétaires réels, ce qui pose la question de savoir pourquoi les banques centrales ont tendance à plafonner la densité de leur série à trois divisions par décade.

En particulier, nos conclusions remettent en question l'idée largement répandue selon laquelle ce sont les coûts de la banque centrale qui imposent une limite sur le nombre de valeurs faciales. Au contraire, notre constat suggère qu'il existe des coûts du côté du public qui ne sont pas encore pris en compte dans le modèle, et que, compte tenu de ces coûts, les banques centrales supposent que le public ne serait pas en mesure de traiter davantage de divisions.

Plus généralement, ces résultats questionnent la composition de ce qui constituerait une série de divisions monétaires optimale, c'est-à-dire la série qui répondrait à la demande du public tout en maximisant le bien-être social. Ils ouvrent la voie à des recherches futures à travers la prise en compte de ces nouveaux coûts dans la détermination du SDM optimal.

⁴³ En cas de forte multiplication du nombre de dénominations, les autres intervenants du circuit du cash, incluant les banques commerciales et les commerçants, pourraient réagir en conséquence pour faire baisser les coûts logistiques en minimisant dans les faits le nombre de dénominations en vue de faciliter la gestion et les coûts correspondants.

Annexe : Etapes de Calcul du modèle

Les étapes de l'algorithme utilisé pour calculer l'état stationnaire sont les suivantes :

- Étape 1 : Configurer les estimations initiales pour $s, \bar{m}, B, N, \beta, \gamma$ et μ , puis spécifier une forme fonctionnelle pour la fonction d'utilité.
- Etape 2 : Configurer une estimation initiale, ω_0 , pour la valeur fonction, ω , pour chaque valeur sur B_s et calculer $\tilde{\omega}_0$ utilisant (13).
- Etape 3 : Déterminer $f(y_c, y_p, \omega)$ et $optT(y_c, y_p, \omega)$ en résolvant (3) et sélectionner une mesure de probabilité λ sur B_s qui satisfait (5).
- Etape 4 : Etablir une estimation initiale θ_0 , pour la distribution des portefeuilles θ_t , et calculer l'utilité actualisée attendue après avoir choisi un portefeuille donné $y, h(y, \omega)$ en utilisant (6).
- Etape 5 : Mettre à jour la fonction ω_t en utilisant (7), et déterminer l'éventail de portefeuilles optimaux $optY(z, \omega)$.
- Etape 6 : Calculer la distribution de richesse avant la création de monnaie, $\tilde{\pi}_t$, et après la création monétaire, π_t , en utilisant (10) et (14) respectivement.
- Etape 8 : Sélectionner une mesure de probabilité sur Y_s qui satisfait (8). Mettre à jour la distribution des portefeuilles, θ_t en utilisant (9).
- Etape 9 : Calculer les normes $\|\omega - \omega\|, \|\pi - \pi_t\|$ et $\|\theta - \theta_t\|$. Suspendre la boucle lorsqu'ils sont tous en dessous d'un seuil $\epsilon = 1E - 4$. Autrement, initialiser ω_t, π_t et θ_t comme nouvelle estimations pour la prochaine itération, et répéter les étapes 2 et 8 jusqu'à atteindre un point de convergence.

Bibliographie

Bagnall J., Bounie D., Huynh K. P., Kosse A., Schmidt T., Schuh, S. D. & Stix H., 2016, “Consumer Cash Usage: A Cross-Country Comparison with Payment Diary Survey Data,” *International Journal of Central Banking*, 12(1): 1-61.

Bartzsch N. & Seitz F., 2016, “Cash Holdings in Germany and the Demand for ‘German’ Banknotes: What Role is There for Cashless Payments?,” dans Gorka, J. (ed.), *Transforming Payment Systems in Europe*, Palgrave Macmillan Studies in Banking and Financial Institutions, 111-148.

Bouhdaoui Y. & Bounie D., 2012, “Efficient Payments: How Much Do They Cost for the Central Bank?,” *Economic Modelling*, 29(5): 1579-1584.

Bouhdaoui Y., Bounie D., François A., 2014, “Convenient Prices, Cash Payments and Price Rigidity,” *Economic Modelling*, 41: 329-337.

Bouhdaoui Y., Bounie D., Van Hove L., 2011, “Central Banks and their Banknote Series: The Efficiency-Cost Trade-off,” *Economic Modelling*, 28(4): 1482-1488.

Bouhdaoui Y., Bounie D., Van Hove L., 2013, “When Do Plastic Bills Lower the Bill for the Central Bank? A Model and Estimates for the U.S,” *Journal of Policy Modeling*, 35: 45-60.

Bouhdaoui Y. & Van Hove L., 2017, “On the socially optimal density of coin and banknote series: Do production costs really matter?,” *Journal of Macroeconomics*, 52: 252-267.

Cramer J.S., 1983, “Currency by Denomination,” *Economics Letters*, 12: 299-303.

Franses P. H. & Kippers J., 2007, “An Empirical Analysis of Euro Cash Payments,” *European Economic Review*, 51: 1985-1997.

Garcia Swartz D., Hahn R., Layne-Farrar A., 2006, “The Move Toward a Cashless Society: A Closer Look at Payment Instrument Economics,” *Review of Network Economics*, 5: 175-198.

Karoubi B. & Chenavaz R., 2015, “Prices for Cash and Cash for Prices? Theory and Evidence on Convenient Pricing,” *Applied Economics*, 47: 4102- 4115.

Kim E. & Turton T., 2014, “The Next Generation Banknote Project,” *Bulletin, Reserve Bank of Australia*, March Quarter, 1-12.

Knotek E. S., II, 2008, "Convenient Prices, Currency, and Nominal Rigidity: Theory with Evidence from Newspaper Prices," *Journal of Monetary Economics*, 55(7), 1303-1316.

Lambert M., Ferrari S., Wajert B., 2013, "Costs and Benefits of Replacing the \$1 Federal Reserve Note with a \$1 U.S. Coin," Staff Working Paper, Federal Reserve, December.

Lee M., 2010, "Carrying Cost of Money and Real Effects of Denomination Structure," *Journal of Macroeconomics*, 32(1): 326-337.

Lee M. & Wallace N., 2006, "Optimal Divisibility of Money when Money is Costly to Produce," *Review of Economic Dynamics*, 9(3): 541-556.

Lee M., Wallace N., Zhu T., 2005, "Modeling Denomination Structures," *Econometrica*, 73(3): 949-960.

Massoud N., 2005, "How Should Central Banks Determine and Control their Bank Note Inventory?" *Journal of Banking and Finance*, 29: 3099-3119.

Menzies G., 2004, "Money to Burn, or Melt? A Cost-Benefit Analysis of Australian Polymer Banknotes," *The North American Journal of Economics and Finance*, 15(3): 355-368.

Schmiedel H., Kostova G., Ruttenberg W., 2012, "The Social and Private Costs of Retail Payment Instruments," Occasional Paper Series No. 137, European Central Bank.

van der Crujssen C., Hernandez L. and Jonker N., 2015, "In Love With the Debit Card but Still Married to Cash," Working Paper No. 461, De Nederlandsche Bank.

Van Hove L., 2001, "Optimal Denominations for Coins and Bank Notes: In Defense of the Principle of Least Effort," *Journal of Money, Credit and Banking*, 33(4): 1015-1021.

Van Hove L., 2007, "Central Banks and Payment Instruments: A Serious Case of Schizophrenia," *Communications & Strategies*, 66: 19-46.

Van Hove L., 2015, "Modelling Banknote Printing Costs: Of Cohorts, Generations, and Note-Years," *Economic Modelling*, 46: 238-249.

Van Hove L., 2016, "Could 'Nudges' Steer Us Towards a 'Less-Cash Society'?" in Gorka, J. (ed.), *Transforming Payment Systems in Europe*, Palgrave Macmillan Studies in Banking and Financial Institutions, 70-110.