Monetary denominational systems, technological changes and cost of cash

Yassine Bouhdaoui

▶ **To cite this version:**

Yassine Bouhdaoui. Monetary denominational systems, technological changes and cost of cash. Economics and Finance. Télécom ParisTech, 2011. English. NNT : 0G5EUG00168 . pastel-00686655

HAL Id: pastel-00686655

https://pastel.hal.science/pastel-00686655

Submitted on 10 Apr 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

2012-ENST-00xx

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Sciences économiques »

présentée et soutenue publiquement par

Yassine BOUHDAOUI

le 15 décembre 2011

Systèmes de divisions monétaires, changements technologiques et coût social des espèces

Directeur de thèse : **David BOUNIE**

Jury

M. Cyril MONNET, Professeur d'économie, Université de Berne	Rapporteur
M. Xavier RAGOT, Professeur associé, Ecole d'économie de Paris	Rapporteur
M. Leo VAN HOVE, Professeur d'économie, Vrije Universiteit Brussel	Examinateur
M. Laurent GILLE, Professeur d'économie, Telecom ParisTech	Examinateur
M. Patrick WAELBROECK, Professeur associé, Telecom ParisTech	Examinateur

T H È S E

TELECOM ParisTech

école de l'Institut Télécom - membre de ParisTech

TABLE DES MATIÈRES

À mes Parents,

REMERCIEMENTS

Mes remerciements sincères vont tout d'abord à mon directeur de thèse, David Bounie, qui m'a permis d'effectuer cette thèse sous sa direction. J'ai eu l'honneur de bénéficier de ses conseils et critiques avisés tout au long de mes années de thèse. Ses encouragements et sa confiance m'ont été très précieux dans la réalisation de ce travail.

Ce travail doit également beaucoup au Professeur Leo Van Hove, à ses idées et suggestions, toujours pertinentes et constructives. Notre collaboration a été riche et fructueuse. J'espère que nos échanges se poursuivront avec la même intensité par la suite.

J'adresse tous mes remerciements à Cyril Monnet, Xavier Ragot, Laurent Gille et Patrick Waelbroeck pour l'honneur qu'ils m'ont fait en acceptant de participer à ce jury de thèse et d'évaluer ce travail.

Ce manuscrit de thèse doit énormément à Geoffrey Gerdes, Nicolas Houy, David Humphrey, Kristin Langwasser et Harry Leinonen qui ont eu la gentillesse de lire différentes parties de ce manuscrit et de me faire part de leur commentaires, ainsi que toutes les personnes de la Banque Centrale de Belgique, de la Réserve Fédérale Américaine et de la Banque Centrale d'Australie qui ont eu la gentillesse de mettre à ma disposition les données qui ont permis à ce travail d'aboutir.

Je suis également reconnaissant à mon Ecole Doctorale l'EDITE pour m'avoir attribué une allocation de thèse et de monitorat. Enseigner sous la direction de Sylvie Gillot et de Jacques Féjoz et auprès de mes collègues Corentin Herbert, Selma Engin et Yuanmi Chen de l'UPMC a été un grand plaisir et une expérience enrichissante. Mes élèves m'ont également beaucoup appris.

Je tiens à remercier tous mes collègues de bureau et du département qui, même s'il n'ont pas suivi mon travail, ont contribué à rendre cette période plus qu'agréable. Sans prétendre être exhaustif, je pense notamment à Laurie Marrauld, Bora Eang, Germain Gaudin, Dominique Freard, Stéphane Couture, Alex White et tous les autres que je n'ai pas cités.

Merci à Marie Baquero et Marie-Josée Vatin pour leur aide administrative tout au long de mon doctorat.

Enfin, j'adresse un profond remerciement à mes parents et à mon frère Anis pour m'avoir écouté, encouragé et soutenu tout au long de ce travail.

CHAPTER 1

INTRODUCTION

Les pièces et billets, bien que millénaires, constituent les plus anciens supports de la monnaie aujourd'hui en circulation dans les économies modernes.[1] Mais, à travers l'histoire, les formes de la monnaie et de ses supports ont considérablement évolué en raison notamment de l'évolution de la nature des échanges. De la monnaie marchandise (coquillage, etc.) à la monnaie métallique (or, etc.) puis papier (billet, etc.), la monnaie est devenue avant tout scripturale (carte de paiement, chèque, etc.) et marginalement électronique (porte-monnaie électronique) pour servir d'intermédiaires des échanges sur les réseaux de télécommunications.[2]

Les supports de la monnaie sont appelés plus communément les instruments de paiement. Ces derniers permettent les transferts des valeurs monétaires entre les agents et recouvrent différentes formes. Les innovations technologiques liées aux instruments de paiement sont considérables depuis plusieurs décennies. Le développement des automates, du web commercial et des réseaux mobiles ont incité de nombreux acteurs bancaires et non-bancaires à proposer des instruments ou des technologies adaptées à ces nouveaux lieux de commerce à l'image du porte-monnaie électronique (Van Hove, 2004b), des systèmes de paiement électronique sur Internet (Bounie, 2002) ou encore des paiements sur téléphone mobile. Ces instruments de paiement, bien qu'innovants et originaux, restent cependant encore peu utilisés par rapport aux instruments de paiement plus traditionnels que sont les espèces et les cartes de paiement. Selon les dernières statistiques publiées par la Banque Centrale Européenne (BCE, 2010), la carte de paiement

[1] L'existence de la monnaie est à l'origine de nombreuses controverses. Depuis plusieurs décennies en effet, les économistes s'interrogent sur les motifs de détention de la monnaie. Pourquoi un agent accepterait-il de détenir un billet de banque dépourvu de toute valeur intrinsèque ? La réponse à cette question fait toujours débat même si la plupart des économistes reconnaissent que la monnaie existe avant tout pour éviter le troc qui exige la double coïncidence des volontés entre les partenaires de l'échange (Williamson et Wright, 2011).

[2] Au-delà de sa fonction d'intermédiaire des échanges, la monnaie a également d'autres fonctions comme celles d'unité de compte et de réserve de valeur.

est l'instrument de paiement scriptural le plus fréquemment utilisé même si sa part de marché en valeur est relativement faible puisque les paiements par carte représentent seulement 0,7% de la valeur totale des paiements effectués dans la zone euro, loin derrière les virements manuels et automatiques (84%). Mais si la carte est l'instrument de paiement scriptural le plus utilisé, il est important de souligner que les espèces demeurent toujours, dans la plupart des économies développées, les instruments de paiement les plus utilisés au point de vente. A l'aide d'une enquête sur les comportements de paiement d'un échantillon représentatif de la population française en 2005, Bounie et François (2006) évaluent par exemple la part de marché des espèces à 60% environ. De même, Stix (2004) estime pour le cas de l'Autriche cette même part à 70%.

Les arbitrages réalisés par les agents économiques entre les instruments de paiement font l'objet de recherches soutenues depuis ces quelques dernières années. Après les contributions pionnières de Mot et Cramer (1982), Boeschoten (1989) et Humphrey *et al.* (1996), de nombreuses études économétriques se sont attachées à mettre en évidence à tour de rôle l'influence des caractéristiques des transactions (Hayashi et Klee, 2003 ; Bounie et François, 2006 ; etc.), des caractéristiques socio-démographiques (Stavins, 2001) et des caractéristiques des instruments de paiement (Kiser *et al.*, 2008) sur les choix des instruments de paiement par les consommateurs.[3] Ces études sont particulièrement importantes pour les banques et les banques centrales dans la mesure où les instruments de paiement ne sont pas équivalents en termes de coût. En effet, si certains instruments de paiement comme les cartes génèrent des revenus provenant des diverses taxes, commissions et abonnements, d'autres, en revanche, ne donnent pas lieu à rémunération. C'est le cas par exemple des espèces qui sont la plupart du temps mis gratuitement à disposition du public et des chèques, en France, qui ne donnent pas lieu à facturation.

Dans ce contexte, des études empiriques ces dernières années ont tenté de mesurer le coût social des différents instruments de paiement utilisés par les agents économiques.[4]

[3] Pour une revue de la littérature sur les choix des instruments de paiement, le lecteur consultera Bounie et François (2008).

[4] Le coût social d'un instrument de paiement se réfère à l'ensemble des ressources que la société, *i.e.*

Quelles que soient les études réalisées aux Etats-Unis, dans les pays de la zone euro ou dans d'autres pays, un constat émerge : les instruments de paiement de type papier que sont les espèces et les chèques sont plus coûteux pour la société que les instruments de paiement électronique tels que la carte de débit et le porte-monnaie électronique (Garcia-Swartz *et al.*, 2006; De Grauwe *et al.*, 2006; Guibourg et Segendorff, 2007). Par exemple, le coût social annuel des paiements en espèces en Belgique et aux Pays-Bas est évalué en moyenne à 0,7% du produit intérieur brut (De Nederlandsche Bank, 2004; Banque Nationale de Belgique, 2006) et s'élève environ 75% du coût social total des paiements.[5] En d'autres termes, la société gagnerait à accélérer la migration des paiements papier vers les paiements électroniques. Ce constat est particulièrement vrai pour certains acteurs qui supportent la plus grande partie des coûts des paiements en espèces à l'image des marchands. Ce dernier constat est toutefois moins vrai pour les banques centrales en général qui bénéficient de revenus de seigneuriage supérieurs à leur coût de production, de traitement et de distribution.

Il résulte de ces développements que la part de marché des espèces (et des chèques dans certains pays) par rapport à celle des cartes de débit par exemple sont plus importantes à l'équilibre que celles socialement souhaitables. Cette situation traduit donc un échec du marché qui résulte en partie d'une non prise en compte par les consommateurs des coûts qu'ils font supporter à la société lorsqu'ils choisissent par exemple les espèces à la place de la carte de paiement. Pour modifier les comportements de paiement des consommateurs, des stratégies ont été mises en œuvre dans de nombreux pays.

1.1 Les stratégies de réduction des coûts des paiements

Deux types de stratégies liées à l'usage des instruments de paiement en général et à la détention et à l'usage des espèces en particulier ont été mises en place pour réduire le coût social des paiements : des stratégies tarifaires et non-tarifaires. Nous détaillons et commentons à présent ces deux types de stratégies.

---

les consommateurs, les commerçants, les banques, etc., mobilise pour utiliser un instrument de paiement.

[5]Pour donner un ordre d'idée, le coût social des paiements en espèces aux Pays-Bas en 2002 était évalué à 400 euros par an et par famille (De Nederlandsche Bank, 2004).

1.1.1 Les stratégies tarifaires

L'objectif de ces stratégies est de tarifer l'usage des instruments de paiement pour les particuliers de manière à en limiter leur usage. S'agissant des espèces, ces stratégies s'inspirent en partie du cadre théorique formulé par Baumol (1952). L'auteur décrit le comportement d'un agent qui effectue régulièrement des dépenses d'un montant connu à l'avance. Pour réaliser ses achats, le consommateur peut retirer des espèces à intervalles réguliers mais il supporte, d'une part, une taxe liée au retrait et, d'autre part, des coûts liés à la détention des espèces (perte des revenus des intérêts issus de la rémunération des dépôts). Le problème de l'agent est alors de décider d'un montant optimal d'espèces à détenir qui minimise les coûts de détention et de retrait des espèces. Baumol (1952) montre que le montant d'espèces optimal d'un agent varie proportionnellement au niveau de ses dépenses, au coût des retraits et au niveau des taux d'intérêt ; en particulier, plus le coût des retraits est élevé plus le montant d'espèces détenu est faible.[6]

Pour limiter la détention d'espèces, les banques ont alors tenté de mettre en place des stratégies tarifaires pour limiter la détention et l'usage des espèces. Par exemple, en France et aux Etats-Unis, une taxe sur les retraits aux Distributeurs Automatiques de Billets (DAB) a été instituée. Bounie et François (2008) étudient cette stratégie à l'aide de données d'un sondage réalisé auprès d'un échantillon représentatif de la population française en 2005. Ils montrent que les personnes soumises à une taxe sur les retraits effectués hors de leur réseau bancaire sont incités à retirer des montants plus élevés au DAB. Ce résultat prolonge une première étude empirique de Gowrisankaran et Krainer en 2005 sur les populations de l'Iowa et du Minnesota aux Etats-Unis. Les auteurs avaient alors établi qu'une augmentation de 10 centimes de dollar du coût d'un retrait au DAB diminuait la probabilité de retrait d'un individu de 4%.

Dans le prolongement de ces travaux, des études ont tenté également de mesurer l'effet d'une tarification à l'usage de certains instruments de paiement tels que la carte et le chèque. La mise en place d'une telle stratégie a pour objet de tarifer en partie les

[6]Pour éviter de détenir des encaisses supplémentaires, les agents peuvent décider de retirer moins fréquemment.

usages des instruments de paiement à leurs coûts pour les consommateurs (Van Hove, 2004a). Deux études sur le chèque et les cartes de débit montrent que les agents sont sensibles à une telle stratégie. La première est à l'initiative d'Humphrey, Kim et Vale (2001). Les auteurs exploitent des données agrégées sur les comportements de paiement par chèque et par carte de débit en Norvège pour étudier l'impact d'une tarification du chèque sur l'usage de la carte de débit. Les résultats des estimations confirment que la tarification à l'acte du chèque permet de réduire substantiellement son usage au profit de la carte de débit. Ces résultats ont été confirmés récemment par Kiser *et al.* (2008) qui montrent que lorsque les consommateurs américains paient une taxe supplémentaire sur les paiements par carte de débit basée sur le code PIN, ceux-ci utilisent plus fréquemment la carte de débit basée sur la signature manuscrite.

Cette problématique de la tarification des instruments de paiement est également très proche des études sur le "surcharging" dans les travaux sur les "two-sided markets" (Rochet et Tirole, 2002, 2003). L'idée de départ est que si un paiement par carte est certes moins coûteux pour la société dans son ensemble, il peut être plus coûteux pour le marchand qu'un paiement en espèces. Le "surcharging" consiste donc pour un marchand à faire payer une taxe au consommateur de manière à ce que ce dernier paie tout ou partie de la différence de coût entre les instruments de paiement. Cette pratique du "surcharging", généralement condamnée jusqu'à présent par les diverses autorités de la concurrence, a été autorisée dans plusieurs pays comme au Pays-bas et en Australie. Bolt *et al.* (2010) ont étudié précisément l'impact de cette pratique tarifaire sur le coût social des paiements aux Pays-bas. Les auteurs montrent que le "surcharging" conduit les consommateurs hollandais à préférer les paiements en espèces et qu'une suppression du surcharging permettrait des économies de coût de plus de 100 millions d'euros à terme.

1.1.2 Les stratégies non tarifaires

La mise en place d'une tarification ou d'une surcharge est cependant souvent mal perçue par les agents ou parfois non recommandée par certains acteurs des paiements.

En France par exemple, le Groupement des Cartes Bancaires n'autorise pas les commerçants qui acceptent leurs cartes à pratiquer des discriminations tarifaires en fonction de la carte utilisée par le consommateur. Une alternative consiste alors à introduire des mécanismes incitatifs non tarifaires pour encourager ou dissuader les agents à utiliser les instruments de paiement les plus couteux. Plusieurs stratégies sont également pratiquées. La première est relative à la modification des conditions de paiement et de retrait des agents économiques. Dans la mesure où les instruments de paiement sont en partie substituables, l'amélioration des conditions de paiement avec un instrument de paiement pénalise l'usage des autres instruments de paiement. C'est le cas par exemple de la carte de paiement et des espèces : en améliorant les conditions de paiement par carte on désincite les consommateurs à utiliser les espèces. Cette stratégie passe par exemple par un développement du nombre de terminaux de paiement électronique puisqu'en augmentant la taille du réseau d'acceptation de la carte on incite les consommateurs à payer par carte. Certains auteurs ont tenté explicitement de mesurer l'influence du nombre de terminaux de paiement électronique (TPE) sur l'usage de la carte. A un niveau agrégé par exemple, Humphrey *et al.* (1996) montrent que le développement des TPE induit un accroissement de l'usage des cartes de paiement. Ce résultat est confirmé dans le cadre de données individuelles par Stavins (2001) qui établit qu'un individu utilise d'autant plus un instrument de paiement que celui-ci est très utilisé par le voisinage géographique. De même, Rysman (2007) met en évidence une corrélation positive entre l'usage des cartes de paiement et leur acceptation par les marchands.

Cet effet de réseau existe également dans le cadre des retraits d'espèces dans la mesure où plus le nombre de distributeurs automatiques de billets (DAB) est élevé et plus les coûts d'accès aux espèces sont faibles (Attanasio *et al.*, 2002). Mais les études empiriques relatives à l'impact des DAB sur les retraits d'espèces ne sont pas concluantes en raison principalement de la qualité des données utilisées. De nombreux travaux font usage de données agrégées à l'échelle d'un pays. Rinaldi (2001) étudie par exemple les effets du développement du nombre de DAB sur la quantité de monnaie en circulation à partir de séries temporelles sur la période 1960-1999 en Belgique. L'auteur montre que le nombre de DAB a un impact négatif sur la monnaie en circulation mais qu'un choc à court terme du nombre de DAB peut augmenter la monnaie en circulation. Snellman *et al.* (2000) conduisent également une étude de panel sur 10 pays européens et trouvent que le nombre de DAB a un effet négatif sur la quantité de monnaie en circulation. Enfin, Drehman *et al.* (2002) analysent un panel de 16 pays de l'OCDE et ne trouvent aucun impact du nombre de DAB sur la demande de monnaie. L'effet ambigu du nombre de DAB sur la quantité de monnaie en circulation peut être du à des effets qui se compensent sur les données agrégées et plaide pour une analyse sur données individuelles. Dans cet esprit, Stix (2004) a proposé une estimation de l'impact de l'usage des fonctions retrait des cartes de débit sur la détention d'espèces détenue par les individus. L'étude confirme que l'usage de la fonction de retrait de monnaie des porteurs de carte de paiement réduit la détention de pièces et billets de 25% environ en valeur.

La deuxième stratégie est relative au changement technologique. Dans le cadre des espèces, deux solutions ont été entreprises pour réduire les coûts de production annuels des pièces et billets. Un exemple bien connu est le changement de matériau pour les billets de faible valeur. Les Etats-Unis par exemple ont introduit le Golden Dollar en 2000, une pièce d'un dollar, pour limiter l'usage du billet jugé trop couteux (Lotz et Rocheteau, 2004). De même, depuis la fin des années 80, de nombreux pays ont adopté partiellement ou complètement une technologie de billet polymère (plastique) caractérisée par une plus grande résistance à l'usage et à la contrefaçon (Coventry, 2001). Peu de recherches ont été entreprises à ce jour pour évaluer les pertes ou les bénéfices de l'adoption d'un changement technologique.

La troisième stratégie a trait à la structure du système de divisions monétaires. Dans les économies modernes, les espèces sont composées d'une série finie de valeurs faciales matérialisées par des pièces et billets. Chaque valeur faciale est appelée division et forment ensemble le système de divisions monétaires (SDM). Il existe aujourd'hui une grande diversité de SDM tant sur le plan des structures que sur le plan des valeurs des divisions. Certains pays disposent de 10 divisions (Etats-Unis) alors que d'autres en ont 15 (zone euro). De même, certains pays comme les Pays-Bas ont cessé l'émission des divisions de 1 et 2 centimes d'euros au motif que les prix en espèces sont arrondis à

5 centimes d'euros. De même, le Canada a suspendu depuis 2001 l'émission de la valeur faciale la plus élevée, le billet de 1000 dollars, à l'image des Etats-Unis, alors que dans la zone euro, la BCE produit toujours des billets de 200 et 500 euros. En France, l'initiative du billet de un euro a même été relancée en 2010 par plusieurs parlementaires français avec à leur tête le député Louis Giscard d'Estaing.[7] Ces débats sur la structure des SDM ont fait l'objet de multiples recherches. Des études théoriques ont montré notamment que la structure du SDM comme le facteur d'espacement entre les divisions affectaient l'usage des pièces et billets. Plus particulièrement, Van Hove et Heyndels (1996) et Van Hove (2001) ont montré que le nombre de pièces utilisées dans les échanges variaient selon la structure du SDM ce qui impliquait de fait une augmentation des coûts de production des billets supportés par la banque centrale en charge de leur émission.

1.2 Objectifs de la thèse et contributions à la littérature

L'objectif de cette thèse est d'analyser plus en détails deux des stratégies non tarifaires évoquées précédemment que sont les changements technologiques et les modifications des systèmes de divisions monétaires. Plus particulièrement, nous souhaitons explorer les incidences des comportements de paiement, des technologies et de la structure du SDM sur la distribution des transactions en espèces dans l'économie et leur coût social. Ces stratégies non tarifaires, souvent pratiquées par les autorités d'émission, ont soit peu fait l'objet de recherches académiques soit reposent sur des hypothèses fragiles ou des résultats contestables qui sont au cœur de nos contributions. Nous abordons tout d'abord la littérature sur les SDM et nous discutons ensuite des résultats sur les changements technologiques.

1.2.1 Les systèmes de divisions monétaires

Les travaux sur les systèmes de divisions monétaires (SDM) se sont attachés à comprendre, d'une part, l'évolution de la masse monétaire par division et, d'autre part, à

[7] Le lecteur consultera le lien suivant : http://www.touteleurope.eu/fr/actions/economie/euro/actualite/actualites-vue-detaillee/afficher/fiche/4172/t/43803/from/2277/breve/le-un-contre-un-de-leuro-contre-le-dollar-louis-giscard-destaing-court-pour.html?cHash=77d7040b45 (dernière visite 10/10/2011).

étudier l'efficacité d'un SDM. Ces travaux sont importants si l'on s'intéresse à la question des coûts des paiements en espèces car la banque centrale en charge de l'émission des pièces et billets a besoin d'anticiper et de financer les coûts d'émission de la masse de monnaie en circulation. L'ensemble des travaux académiques dédiés aux SDM reposent tous sur une hypothèse commune : les agents économiques sont censés payer selon le principe du moindre effort (PME). Ce principe énonce que les consommateurs et marchands utilisent un nombre minimum de pièces et billets pour payer l'ensemble des transactions dans l'économie.[8] Dans ce contexte, si les agents se comportent selon le PME, alors il est possible d'établir théoriquement une proposition empirique établie par Hentsch (1973) selon laquelle la masse monétaire en circulation d'une division est proportionnelle à la racine carrée de sa valeur faciale (Caianiello *et al.*, 1982).

Résultat 1 : Si les agents paient selon le PME alors il est possible d'établir un principe d'invariance qui énonce qu'à l'équilibre la valeur moyenne des pièces et billets en circulation est égale à la valeur moyenne d'une transaction en espèce dans l'économie.

Dans un premier article, Bouhdaoui, Bounie et Van Hove (2011a), nous contestons la validité du principe d'invariance établi par Caianiello *et al.* (1982). Plus précisément, nous montrons que la valeur moyenne des pièces et billets en circulation n'est pas égale à la valeur moyenne d'une transaction en espèce dans l'économie et que cette inégalité repose sur une confusion entre le stock et le flux de monnaie en circulation. En conséquence, la validité de la loi empirique proposée par Hentsch manque toujours d'une preuve formelle théorique.

Contribution n°1 : Nous invalidons l'existence d'un principe d'invariance censé justifier la loi empirique de Hentsch qui décrit la valeur de la monnaie en circulation pour chaque division d'un SDM.

Supposons toutefois, pour poursuivre, que les agents paient selon le principe du moindre effort (PME). Dès lors, il devient possible, d'une part, d'étudier comment un

[8] Ce principe a été en partie validé empiriquement à l'aide d'une étude de Franses et Kippers (2007). Les auteurs montrent à l'aide d'une enquête sur les comportements de paiements d'un échantillon de consommateurs Hollandais que 61% des paiements réalisés en espèces selon le principe du moindre effort.

changement de SDM affecte le nombre de pièces et billets échangés entre les agents économiques et, d'autre part, de choisir le SDM optimal, *i.e.* le SDM qui minimise le nombre de pièces et billets échangés entre les agents. Un SDM optimal a en effet la propriété de faciliter les échanges entre les agents. Dans le cadre de plusieurs travaux théoriques, Caianiello *et al.* (1982) et Van Hove et Heyndels (1996) ont montré que le SDM optimal est un système modulaire caractérisé par un ratio constant (le module) entre les divisions. Plus précisément, Caianiello *et al.* (1982) ont démontré qu'un système puissance de deux est optimal lorsque les agents économiques paient en donnant le montant exact du paiement (les retours de monnaie ne sont pas considérés) et Van Hove et Heyndels (1996) ont étendu les résultats en considérant explicitement les retours de monnaie[9]. Ces résultats importants reposent cependant sur une hypothèse cruciale : les montants des transactions sont supposés distribués uniformément sur un intervalle.

Pourtant, les transactions en espèces observées dans la vie réelle sont loin d'être distribuées uniformément sur un intervalle. Nous constatons en effet que les prix arrondis sont plus fréquents que les autres et que tous les prix en espèces, notamment ceux de valeur élevée, sont moins fréquents. Ces observations ont été notamment confirmées par Boeschoten and Fase (1989) qui étalissent que la distribution des transactions en espèces aux Pays-Bas est plus proche d'une distribution log-normale.

Résultat n°2 : Les travaux théoriques sur l'optimalité des SDM reposent sur une distribution uniforme des transactions en espèces. Cette distribution présuppose l'existence de tous les prix sur un intervalle et un même nombre d'occurrence de chaque prix ce qui a pour incidence de surévaluer l'utilisation des coupures de valeurs faciales élevées.

Dans le deuxième travail, Bouhdaoui et Bounie (2012), nous exploitons une distribution empirique des transactions en espèces collectées auprès d'un échantillon représentatif de la population française en 2005. Nous comparons en particulier l'efficacité du SDM européen sur cette distribution empirique et sur une distribution uniforme. Nous montrons que la distribution uniforme caractérisée par une plus grande proportion de

[9]Les résultats de Van Hove et Heyndels (1996) ont contesté un précédent résultat de Sumner (1993) qui soutenait qu'un système puissances de 3 était optimal.

montants élevés induit une sur-utilisation des grosses coupures et, également dans une certaine mesure, des plus petites divisions qui sont utilisées pour rendre la monnaie.

Contribution n°2 : L'utilisation d'une distribution uniforme est erronée et conduit à surestimer l'usage des pièces et billets dans les échanges.

Ces travaux théoriques sur l'optimalité des SDM ont été étendus notamment par Van Hove (2001). Dans cette dernière contribution, l'auteur défend l'idée qu'un système puissance de deux qui possède un facteur d'espacement plus faible entre les divisions est préférable à un système puissance de trois dans la mesure où, les agents payant selon le principe du moindre effort, un nombre plus faible de pièces et billets seront échangés entre les agents. Dit autrement, si l'on compare deux SDM sur le même intervalle des prix, plus le facteur d'espacement entre les valeurs faciales d'un SDM est faible et plus le nombre de pièces et billets échangés entre les acteurs est faible. A titre d'illustration, considérons l'intervalle des prix [1, 10]. Sur cet intervalle, supposons les deux SDM suivants: 1, 2, 4, 8 et 1, 3, 9.[10] Pour payer un achat au prix de 4, une unité de la division du premier SDM est nécessaire alors que 2 le sont avec le second SDM. Cette propriété du PME qui garantit un nombre minimum de pièces et billets échangés entre les acteurs est également importante en ce qui concerne les coûts de production supportés par les banques centrales. En effet, en diminuant le nombre de pièces et billets échangés entre les agents, le PME est censé contribuer à assurer également de faibles coûts de production des pièces et billets pour les banques centrales (Van Hove, 2001).

Résultat n°3 : Un SDM optimal minimise le nombre de pièces échangées et réduit les coûts de production pour la banque centrale. Toutes choses égales par ailleurs, un SDM optimal réduit le coût social des paiements en espèces.

Dans un troisième travail, Bouhdaoui et Bounie (2011), nous montrons que le principe du moindre effort ne participe pas nécessairement à réduire les coûts de production de la

[10]Nous adoptons l'approche de Cramer (1983) qui consiste à définir un intervalle des prix couvrant un nombre entier de cycles de toutes les divisions à l'exception de la plus grande, *i.e.* $[v(1); v(J) - v(1)]$ avec $v(1)$ et $v(J)$ les valeurs faciales de la plus petite et de la plus grande division. Dans notre cas, $J = 4$ et $J = 3$ pour le premier et second SDM, $v(1) = 1$ pour les deux SDM, $v(4) = 8$ et $v(3) = 9$ pour le premier et le second SDM.

banque centrale. Pour ce faire, nous développons un modèle qui lie les coûts des transactions aux coûts de production des valeurs faciales supportés par la banque centrale et qui permet de comparer les coûts des transactions en espèces lorsque les agents utilisent le principe du moindre effort à ceux d'un autre principe baptisé le "principe du moindre coût". Ce dernier minimise les coûts des transactions en espèces sans considérer le nombre de pièces et billets échangés dans les transactions; ce principe ne minimise donc pas le nombre de pièces et billets utilisés dans les échanges et n'est donc pas nécessairement efficace au sens du principe du moindre effort. Nous simulons des paiements en espèces selon ces deux principes à partir d'une distribution log-normale représentative des transactions en espèces aux Etats-Unis et en se basant sur les données de coût de production du SDM américain de l'année 2010. Les résultats obtenus montrent que si le principe du moindre effort réduit de 6% le nombre de pièces et billets utilisés dans les échanges, il augmente en contrepartie d'environ un quart les coûts des transactions en espèces comparativement au principe du moindre coût. En conséquence, si le principe du moindre effort facilite les échanges entre consommateurs et marchands, il est susceptible d'accroître les coûts des transactions des espèces et ainsi les coûts de production de la banque centrale. Il existe donc un arbitrage pour les banques centrales qui opposent les intérêts des agents économiques (commodité des paiements) et leur propre intérêt (coût de production des billets).

Contribution n°3 : Le principe du moindre effort ne garantit pas une réduction des coûts des transactions et des coûts de production pour la banque centrale.

Dans un quatrième travail, Bouhdaoui, Bounie et Van Hove (2011b), nous évaluons plus généralement comment cet arbitrage est sensible à la structure du SDM et aux coûts de production des pièces et billets et des valeurs faciales. En effet, un SDM dont le facteur d'espacement est faible réduit le nombre de billets échangés et de fait le nombre de billets à produire par la banque centrale ; ainsi, les coûts variables de production pour émettre ces billets sont donc également plus faibles pour la banque centrale. Cependant, dans le même temps, la banque centrale a besoin d'émettre un nombre de valeurs faciales plus important (4 contre 3 dans l'exemple ci-dessus) ce qui induit une augmentation, d'une part, des coûts fixes de production pour chaque valeur faciale supplémentaire et, d'autre part, des coûts fixes de production pour toutes les unités produites pour chaque valeur faciale (éléments de sécurité, coût de R&D, etc.). A l'aide de simulations, nous montrons qu'un SDM dont le facteur d'espacement est plus faible n'induit pas nécessairement des coûts de production plus faibles. Ce résultat confirme à nouveau qu'il existe un arbitrage fondamental pour la banque centrale entre les intérêts du public et les coûts de production qu'elle supporte pour permettre la réalisation des transactions.

Contribution n°4 : Un SDM caractérisé par un facteur d'espacement plus faible n'induit pas nécessairement une baisse des coûts de production pour la banque centrale.

Enfin, dans ces travaux, une dernière hypothèse centrale formulée est qu'il existe une indépendance entre le comportement de paiement et la structure du SDM. Or, les enquêtes menées montrent que la structure du SDM est liée non seulement à l'environnement économique et aux comportements de paiement mais également à la distribution des transactions et au bien être social (Knotek, 2008). Dans ce contexte, l'étude de Lee *et al.* (2005) constitue une contribution majeure dans la mesure où il introduit le premier modèle d'équilibre général qui endogènéise la structure du SDM à travers les coûts de transport de chaque coupure détenue dans le portefeuille des agents. Les auteurs utilisent un "modèle de search" dans lequel des consommateurs et des producteurs se rencontrent aléatoirement pour réaliser des transactions. Pour faire face à leurs achats, les agents détiennent des coupures et supportent alors des coûts de transport. Sous certaines conditions et après une série de rencontres, ce modèle converge vers un état stable dans lequel la répartition des richesses et l'utilité espérée des agents deviennent stationnaires. Ce dernier état est pris comme référence pour mesurer l'impact d'une structure de SDM sur le bien-être social. Cependant, ce cadre possède des limites qui tiennent à une définition réduite du coût social à travers le seul élément du coût de transport. Dans Lee et Wallace (2006), cette limite est levée en introduisant en particulier les coûts de maintenance du stock de monnaie en circulation et du coût de production initial. En revanche, le modèle considère un SDM avec une seule division et n'endogénéise pas les coûts de transport des coupures monétaires.

Résultat n°4 : La littérature étudie le coût social des espèces soit en prenant partiellement en compte les coûts avec un SDM à plusieurs divisions comme c'est le cas pour Lee *et al.* (2005), soit en définissant le coût social de manière plus complète mais dans le cadre d'un SDM à une seule division (Lee et Wallace (2006)). Il n'existe pas de modèle général qui associe un coût social plus complet et un SDM à plusieurs divisions.

Nous étendons le modèle de Lee *et al.* (2005) en introduisant, d'une part, les coûts de maintenance des espèces et le coût de production du stock initial pour les SDM à plusieurs divisions. D'autre part, nous intégrons un coût fixe par division qui matérialise le coût de conception d'une nouvelle division que nous supposons indépendant du volume en circulation.[11] Ces trois éléments de coût définissent le coût social d'un SDM et peuvent être comparés entre plusieurs systèmes de divisions. A ce titre, nous réalisons une simulation avec les structures puissances de deux et puissances de trois étudiées plus généralement dans la littérature sur les SDM. Nous trouvons alors que le premier SDM (puissances de deux), tout en étant plus efficace dans les paiements, peut être socialement plus coûteux lorsque le coût fixe par division est élevé.

Contribution n°5 : Nous proposons un modèle général qui élargit la définition du coût social dans le cadre d'un SDM à plusieurs divisions. A l'aide d'un jeu de simulations, on analyse les interactions entre les différentes composantes de coût et la composition des portefeuilles des agents.

1.2.2 Les changements technologiques

Dans le cadre des changements technologiques, nous avons indiqué que la Réserve Fédérale des Etats-Unis a introduit en 2000 la pièce d'un dollar en vue de réduire la production de billets de la même valeur jugée plus coûteuse. En dépit des stratégies adoptées, le public n'a jamais adopté cette pièce pour l'utiliser dans des transactions quotidiennes en raison des nombreux coûts matériels et psychologiques (Lotz et Rocheteau, 2004).

[11] L'introduction d'un coût fixe par division a également été recommandé dans Lee *et al.* (2005) p.956.

De même, l'Australie a adopté progressivement une gamme complète de "billets plastique" depuis la fin des années 1980. A ce jour, plus de 30 pays dans le monde ont adopté partiellement ou complètement cette nouvelle technologie (Canada, Inde, etc.). La question qui se pose pour la banque centrale est celle des gains potentiels de l'adoption de cette innovation technologique. Le billet plastique a en effet plusieurs avantages pour la banque centrale. Tout d'abord, sa durée de vie est plus longue (entre 2 et 5 fois) - qu'un billet papier traditionnel ce qui réduit les volumes de billets à remplacer chaque année. En outre, il rend plus complexe la contrefaçon ce qui augmente la demande pour les billets "authentiques" émis chaque année. Mais le billet plastique a également des désavantages. Son coût unitaire de production est plus élevé (entre 2 et 3 fois) ce qui augmente les coûts des billets remplacés chaque année et qui diminue les revenus du seigneuriage provenant des nouvelles émissions de billets. Au final, l'adoption du billet plastique a donc des effets à la fois positifs et négatifs pour une banque centrale et un cadre formel est donc nécessaire pour les évaluer.

Ces effets ont été partiellement pris en compte dans le cadre d'une simple analyse coût-bénéfice réalisée par Menzies (2004). Ce dernier ne propose pas un cadre général pour évaluer les effets ambivalents de l'adoption d'une nouvelle technologie de paiement. Ce cadre formel fait en outre défaut aux rares travaux dédiés à cette question. Or, ce cadre est nécessaire car si certains pays ont réussi avec succès l'introduction des billets plastique par exemple, d'autres pays tels que le Brésil sont revenus sur leur décision.

Résultat n°5 : La littérature académique ne dispose pas d'un cadre formel pour évaluer les gains et les bénéfices potentiels de l'introduction d'une nouvelle technologie.

Dans Bouhdaoui, Bounie et Van Hove (2011c), nous proposons un cadre formel pour évaluer les effets de l'adoption d'une nouvelle technologie sur les revenus de la banque centrale liés à l'activité de production des divisions monétaires. Notre modèle additionne les variations potentielles des coûts de production et des revenus de seigneuriage suite à la migration vers une nouvelle technologie de fabrication des coupures monétaires. Les paramètres observés par le modèle sont notamment les coûts unitaires de production, la durée de vie des coupures et leur résistance à la contrefaçon. Nous utilisons ce cadre formel pour évaluer les gains potentiels pour la Réserve Fédérale des États-Unis liés à l'adoption de la technologie polymère pour l'ensemble des billets de banque en circulation. A l'aide de données détaillées sur la période 2003-2008, les simulations montrent qu'une migration complète apporterait un revenu supplémentaire de 140 millions de dollars par an à la Réserve Fédérale sur l'activité de production des divisions monétaires.

Contribution n°6 : Nous proposons un modèle pour évaluer les gains et les pertes potentielles liées à l'adoption par la banque centrale d'une nouvelle technologie de paiement destinée à remplacer la production de pièces et/ou billets. Nous appliquons ce modèle au cas d'une migration vers la technologie polymère des billets de banque aux Etats-Unis.

1.3 Nos papiers de recherche

Ce manuscrit est articulé autour de six papiers de recherche. Trois de ces papiers ont été acceptés pour publication, un autre a été soumis et les deux autres ont le statut de document de travail. Les articles de recherche sont :

1. Bouhdaoui Y., Bounie D. and Van Hove L., 2011a, "The 'Principle of Invariance' in Currency Systems: a Comment on Caianiello *et al.*", forthcoming in *International Journal of General Systems*.

2. Bouhdaoui Y. et Bounie D., 2012, "Distribution des Transactions en Espèces et Efficacité des Paiements en Euros", Forthcoming in *Revue Economique*.

3. Bouhdaoui Y., Bounie D. and Van Hove L., 2011b, "Central Banks and their Banknote Series: the Efficiency-Cost Trade-Off", *Economic Modelling*, 8(4): 1482-1488.

4. Bouhdaoui Y. and Bounie D., 2011, "Efficient Payments: How Much Does It Cost for the Central Bank?", Telecom ParisTech Working Paper.

5. Bouhdaoui Y., Bounie D. and Van Hove L., 2011c, “Production Costs, Seigniorage and Counterfeiting: Central Banks’ Incentives for Improving their Banknote Technology” (soumis à l’*International Journal of Central Banking*, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1652173).

6. Bouhdaoui Y., 2011, “The Social Cost of Denomination Structures”, Working paper Telecom ParisTech, http:papers.ssrn.com/sol3/papers.cfm?abstract_id=1876367.

CHAPTER 2

THE "PRINCIPLE OF INVARIANCE" IN CURRENCY SYSTEMS: A COMMENT ON CAIANIELLO *ET AL.*

Bouhdaoui Y.[1] , Bounie D.[2] and Van Hove L.[3]

2.1 Introduction

One of the responsibilities of central banks is to ensure the smooth functioning of payment systems. Any central bank is thus concerned with its "currency system"; that is, with how many different coins and banknotes it should provide to the public, as well as with their denominational values. In an article in this journal, Caianiello *et al.* (1982) analyze such currency systems by means of a systemic approach.[4] Specifically, they examine three aspects of "denominational structures". First, they demonstrate that, under certain assumptions, modular currency systems - systems in which each denomination is M times the one below it, with M a constant, integer number - are the more efficient, in that the average number of tokens (coins and/or banknotes) needed in a payment is minimized. Secondly, Caianiello *et al.* posit that the distribution of the tokens in circulation over the different denominations obeys a simple "principle of invariance", which holds that, in conditions of equilibrium, the average value of the coins and banknotes in circulation must equal the average value of cash transactions in the economy. Caianiello *et al.* rely on this principle to provide a theoretical justification for the "empirical law" advanced by Hentsch (1973), according to which the value in circulation of any given

[1]Telecom ParisTech, Department of Economics and Social Sciences, 46 rue Barrault, 75634, Paris Cedex 13; tel: +33 1 45 81 82 95; fax: +33 1 45 65 95 15; e-mail: yassine.bouhdaoui@telecom-paristech.fr

[2]Telecom ParisTech, Department of Economics and Social Sciences, 46 rue Barrault, 75634, Paris Cedex 13; tel: +33 1 45 81 73 32; fax: +33 1 45 65 95 15; e-mail: david.bounie@telecom-paristech.fr.

[3]ECON - 2C147, Vrije Universiteit Brussel (Free University of Brussels), Pleinlaan 2; B-1050 Brussels, Belgium; tel: +32 2 629 21 25; fax: +32 2 629 20 60; e-mail: Leo.Van.Hove@vub.ac.be; Corresponding author.

[4]Note that Caianiello *et al.* use the term "monetary systems", which we think is too broad.

denomination would be proportional to the square root of its face value. Thirdly and finally, Caianiello *et al.* turn their thermostatic approach into a thermodynamic one, and argue that the latter yields an improved understanding of the way in which a currency system reorganizes itself in the face of inflation.

In this comment, we concentrate on the second issue and challenge the principle of invariance. In Section 2, we first introduce Caianiello *et al.*'s notations and summarize their approach. In Section 3, we demonstrate that there is no identity between the average value of the coins and banknotes in circulation on the one hand and the average value of cash transactions on the other. In Section 4, we demonstrate that there is no invariance either; that is, contrary to what Caianiello *et al.* claim, the average value of the coins and banknotes in circulation *is* affected by changes in the module M. As a result, to the best of our knowledge, Hentsch's empirical "square root rule" still - after all these years - lacks a sound theoretical underpinning.

2.2 Caianiello *et al.*'s principle of invariance

The starting point of Caianiello *et al.*'s reasoning is that a currency system is a prime example of a system that is both hierarchical and self-organizing. They point out that, all over the world, central banks issue their currency in a limited number of denominations that recall the levels of a hierarchical system. As mentioned in the Introduction, in the first part of their article Caianiello *et al.* demonstrate that if one assumes payment with the exact amount and a uniform distribution of prices, modular currency systems are the more efficient.[5] This is why they concentrate on such systems throughout their article. Also, Caianiello *et al.* point out that hierarchical modular systems are a particularly interesting subclass "*as it has been possible to construct a "thermodynamics" for them. Thus, by using a principle of invariance, it is possible to learn to derive the structure through which the system is organized from the effect of its interaction with the external world*" (*o.c.*, p. 81).

[5] For the case where overpayment and the return of change is allowed, see Van Hove and Heyndels (1996) and Van Hove (2001).

More specifically, Caianiello *et al.* examine currency systems with $L+1$ levels and face values v_0, v_1,..., v_L such that the face value of the smallest denomination, v_0, equals 1, and the face value of the denomination on level $h+1$ is given by $v_{h+1} = M \cdot v_h$, with M the module. M is taken to be integer and strictly greater than 1 (otherwise the system would consist of one single denomination - of face value 1). Caianiello *et al.* then consider so-called refinements of the currency system. Refinements lower the module so as to introduce additional levels into the system besides the already existing ones (*o.c.*, p. 85). In general terms, in a refinement of the order of p the module goes from M to $M' = M^{1/p}$ (with p an integer) and the number of levels increases from $L+1$ to $pL+1$, with v_L fixed by hypothesis.[6] Obviously, p has to be greater than 1; otherwise there is no refinement. As an illustration, suppose that initially $M = 9$ and $L+1 = 3$ - implying that there are three denominations with face values of 1, 9, and 81 units. After a refinement of order 2, the module, M', is now 3 and two denominations have been added to the series, with face values of 3 and 27, respectively.

Caianiello *et al.* further define n_h as "*the number of elements belonging to level h*" (*ibidem*); that is, the number of tokens of face value v_h that are in circulation. The total number and value of the tokens in circulation are then given by:

$$N = \sum_{h=0}^{L} n_h,$$

and

$$V = \sum_{h=0}^{L} n_h \cdot v_h.$$

Hence, the "average value of an element of the system" - that is, the average value of the coins and banknotes in circulation - is:

$$< v > = \frac{1}{N} \sum_{h=0}^{L} n_h \cdot v_h. \tag{2.1}$$

[6]The value of the largest denomination is fixed in order to compare denominational structures over identical intervals.

Crucially, Caianiello *et al.* then conjecture that this average value $<v>$ "*cannot be an arbitrarily chosen quantity; it is fixed by the average value of the values of all the goods exchanged in an assigned time interval*" (*ibidem*) and, one should add, paid in cash. They also stress that this average price is assigned outside the currency system. It depends, among other things, on the general price level and on the popularity of competing payment instruments such as debit and credit cards, but not, Caianiello *et al.* stress, on the choice of the module of the currency system.

This is where the self-organizing nature of currency systems enters into play. Caianiello *et al.* emphasize that currency systems can, if the need arises, alter their structure in response to external shocks. Concretely, the distribution of the tokens in circulation will change when the distribution of cash payments changes. Caianiello *et al.* build on this to formulate their fundamental "principle of invariance": the average value of the coins and banknotes in circulation is not only equal to the average value of cash transactions in conditions of equilibrium, this will also remain the case when the currency system is refined. In their own words: "*the principle of invariance of* $<v>$ *for refining transformations*" holds that "*the mean value of an element of a monetary system is invariant under* $M \to M^{1/p}$ *refinement, regardless of what p is*" (*ibidem*). According to Caianiello *et al.*, this is very promising as "*the discovery of a principle of invariance has always been a powerful analytical tool in all fields of scientific research*" (*ibidem*).

From this principle of invariance, Caianiello *et al.* then derive the law of distribution of the tokens over the different levels of a modular currency system. Interestingly, they find - as suggested by the empirical findings in Hentsch (1973, 1975, 1983 and 1985) - that the values in circulation are indeed proportional to the square roots of the face values of the denominations; the factor of proportionality being the number of units of the smallest denomination:

$$A_h = n_0 \cdot \sqrt{v_h}, \tag{2.2}$$

with A_h, the value in circulation in the form of tokens of face value v_h.[7]

In short, Caianiello *et al.* appear to provide a theoretical justification for Hentsch's

[7] Note that Hentsch does not specify the factor of proportionality.

square root law. In a final step, they look at real-life data for 11 countries and in all cases find "substantial confirmation" of Hentsch's currency distribution rule.[8]

2.3 No identity

The theoretical approach that we have just described is clearly innovative and intellectually attractive. However, there are major problems with the proof. In this section, we demonstrate that the identity between on the one hand the average value of the coins and banknotes in circulation and on the other hand the average value of cash transactions, does not hold. We start by assuming, as Caianiello *et al.* implicitly do, that the velocities of circulation of all tokens - the number of times during the observation period that a specific denomination is involved in transactions - are equal.[9]

Let us start by denoting by $\bar{a}_h$ the average frequency of use per transaction of the token with face value v_h. We also introduce the velocity of circulation q_h:

$$q_h = \frac{\bar{a}_h \cdot N_T}{n_h},$$

with N_T the number of transactions during the observation period. Note that the numerator, $\bar{a}_h \cdot N_T$, is the total demand for denomination h during the same period.

[8] As an aside, while many of the circulation figures per denomination are - when transferred into logs - to a greater or less extent aligned along a line with a slope of 0.5, Caianiello *et al.* fail to check whether the intercept is also in accordance with the distribution law that they have derived. Indeed, given Equation (2.2), $\log A_h = \log n_0 + 0.5 \log v_h$, so that the intercept should equal $\log n_0$.

[9] Caianiello *et al.* assume that the velocities of circulation of all tokens equal 1, which is a particular case of ours. Indeed, they seem to overlook that there is a difference between the number of tokens exchanged in a given period (the flow) and the number of tokens in circulation at a given point in time (the stock). As a matter of fact, in Section 2 of their article (where they derive formulas for the flows), they define N as "*the minimum number of coins which are necessary to obtain all the integers between zero and* $M^L - 1$" (*o.c.*, p. 84). In Section 3 (where they reason in terms of stocks), they use the same symbol to designate "the total number of elements" in the system; that is, in circulation (*o.c.*, p. 85). [As will be shown below, this confusion lies at the heart of the fatal error in their proof.] The gap between their Sections 2 and 3 can only be closed by assuming that the velocity of circulation of all tokens is 1; that is, that all tokens are used once and only once in the period considered.

Rearranging the terms, we have:

$$n_h = \frac{\bar{a}_h \cdot N_T}{q_h}. \tag{2.3}$$

Since the velocity of circulation is assumed to be the same for all denominations, replacing (2.3) in (2.1) gives:

$$< v > = \frac{\sum_h \bar{a}_h \cdot v_h}{\sum_h \bar{a}_h}.$$

Under the assumption of exact payment, we recognize the average price $\bar{x} = \sum_h \bar{a}_h \cdot v_h$ and the average number of tokens per transaction $\bar{n} = \sum_h \bar{a}_h$. Thus:

$$< v > = \frac{\bar{x}}{\bar{n}}.$$

Hence for $<v>$ to equal $\bar{x}$ as Caianiello et *al.* claim, $\bar{n}$ has to equal 1. This will only be the case when there is a denomination for each price; otherwise $\bar{n}$ is strictly greater than unity. In short, the identity formulated by Caianiello et *al.* is far from a general rule. Obviously, identical velocities of circulation are not observed in real life, but this assumption is not crucial for our proof.

2.4 No invariance

This said, Caianiello *et al.*'s derivation of their distribution law rests essentially on the invariance of $<v>$, and less so on the link with the average price.[10] Hence, if there are other explanations for the invariance of $<v>$ besides the (defective) link with the average price, Caianiello *et al.*'s distribution law still stands.

However, in this Section we demonstrate that the principle of invariance is shaky, and, more in particular, that $<v>$ increases when the denominational structure is refined. We present two alternative proofs. The first is formal; the second is more intuitive. The

[10] Note that Caianiello *et al.* do not demonstrate that $<v>$ equals $\bar{x}$; they simply state it.

first needs the assumption that the distribution of prices is uniform; the second is more general. But the two proofs have the same starting point, namely the observation that under exact payment, the total value of coins and notes exchanged in transactions is not affected by refinements (because it equals the sum of all prices). Hence, we can focus on the number of tokens exchanged, and if we can demonstrate that this number is affected, then the average value of the coins and notes that are exchanged is also affected.[11] If we assume in addition that the velocity of circulation is the same for all denominations (and is not affected by refinements), then the same is true for the average value of coins and notes in circulation.

Given this, a first way to demonstrate that the number of tokens exchanged is sensitive to refinements builds on the first part of Caianiello *et al.*'s article. In that part, Caianiello *et al.* (o.c., p. 84) demonstrate that, assuming exact payment, a modular currency system and a uniform distribution of prices, the minimum number of tokens needed to form all integer values up to the largest denomination is given by:

$$N = \frac{L}{2}(M-1)v_L.$$

Let us then consider a refinement from M to $M' = M^{1/p}$, with $p > 1$. The goal is then to demonstrate that this will lower the average number of tokens exchanged - the intuition being that an increase in the number of denominations in a given interval, which is the essence of a refinement, will decrease the number of tokens exchanged. Algebraically, we need to demonstrate that:

$$\frac{L'}{2}(M'-1)v_{L'} < \frac{L}{2}(M-1)v_L. \tag{2.4}$$

Given that, by construction, a refinement leaves unchanged the nominal value of the

[11]Note that this is where Caianiello *et al.* go wrong. In Equation (1), on p. 86, they (implicitly) make the jump from N as designating the number of tokens exchanged to N as shorthand for the number of tokens in circulation; see footnote 6 on this. The same equation is also the starting point of their proof that the mean value remains unchanged under a p-refinement. Unfortunately, they use the same N in their expressions for both $<v>$ and $<v>'$, and thus in effect only demonstrate that the value of the coins and notes exchanged remains the same - which is self-evident; see main text.

largest denomination (so that $v_{L'} = v_L$), Equation 2.4 can immediately be simplified to:

$$L'(M'-1) < L(M-1). \tag{2.5}$$

The invariance of the value of the largest denomination can also be exploited to derive a relationship between the number of denominations in the original and in the refined denomination structure, namely:

$$M^L = (M')^{L'} = (M^{\frac{1}{p}})^{L'},$$

so that:

$$L' = pL. \tag{2.6}$$

Substitution of (2.6) in (2.5) gives:

$$pL(M^{\frac{1}{p}} - 1) < L(M-1),$$

$$p(M^{\frac{1}{p}} - 1) < M - 1.$$

If we then define a function $f(M)$ as:

$$f(M) = p(M^{\frac{1}{p}} - 1) - M + 1,$$

we only need to demonstrate that this function is strictly negative for (2.4), and thus the entire proof, to hold. To that end, it can be noted that:

$$f'(M) = M^{\frac{1}{p}-1} - 1 = \frac{M^{\frac{1}{p}} - M}{M}.$$

Given that $M > 1$ and that $p > 1$, $M > M^{1/p}$ so that the expression in (7) is negative, and $f(M)$ is downward sloping. Moreover, since $f(1) = 0$, all values in the relevant interval are negative.

As announced, there is also a more intuitive way to demonstrate that the number of tokens exchanged is sensitive to refinements. As our example in Section 2 illustrates, a refined currency system, by construction, includes all the denominations of the non-refined one. As a result, the refined system will be at least as efficient as the original, and for it to be more efficient, it is sufficient that there is one amount in any given price distribution that involves strictly fewer tokens.

Actually, there are many such amounts. As explained in Section 2, refinements add denominations to a currency system. As a result, amounts that are equal to the face value of one of these additional denominations can be paid with a single token under the refined series, but will require $M^{1/p}$ tokens of the next-lower denomination under the original series. In our example in Section 2, an amount of 27 requires three tokens of 9 prior to the refinement, but only one token of 27 with the refined series. Given its higher "density", a refined denominational structure will thus clearly always be more efficient than the original. We provide a formal proof of this intuition in the Appendix.

To sum up, we have demonstrated that a p-refinement lowers the number of tokens exchanged. Conversely, under exact payment, their total value remains the same. Hence, the average value of the tokens that are exchanged increases, and assuming identical velocities of circulation across denominations, the same is true for the average value of the tokens in circulation. Hence, the principle of invariance does not hold.

2.5 Conclusion: no identity, no invariance, no distribution law

In this comment, we have demonstrated, first, that there is no identity between the average value of the coins and banknotes in circulation on the one hand and the average value of cash transactions on the other, and, second, that the former is not invariant to refinements of a modular currency system. As a result, Caianiello *et al.*'s law of distribution of the tokens over the different levels of a denominational structure - which builds on the principle of invariance - is incorrect, and the theoretical justification for Hentsch's empirical "square root law" falls apart. Finally, since Caianiello *et al.*'s thermostatic approach is flawed, their thermodynamic analysis of how a currency system reorganizes itself in the face of inflation is also flawed. This does not, however, preclude that an amended version of the thermodynamic approach - initially sketched in Caianiello (1977) - might not be a valid framework to study hierarchical modular systems in general and, who knows, even currency systems.

Acknowledgments

We are grateful to Steven Vanduffel and the three referees for helpful comments.

Literature cited

Caianiello E.R., 1977. Some remarks on organization and structures. *Biological Cybernetics*, 26, 151-158.

Caianiello E.R., Scarpetta G. and Simoncelli G., 1982. A systemic study of monetary systems. *International Journal of General Systems*, 8(2), 81-92.

Hentsch J.C., 1973. La circulation des coupures qui constituent une monnaie. *Journal de la Société de Statistique de Paris*, 114(4), 279-286.

Hentsch J.C., 1975. Calcul d'un critère qualitatif pour les séries de valeurs définissant l'échelonnement des signes monétaires. *Journal de la Société de Statistique de Paris*, 116(4), 310-315.

Hentsch J.C., 1983. Distribution de la monnaie fiduciaire entre les coupures qui la représentent. *Journal de la Société de Statistique de Paris*, 124(4), 263-272.

Hentsch J.C., 1985, Distribution de la monnaie fiduciaire entre les coupures. Les paramètres et l'indice des prix. La distribution normalisée. *Journal de la Société de Statistique de Paris*, 126(4), 139-144.

Van Hove L. and Heyndels B., 1996. On the optimal spacing of currency denominations. *European Journal of Operational Research*, 90, 547-552.

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.

Appendix: the efficiency of refined and non-refined currency systems

In this Appendix, we demonstrate that the number of tokens exchanged decreases when the currency system is refined, *i.e.* that $\bar{n'} < \bar{n}$. To that end, we show first that $\forall x,\ n'(x) \leq n(x)$.

Assuming exact payment, we denote by $a_h(x)$ the frequency of use of denomination h in composing an amount x under the non-refined currency system:

$$x = \sum_{h=1}^{L} a_h(x) M^h,$$

such that $n(x) = \sum_{h=1}^{L} a_h(x)$ is minimum.

It is then useful to introduce the notation $m_h(x) = a_h(x)M^h$. These $m_h(x)$ can be thought of as "sub-amounts" of x:

$$x = \sum_{h=1}^{L} m_h(x). \tag{2.7}$$

To demonstrate that $n'(x) \leq n(x)$, it is enough, given (2.7), to show that the sub-amounts $m_h(x)$ can be paid with fewer than $a_h(x)$ tokens using the p-refined currency system.

Denoting by $(v'_h)_h$ the face values of the p-refined currency system, which is composed of $L' = pL$ denominations, we have:

$$v'_h = M'^h = M^{h/p}.$$

In words, all "old" denominations are p-multiples of the new denominations. The $a_h(x)$ units of old denominations can thus be composed with units of the new denominations in the following way:

$$a_h(x) = \sum_{k=1}^{k_0} a'_k(x) \cdot M'^k,$$

with $M'^{k_0} \leq a_h(x) \leq M'^{k_0+1}$.

Therefore:

$$m_h(x) = a_h(x) \cdot M'^{ph} = \sum_{k=1}^{k_0} a'_k(x) \cdot M'^{ph+k},$$

that is, with the p-refined system, $m_h(x)$ is payable with $\sum_{k=1}^{k_0} a'_k(x)$ tokens.

By definition we have:

$$0 \leq a'_k(x) \leq M'.$$

More intuitively, the number of tokens of level k that are used, $a'_k(x)$, cannot exceed the module M' because otherwise they are partly replaced by one token of the higher-level denomination.

Therefore:

$$\sum_{k=1}^{k_0} a'_k(x) \leq k_0 M^{1/p}. \tag{2.8}$$

On the other hand, we have:

$$M^{k_0/p} \leq a_h(x). \tag{2.9}$$

Given that $k_0 M^{\frac{1}{p}} \leq M^{\frac{k_0}{p}}$, we deduce from (2.8) and (2.9) that:

$$\sum_{k=1}^{k_0} a'_k(x) \leq a_h(x).$$

That is, $m_h(x)$ is payable with the same number or fewer tokens using the p-refined system.

Hence, $n'(x) \leq n(x)$ and the strict inequality depends on the amount x. For instance, amounts that are equal to the face value of one of the additional denominations of the refined system can be paid with a single token under the refined series but need strictly more than one token under the non-refined system.

CHAPTER 3

DISTRIBUTION DES TRANSACTIONS EN ESPÈCES ET EFFICACITÉ DES PAIEMENTS EN EUROS

Bouhdaoui Y.[1] et Bounie D.[2]

3.1 Introduction

De nombreuses contributions théoriques et empiriques se sont interrogées ces dernières années sur la structure optimale d'un système de divisions monétaires (SDM ci-après).[3] Dans ces études, un SDM est un ensemble de divisions ou valeurs faciales - par exemple 1€, 2€, 5€, etc. - et son efficacité se réfère au nombre moyen de pièces et billets utilisés pour régler les transactions dans une économie. Plus précisément, les agents économiques sont censés payer selon le "principe du moindre effort", *i.e.* en échangeant le nombre minimum de coupures par transaction. Un SDM est alors considéré plus efficace qu'un autre si le nombre moyen de pièces et billets qu'il nécessite dans les échanges est plus faible. Après de nombreuses controverses, Van Hove et Heyndels (1996) et Van Hove (2001) ont montré que le SDM le plus efficace est celui dont le facteur d'espacement (entier) entre les divisions est égal à deux (1, 2, 4, etc., dit "puissances de 2"). De même, Bouhdaoui *et al.* (2011) ont démontré que l'efficacité du SDM européen, caractérisé par un facteur d'espacement moyen de 2,2, est très proche de celle de la structure optimale "puissances de 2".

Les motivations et les enjeux de ces recherches sont multiples. Tout d'abord, la structure d'un SDM, *i.e.* le nombre total et la valeur faciale de chaque division, affecte le nombre de pièces et billets échangés entre les agents économiques. Or, un SDM qui induit un usage moyen plus élevé de pièces et billets dans les échanges peut être plus

[1]Telecom ParisTech, Département d'Economie et de Sciences Sociales.

[2]Telecom ParisTech, Département d'Economie et de Sciences Sociales.

[3]Caianiello *et al.* (1982), Sumner (1993), Telser (1995), Van Hove et Heyndels (1996), Tschoegl (1997), Wynne (1997), Van Hove (2001), Lee *et al.* (2005), Franses et Kippers (2007), Bounie et Houy (2009), etc.

coûteux pour ces acteurs (coûts de détention, fraude, etc.) (Van Hove, 2001). Ensuite, Bouhdaoui *et al.* (2011) et Bouhdaoui et Bounie (2010) ont montré, sous certaines hypothèses, qu'un SDM plus efficace implique un volume de production moins élevé de pièces et billets, d'où une baisse des coûts de production pour la banque centrale qui induit, de causes à effets, une baisse des coûts des transactions payées en espèces. Enfin, Chen (1976), Chen et Tsaur (1983) et Boeschoten et Fase (1989) ont montré que l'efficacité d'un SDM affecte la demande de monnaie et, *in fine*, le niveau général des prix.

Dans cet article, nous levons une hypothèse restrictive émise généralement dans les travaux théoriques, celle de distribution uniforme des transactions en espèces dans l'économie. Van Hove (2001, p. 1020) souligne en effet que *"l'hypothèse d'une distribution uniforme des transactions en espèces, dans le cadre de l'approche du principe du moindre effort, ne reflète pas la réalité. En Hollande, des sondages (Boeschoten 1992, p. 77) ont montré que la distribution est en réalité lognormale. Or, puisque les transactions de faibles montants sont beaucoup plus fréquentes, les petites divisions contribuent davantage à l'efficacité globale du SDM que les grandes. En d'autres termes, les écarts dans les régions supérieures du SDM ne doivent pas affecter grandement la capacité à procéder efficacement aux paiements en espèces"*. Pour ce faire, nous utilisons une distribution empirique des transactions en espèces dans l'économie française établie dans le cadre d'une enquête sur les comportements de paiement des français en 2005, et nous comparons l'efficacité des paiements obtenue à celle d'une distribution uniforme.

Les résultats des simulations montrent, premièrement, que le degré de concurrence entre les instruments de paiement et la proportion élevée de montants arrondis dans la distribution empirique améliorent l'efficacité du SDM européen en comparaison de celle obtenue avec la distribution uniforme. Plus précisément, nous établissons que le nombre moyen de coupures utilisées dans les transactions pour les montants observés est environ deux fois plus faible que celui obtenu lorsque les montants sont répartis uniformément. Deuxièmement, les simulations mettent en évidence qu'une distribution uniforme des montants payés en espèces surestime la fréquence moyenne d'utilisation des valeurs faciales élevées et sous-estime celle des petites, du moins la division de “un euro” en raison de la proportion élevée de montants faibles dans la distribution empirique.

Ces résultats contribuent à la littérature sur deux points en particulier. Premièrement, l’analyse des données collectées permet d’établir une occurence élevée des montants multiples de 5 et 10 dans la distribution observée des transactions en espèces, et de fait, qu’il existe une interaction entre les valeurs faciales du SDM européen et les montants payès en espèces. Ce résultat tend à remettre en cause les travaux théoriques qui comparent l’efficacité de plusieurs SDM en se basant sur une même distribution des transactions en espèces. En ce sens, notre enquête confirme en partie le modèle théorique de Lee *et al.* (2005) qui proposent un modèle qui endogénise les prix et les valeurs faciales du SDM. Deuxièmement, nos simulations montrent que les divisions de valeurs faciales supérieures à 10 euros contribuent relativement peu à l’efficacité globale d’un SDM (environ 17%), confirmant ainsi l’assertion de Van Hove (2001).

Le reste du document est structuré comme suit. Dans la section 2, nous introduisons quelques notations, nous précisons les principales hypothèses et nous décrivons les données utilisées. Dans la section 3, nous présentons et discutons les résultats des simulations. La section 4 conclut ce travail.

3.2 Notations, hypothèses et données

Nous souhaitons étudier l’influence des montants payés en espèce sur le nombre moyen de coupures échangées par transaction. Dit autrement, nous nous intéressons à l’impact d’un changement de distribution des transactions sur l’efficacité d’un système de divisions monétaires. Pour ce faire, il est nécessaire de définir au préalable quelques notations et hypothèses.

Tout d’abord, nous définissons un SDM comme un ensemble d’entiers strictement positifs qui comprend l’entier 1 et qui représente les valeurs faciales des divisions. Ainsi, par SDM^{EU}, on désigne le SDM actuellement utilisé dans la zone euro auquel on a exclu les trois premières divisions : $SDM^{EU} = \{1,2,5,10,20,50,100,200,500,1000,2000,5000\}$

(en dixièmes d'euros).[4] Ensuite, nous supposons qu'un agent représentatif dispose d'un stock infini d'unités des différentes valeurs faciales du SDM européen qui lui permet de réaliser les transactions sans contraintes sur les divisions.

Cet agent est confronté à une distribution de montants à régler en espèces. Dans notre cas, nous comparons deux distributions : une distribution où les montants sont répartis uniformément sur un segment et une distribution empirique des transactions obtenue à partir d'une enquête administrée de Mars à Mai 2005 sur un échantillon de personnes représentatives de la population française. Plus précisément, nous avons suivi 1386 personnes sur une période de huit jours durant lesquels elles ont répertorié les montants de tous leurs achats payés en espèces (arrondis à la première décimale).[5]

L'enquête réalisée en partenariat avec l'institut de sondage GfK/ISL se déroulait plus précisément en deux étapes. Dans la première, les personnes étaient interrogées en face-à-face afin de recueillir des informations sur leur détention en matière de moyen de paiement, sur leur appréciation de ces différents moyens de paiement, etc.[6] Dans une deuxième étape, les personnes s'engageaient à remplir un carnet de dépenses sur lequel elles devaient reporter plusieurs informations telles que la valeur de l'achat, le type de bien acheté, le type de commerce fréquenté, le moyen de paiement utilisé, etc.

Après saisie des différentes informations sur les carnets de dépenses, nous constatons que les personnes ont réalisé 16,692 achats au total pour une valeur de 541 583 euros ; en d'autres termes, une personne réalise en moyenne 12 achats sur huit jours pour une valeur de 32 euro environ. Les espèces sont les moyens de paiement les plus fréquemment utilisés par les Français ; 62,4% de l'ensemble des paiements sont réglés à l'aide des espèces (soit 10 419 paiements). La valeur moyenne d'un achat réglé en espèces s'élève

[4]Nous précisons ci-après les raisons pour lesquelles nous excluons les trois premières divisions.

[5]En vue de simplifier le report des informations sur les carnets de dépenses, nous avons demandé aux enquêtés de ne reporter que la valeur de leurs achats arrondie à la première décimale. Cette contrainte implique et justifie donc l'abandon dans le SDM européen des trois premières divisions à savoir les divisions 1, 2 et 5 centimes d'euro.

[6]Les personnes enquêtées ont été recrutées directement en face-à-face et n'avaient fait l'objet d'aucune panélisation antérieure. La méthode d'échantillonnage utilisée est celle des quotas. Les critères utilisés sont la région (région ZEAT (Zone d'Etudes et d'Aménagement du Territoire)), l'agglomération, le sexe croisé avec la situation professionnelle (actif/inactif), l'âge, la profession et le type de logement.

à 10,8 euro et reste très inférieur à la valeur moyenne de tous les achats en raison de l'usage important des chèques en France. Le chèque reste en effet le troisième moyen de paiement le plus fréquemment utilisé avec une part de marché de 13,8%.

La répartition du nombre des montants en espèces en fonction de la valeur de la transaction est représentée à la Figure 3.1. Nous observons que les montants sont répartis entre 0,2 et 1100 euros (2 et 11000 dixièmes d'euros) et que le nombre de transactions payées en espèces décroît avec la valeur de l'achat. En outre, nous constatons que le montant le plus fréquemment payé en espèces dans l'économie française est celui de 80 centimes d'euros (8 dixièmes d'euros) et qu'il existe des pics sur les valeurs arrondies qui correspondent, le plus souvent, à des multiples de 5 et de 10.

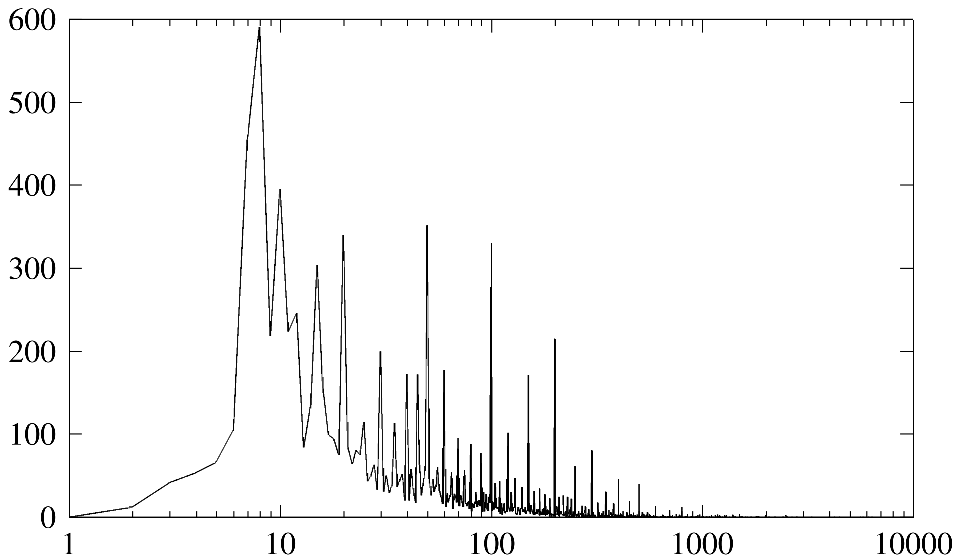

Figure 3.1 – Nombre d'occurences de chaque montant de la distribution empirique (en dixièmes d'euros, échelle logarithmique).

Enfin, notre dernière hypothèse porte sur le comportement de paiement des agents économiques. Nous supposons que les agents, consommateurs et marchands, règlent les transactions en suivant le "principe du moindre effort", *i.e.* en échangeant un nombre minimum de pièces et billets pour chaque transaction. Plus formellement, étant donné un SDM composé de J divisions de valeurs faciales $\{v(j)\}_{j\in[\![1,J]\!]}$, le montant x est alors payé efficacement en combinant des unités de valeurs faciales représentées par la série d'entiers $\{n(x,j)\}_{j\in[\![1,J]\!]}$ tel que :

$$x = \sum_{j=1}^{J} n(x,j)v(j),$$

avec un nombre de coupures échangées :

$$n(x) = \sum_{j=1}^{J} |n(x,j)|,$$

minimal.

Cependant, pour la plupart des montants, plusieurs solutions efficaces peuvent exister. On distingue les $K(x)$ solutions du montant x par le suffixe k :

$$x = \sum_{j=1}^{J} n_k(x,j)v(j),$$

et on définit donc la fréquence d'utilisation a de la division j pour le montant x comme suit :

$$a(x,j) = \frac{1}{K(x)} \cdot \sum_{k=1}^{K(x)} |n_k(x,j)|.$$

Il en résulte que la fréquence moyenne d'utilisation, $a(j)$, d'une division j pour une distribution de transactions en espèces $\mathscr{D}$ est :

$$a(j) = \frac{1}{N_T} \cdot \sum_{x\in\mathscr{D}} a(x,j),$$

avec N_T, le nombre de transactions.

En sommant les fréquences d'utilisation moyennes des divisions, on obtient le nombre moyen de coupures échangées par transaction. Ce dernier caractérise l'efficacité du SDM sur la distribution étudiée. On le note :

$$\overline{n} = \sum_{j \in [\![1,J]\!]} a(j).$$

Pour illustrer l'algorithme de paiement, considèrons le règlement des transactions 1,60 euros (16 dixièmes d'euros) et 0,30 euros (3 dixièmes d'euros) avec le SDM^{EU}. Pour le premier montant, un paiement efficace peut être effectué de trois manières : *i.* 1 euro plus 50 centimes plus 10 centimes ; *ii.* 2 euros et 10 centimes et un retour de 50 centimes ; *iii.* 2 euros et le retour de deux unités de 20 centimes. De même, pour le second prix, un paiement efficace peut être effectué de deux façons, *i.e.* par le paiement de 20 centimes et 10 centimes et par le paiement de 50 centimes qui est suivi d'un rendu de 20 centimes.

Pour résumer, il existe trois solutions pour la première transaction qui mobilisent trois unités ($n(16) = 3$) et nous obtenons les fréquences d'utilisation suivantes : $a(16,1) = \frac{2}{3}$, $a(16,2) = \frac{2}{3}$, $a(16,3) = \frac{2}{3}$, $a(16,4) = \frac{1}{3}$ et $a(16,5) = \frac{2}{3}$. De même, les deux solutions de la seconde transaction impliquent deux coupures ($n(3) = 2$) et on obtient $a(3,1) = \frac{1}{2}$, $a(3,2) = 1$ et $a(3,3) = \frac{1}{2}$. Ainsi, en considérant la distribution $\mathscr{D}$ constituée des montants 1,60 euros et 30 centimes ($\mathscr{D} = \{16,3\}$), on obtient les fréquences moyennes par division $a(1) = \frac{7}{12}$, $a(2) = \frac{5}{6}$, $a(3) = \frac{7}{12}$, $a(4) = \frac{1}{6}$ et $a(5) = \frac{1}{3}$. Enfin, le nombre moyen de coupures par transaction pour $\mathscr{D}$, obtenu en sommant les fréquences moyennes $a(j)$, s'élève à $2,5$ coupures par transaction.[7]

Dans la section suivante, nous présentons le mode opératoire de la comparaison et nous exposons les résultats obtenus.

3.3 Simulations et résultats

L'objectif de cette partie est de comparer l'efficacité des paiements obtenue dans le cadre d'une distribution uniforme et empirique des transactions en espèces. Avant de

[7]Notons que le poids d'une transaction dans la distribution est indépendant du nombre de ses solutions en appliquant le principe du moindre effort. Le poids de chaque montant est lié à son nombre d'occurences dans la distribution.

commenter les résultats des simulations, nous précisons au préalable la méthodologie.

3.3.1 Méthode

Pour comparer l'efficacité du SDM dans le cadre de deux distributions, nous devons définir tout d'abord un intervalle d'étude. Pour cela, nous adoptons l'approche de Cramer (1983) en définissant un intervalle couvrant un nombre entier de cycles de toutes les divisions à l'exception de la plus grande, *i.e.* $[v(1); v(J) - v(1)]$ avec $v(1)$ et $v(J)$ les valeurs faciales de la plus petite et de la plus grande division. Dans notre cas, $J = 12$, $v(1) = 1$, $v(12) = 5000$ dixièmes d'euros et notre intervalle d'étude s'écrit [1, 4999]. Cette méthode nous conduit à tronquer la distribution empirique des transactions en excluant les montants supérieurs ou égaux à 500 euros. La distribution ainsi obtenue se compose de 10412 transactions.[8] En revanche, sur le même intervalle [1, 4999], la distribution uniforme ainsi définie ne comprend que 4999 montants (espacés de 1 dixième d'euros).[9]

3.3.2 Simulations

Un programme informatique, développé pour cette étude, nous permet de simuler sur les deux distributions le principe du moindre effort. Ce dernier, comme précisé dans les hypothèses, correspond au comportement de paiement des agents effectuant les transactions en espèce et permet ainsi de mesurer l'efficacité du SDM^{EU} sur chaque distribution.

Deux principaux commentaires peuvent être tirés de la simulation. Tout d'abord, le nombre moyen de pièces et billets échangés par transaction est environ deux fois plus élevé pour la distribution uniforme que pour la distribution empirique : $4,78$ contre

[8]Il est intéressant de souligner que cette méthode nous conduit à exclure seulement sept montants de l'échantillon initial, le montant le plus élevé s'élevant à 1 100 euro. La troncature de la distribution empirique réduit donc les écarts entre les fréquences d'utilisation des divisions des deux distributions dans la mesure où la construction d'une distribution uniforme sur l'intervalle [1 ; 1 100] euros implique une augmentation de 5 999 montants supplémentaires. Cet accroissement considérable du nombre de montants pour la distribution uniforme par rapport à la distribution empirique aurait alors pour incidence d'accroitre le nombre moyen de pièces et billets utilisés.

[9]Il est important de souligner toutefois que la différence du nombre de montants dans les intervalles d'étude n'ont pas d'impacts sur les résultats puisque nous raisonnons sur des moyennes.

$2,33$. Ensuite, la Figure 3.2 montre que la fréquence moyenne d'utilisation d'une division dans le cas de la distribution empirique est toujours inférieure à celle de la distribution uniforme à l'exception de la division de "un euro". Globalement, l'écart des fréquences moyennes d'utilisation des divisions s'accentue avec l'augmentation de la valeur faciale.[10] Ce résultat confirme l'assertion de Van Hove (2001, p. 1020) selon laquelle les plus petites divisions contribuent davantage à l'efficacité globale d'un SDM. Nous observons en effet une décroissance rapide des fréquences moyennes d'utilisation des divisions pour la distribution empirique à partir de 2 euros, fréquences au demeurant quasi-nulles pour les divisions supérieures à 100 euros. Dès lors, la contribution des divisions supérieures ou égales à 10 euros à l'efficacité globale du SDM européen s'élève seulement à 17,4%.[11] En comparaison, cette contribution s'élève à 48,5% dans le cas de la distribution uniforme.

Il est important de souligner pour conclure que les faibles fréquences d'utilisation des plus grandes divisions illustrées à la Figure 3.2 pour la distribution empirique sous-estiment la demande du public pour ces divisions. En effet, des travaux empiriques ont montré, à de multiples reprises, que la demande pour les grandes coupures n'est pas nécessairement liée à un motif de transaction pour les achats de la vie courante mais à un motif de précaution ou à un motif de transaction pour des activités illégales et/ou non officielles (Boeschoten et Fase, 1992). Dans un souci de rendre compte de la demande globale du public pour ces divisions, notre mesure de la demande pour ces valeurs faciales élevées devrait donc être logiquement corrigée à l'image des travaux de Hentsch (1973, 1983, 1985) et Van Hove et Vuchelen (1996).

3.3.3 Discussion

Nous pouvons commenter les deux résultats décrits dans la section précédente.

[10] Lorsqu'on réalise un test d'égalité entre les moyennes des 4 premières divisions, les différences sont significatives au seuil de 15% seulement. En revanche, sur l'ensemble des divisions, les différences sont significatives au seuil de 1%.

[11] Nous définissons formellement la contribution de ces divisions à l'efficacité globale du SDM comme le ratio du nombre moyen d'échanges qu'ils induisent et du nombre moyen total échangé par transaction. Formellement, on peut l'écrire comme suit : $\frac{\sum_{j \in [\![v(7),v(12)]\!]} a(j)}{\bar{n}}$. Le Tableau 3.I en annexe donne les valeurs des fréquences d'utilisation moyennes des divisions.

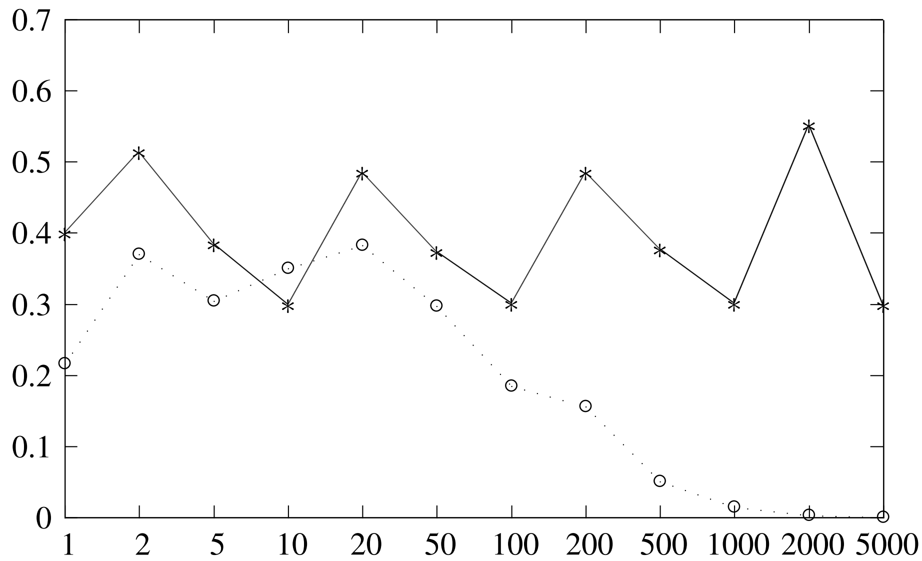

Figure 3.2 – Fréquence moyenne d'utilisation par division pour les ditributions empirique (ligne pointillée) et uniforme (ligne pleine) des transactions en espèces.

D'une part, le faible nombre moyen de pièces et billets échangés par transaction dans la distribution empirique s'explique par la combinaison de deux effets : un faible nombre de montants uniques et une proportion élevée des montants arrondis parmi ces derniers.[12]

Pour le premier effet, on constate qu'il existe seulement 606 montants uniques dans la distribution empirique parmi les 4999 possibles que compte l'intervalle d'étude, soit 12% de la totalité des montants de la distribution uniforme. Ce constat peut s'expliquer notamment par la concurrence des autres moyens de paiement. Nous connaissons en effet pour chaque personne de notre échantillon, les usages de leur carte de paiement et de leur chèque sur le même intervalle d'étude. Un simple calcul montre qu'il existe à présent 1286 montants uniques et que 14% des montants possibles sur l'intervalle sont réglés uniquement à l'aide de moyens de paiement alternatifs. Il est important de souligner que ce constat est probablement spécifique à la France dans la mesure où les espèces sont beaucoup plus utilisées dans d'autres pays comme l'Allemagne, l'Autriche,

[12]Un montant unique désigne un élément de notre intervalle d'étude qui comprend, par définition pour la distribution uniforme, 4999 montants uniques.

les Etats-Unis ou encore le Japon.

Pour le deuxième effet, la Figure 3.1 met en évidence des pics généralement lorsque la transaction est un multiple de 5 et de 10. Ces montants sont dits arrondis car ils correspondent également souvent à la valeur faciale d'une division et sont donc payables en une seule coupure. La Figure 3.3 qui décrit le pourcentage des montants qui impliquent 1, 2 ou plus de coupures dans le cas des distributions empirique et uniforme illustre ce constat : 90% des transactions de la distribution empirique nécessitent moins de trois coupures, alors que cette même proportion s'élève seulement à 13% pour la distribution uniforme.

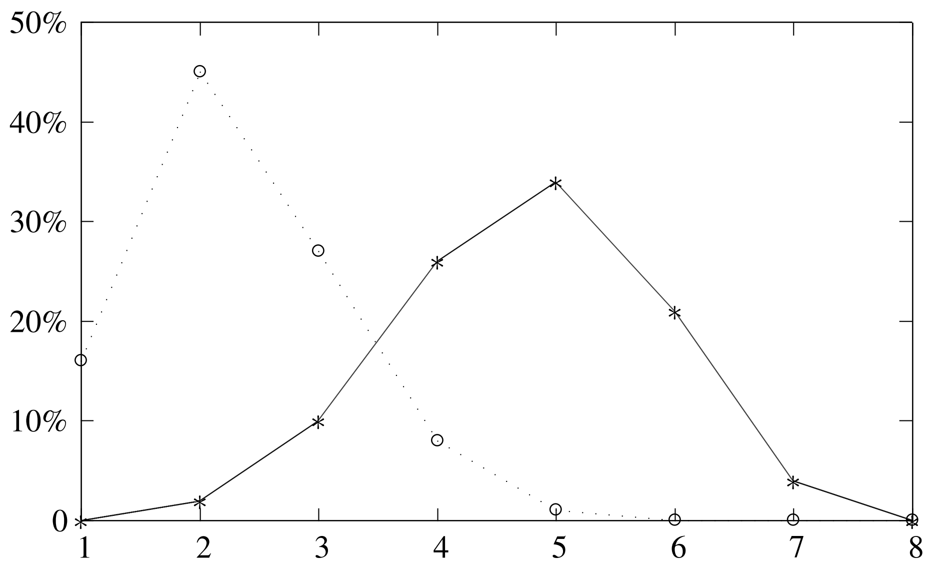

Figure 3.3 – Pourcentage des montants par nombre de coupures échangées dans les cas de distribution empirique (ligne pointillés) et uniforme (ligne pleine).

D'autre part, les écarts observés dans les fréquences d'utilisation des divisions entre les deux distributions (Figure 3.2) s'expliquent par les faibles valeurs des transactions de la distribution empirique : la transaction moyenne s'élève seulement à 10,4 euros et 82% des transactions sont inférieures à 20 euros. En comparaison, la transaction moyenne s'élève à 250 euros pour la distribution uniforme et, par définition, 50% des montants sont supérieurs à la transaction moyenne. La comparaison des Figures 3.4 et 3.5 illustrent ces observations : la proportion très faible des montants supérieurs à 100 euros (1000 dixièmes d'euros) dans la distribution empirique explique les vides observés dans la Figure 3.4 dans cette région, cette même proportion est de 80% dans la Figure 3.5 (elle apparaît plus réduite du fait de l'échelle logarithmique). En termes de nombre de coupures échangées, la Figure 3.5 montre que huit pièces et billets peuvent être nécessaires pour régler certaines transactions dans la distribution uniforme, impactant ainsi directement la demande pour les divisions de valeurs faciales élevées. Ainsi, caractérisée par une plus grande proportion de montants élevés, la distribution uniforme induit une surutilisation des grosses coupures et, également dans une certaine mesure, des plus petites divisions qui sont utilisées pour rendre la monnaie.

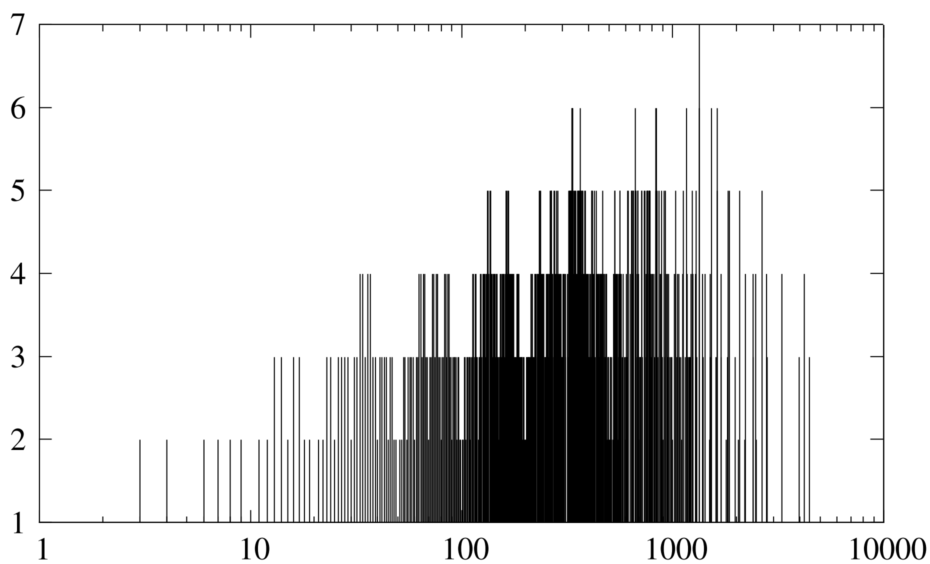

Figure 3.4 – Distribution empirique des transactions en espèces : nombre de coupures échangées en fonction du montant (échelle logarithmique en dixième d'euros)

3.4 Conclusion

Cet article avait pour objectif d'analyser la relation entre la distribution des transactions en espèces et l'efficacité d'un système de divisions monétaires. Plus précisément, nous avons mesuré et comparé l'efficacité du SDM européen obtenue dans le cadre d'une

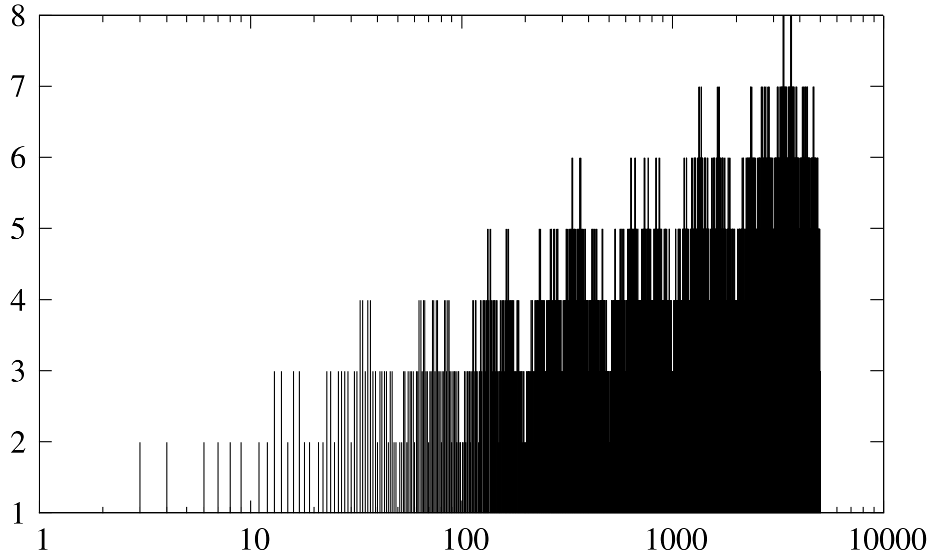

Figure 3.5 – Distribution uniforme des transactions en espèces : nombre de coupures échangées en fonction du montant (échelle logarithmique en dixième d'euros)

distribution empirique des transactions payées en espèces dans l'économie française et d'une distribution uniforme définie sur le même intervalle d'étude.

Les résultats ont mis en évidence une plus grande efficacité du SDM pour la distribution empirique et ont montré comment les différentes divisions étaient impactées. Cet article fournit ainsi des indices utiles afin de mieux comprendre comment et pourquoi la répartition des transactions en espèces peut affecter l'efficacité des paiements. Nous avons indiqué notamment les facteurs liés à la proportion des montants faibles et des montants arrondis dans la distribution empirique en France.

Au final, notre travail questionne les recherches qui concluent à la plus grande efficacité d'un SDM sur la base d'une distribution unique, théorique et exogène des transactions en espèces et appelle ainsi à de nouveaux développements. Notre travail montre en effet que la distribution des montants payés en espèces en France est très contingente à la structure du SDM européen. En particulier, la fréquence d'apparition d'un montant payé en espèces est plus élevée lorsque le nombre de coupures échangées est faible ou lorsque ce dernier correspond à une valeur faciale du SDM utilisé. On peut donc en conclure que les agents économiques endogénéisent la structure du SDM lorsqu'ils fixent les prix ou lorsqu'ils composent différents prix dans leur panier de bien (Lee et *al.*, 2005) et qu'une distribution théorique est mal adaptée à la question de recherche.

Pour contrôler en partie le problème de la contingence des prix et mesurer l'efficacité d'un SDM, il serait donc intéressant dans de futures recherches de comparer deux SDM distincts, européen et américain par exemple, avec deux distributions empiriques des montants payés en espèces provenant de leurs environnements d'usage respectifs. Ce travail constituerait une piste de recherche originale sur l'efficacité comparée des systèmes de divisions monétaires.

Références

Boeschoten W.C. et Fase M.M.G., 1989, "The Way we Pay with Money", *Journal of Business and Economic Statistics*, 7(3) : 319-326.

Boeschoten W. C., 1992, "The Demand for Large Bank Notes", *Journal of Money, Credit and Banking*, 24(3) : 319-337.

Bouhdaoui Y., Bounie D. et Van Hove L., 2011b, "Central Banks and Their Banknote Series : The Efficiency-Cost Trade-Off", *Economic Modelling*, 8(4) : 1482-1488.

Bouhdaoui Y. et Bounie D., 2010, "Optimal Denominations for Coins and Banknotes : Reconciling the Principle of Least Effort with Cost Considerations", Available at SSRN : http ://ssrn.com/abstract=1629192.

Bounie D. et Francois A., 2006, "Cash, Check or Bank Card ? The Effects of Transaction Characteristics on the Use of Payment Instruments", Telecom Paris Economics and Social Sciences Working Paper No. ESS-06-05. Available at SSRN : http ://ssrn.com/abstract=891791.

Bounie D, et Houy N., 2009, “Efficacité des paiements en euros et système de divisions monétaires”, *Revue Française d’Economie*, 24 : 153-167.

Chen C-N., 1976, "Currency Denominations and the Price Level", *Journal of Political Economy*, 84(1) : 179-184.

Chen C-N. et Tsaur T-W, 1983, "Currency denominations, currency substitutions, and the price level", *Journal of Macroeconomics*, 5(4) : 511-513.

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

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

Hentsch J.C., 1975, "Calcul d’un critère qualitatif pour les séries de valeurs définissant l’échelonnement des signes monétaires", *Journal de la Société de Statistique de Paris*, 116(4) : 310-315.

Hentsch J.C., 1983, "Distribution de la monnaie fiduciaire entre les coupures qui la représentent". *Journal de la Société de Statistique de Paris*, 124(4) : 263-272.

Hentsch J.C., 1985, "Distribution de la monnaie fiduciaire entre les coupures. Les paramètres et l'indice des prix. La distribution normalisée." *Journal de la Société de Statistique de Paris*, 126(4) : 139-144.

Klee E., 2008, "How People Pay : Evidence from Grocery Store Data", *Journal of Monetary Economics*, 55 : 526-541.

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 : 1015-1021.

Van Hove L. et Vuchelen J., 1996, "Analyse de la répartition de la monnaie fiduciaire en coupures : la méthode de Hentsch reconsidérée", *Revue Economique*, 47(5) : 1149-1178.

Annexe

Division	Distribution uniforme	Distribution empirique
10c	0,410	0,216
20c	0,515	0,370
50c	0,385	0,304
1€	0,300	0,350
2€	0,486	0,382
5€	0,375	0,297
10€	0,301	0,184
20€	0,485	0,156
50€	0,378	0,049
100€	0,301	0,014
200€	0,552	0,003
500€	0,300	0,000
Total ($\overline{n}$)	4,78	2,33

Tableau 3.I – Fréquence moyenne d'utilisation des divisions pour les distributions uniforme et empirique.

CHAPITRE 4

CENTRAL BANKS AND THEIR BANKNOTE SERIES : THE EFFICIENCY-COST TRADE-OFF

Bouhdaoui Y.[1] , Bounie D.[2] and Van Hove L.[3]

4.1 Introduction

In the theoretical research on optimal denominations for coins and banknotes, efficiency has always been the predominant concern. This is reflected in the popularity of the "principle of least effort", which holds that a denominational structure should make it possible for transactors to economize on the number of tokens exchanged. For the case of exact payment, Caianiello *et al.* [1] demonstrated early on that modular currency systems - systems in which each denomination is X times the one below it (with X an integer) - are the most efficient, and that the number of tokens exchanged is a growing function of the spacing factor. For the case where overpayment and the return of change is allowed, Van Hove and Heyndels [2] and Van Hove [3] eventually, after many a controversy, showed that the optimal spacing factor is *not* three - as claimed by Sumner [4] and Telser [5] - but rather two, as in the case of exact payment.

These theoretical studies supposedly also have important practical implications beyond efficiency. In particular, a reduction in the number of tokens exchanged is assumed not only to be more convenient for transactors, but also to keep down the number of coins and banknotes in circulation, and thus the handling and production costs incurred by the central bank.[4] In short, the most efficient denominational mix would at the same time

[1]Telecom ParisTech, Economics and Social Sciences (ESS), 46 rue Barrault, 75634, Paris Cedex 13 ; tel : +33 1 45 81 82 95 ; e-mail : yassine.bouhdaoui@telecom-paristech.fr

[2]Telecom ParisTech, ESS, 46 rue Barrault, 75634, Paris Cedex 13 ; tel : +33 1 45 81 73 32 ; e-mail : david.bounie@telecom-paristech.fr ; Corresponding author.

[3]ECON - 2C147, Vrije Universiteit Brussel (Free University of Brussels), Pleinlaan 2 ; B-1050 Brussels, Belgium ; tel : +32 2 629 21 25 ; e-mail : Leo.Van.Hove@vub.ac.be.

[4]Many economists support this view ; see Boeschoten and Fase [6], Eriksson and Kokkola [7], Abrams [8], Pedersen and Wagener [9], Van Hove and Heyndels [2] and Van Hove [3].

also be the most cost efficient, at least for the central bank. This matters because the currency expenses of central banks are quite high. In the U.S, for example, the cost for the Federal Reserve System of new currency alone is budgeted at $703 million for 2010, which is equivalent to 16% of the total budget.[5]

The objective of this paper is to show that reducing the spacing factor can actually *increase* the production costs incurred by the central bank. To that end, we build on Cramer's model of efficient payments [10] by incorporating the production costs of the tokens. The resulting model makes explicit how an increase in the "density" of a currency system - that is, an increase in the number of denominations over a given interval - on the one hand improves its efficiency and thus lowers the variable production costs, but on the other hand increases the fixed production costs. Using simulations in which we compare powers-of-two and powers-of-three currency systems, we demonstrate that, under certain conditions, the lower average frequency of use that comes with the powers-of-two system is not sufficient to offset the second effect. In other words, while the powers-of-two system is clearly more efficient, it can be more costly for central banks than a powers-of-three system, and this even with identical cost structures. Obviously, pure powers-of-two or powers-of-three systems - the latter having denominations of 1, 3, 9, 27, 81, etc. - are not really viable in practice.[6] We have therefore also included two real-life currency systems in our simulations, namely the series used by the European Central Bank (ECB) and the Federal Reserve. Again we find that the central bank faces an efficiency-cost trade-off and thus has to weigh the benefits for transactors against those for the bank itself.

The paper contributes to the literature on optimal denominations for coins and banknotes on two crucial points. For one, it is the first to fully consider all production costs. Indeed, while Bouhdaoui and Bounie [13] also show that the principle of least effort fails to minimize the production costs of the central bank and that it should thus be reconciled

[5]Source : Board of Governors of the Federal Reserve System, Annual Report : Budget Review, May 2010, p. 9 < *http : //www.federalreserve.gov/boarddocs/rptcongress/budgetrev/br10.pdf* >.

[6]Tschoegl [11 : 546] notes that today no currency follows the powers-of-three principle, and that even in the past only few examples of the use of the principle can be found. Wynne [12 : 222] finds that only 5 countries have denominations that are either powers or integer multiples of three.

with cost considerations, they use a basic cost function of the central bank that ignores the fixed costs of production. As a result, they do not take into account economies of scale, which are critical in understanding the design of a denominational mix. Second, the present paper looks into the issue of what constitutes a (theoretically) optimal denominational mix. Bouhdaoui and Bounie [13], for their part, only compare the cost of efficient and inefficient payments for a single real-life currency system, namely the U.S. series as it exists today.

The remainder of the paper is structured as follows. In Section 2, we present an original framework that models the currency production costs of the central bank and in particular indicates how these are affected by the frequency of use of the tokens in circulation. In Section 3, we use this framework to study - still on a purely theoretical level - the conditions under which one currency system can be more costly than another. In Section 4, we then illustrate our analysis by performing simulations for selected currency systems. Section 5 discusses the implications of our results.

4.2 General framework

In this section, we introduce our general framework - in two steps. We first present a currency production cost function and discuss the different types of costs involved (2.1). Subsequently, we model how these production costs are affected by parameters such as the number of cash transactions in the economy, the quality standards of the central bank, and especially the efficiency of the currency system (2.2).

4.2.1 The production costs of the central bank

In a given economy, let us consider a central bank that has a currency system V composed of J denominations of face values $v(j)$, with $j \in [1,J]$. Let us assume that the central bank mints its own coins and prints its own banknotes. If we denote by $N_i(j)$ the volume of new tokens of a specific denomination that are issued by central bank in a period of, say, one year, and if we assume uniform amortization, then the total annual production cost C incurred by the central bank for these tokens can be written as :[7]

$$C = C_0^F + \sum_{j=1}^{j=J} \left(C_p^F(j) + c_p^V(j) \cdot N_i(j) \right). \tag{4.1}$$

In this equation, C_0^F stands for the "overall" fixed costs, $C_p^F(j)$ for the fixed costs on the level of an individual denomination, and $c_p^V(j)$ for the unit variable cost of a denomination. Note that the distinction between fixed and variable costs - and especially the presence of $C_p^F(j)$ - will prove crucial for our analysis of the cost efficiency of currency systems in Section 3. Below we discuss and illustrate the different types of costs in more detail. For reasons of brevity, we do this for the case of banknotes, and not for coins. Illustrations are mainly based on data from the U.S. Bureau of Engraving and Printing (BEP) - the government agency responsible for the development and production of U.S. currency notes - and the Federal Reserve, the central bank that is most transparent in this respect.

As is clear from Equation (4.1), overall or "high-level" fixed costs are fixed costs that the central bank incurs regardless of the number of denominations in its currency system. Infrastructure and indirect personnel costs come to mind. In its annual "New Currency Budget" (NCB), the Federal Reserve does not split up its printing costs into fixed vs. variable costs, but it does provide a split-up into "capital investment", "public education", "production support", and "currency production". Of these four cost components, the first two are clearly fixed. The percentage contribution of these and other factors varies from one year to another - depending for example, on the number of notes ordered by the Fed - but in 2009 capital investment represented approximately 6 percent of total currency printing costs and public education 2 percent.[8] The item "production support" (46 percent in 2009) is of a mixed nature : it comprises such fixed outlays as expenses for IT security and business continuity, but costs for "steam, natural gas, water and electricity" are influenced by production volume. Dan Peterson of the BEP estimates that the item is

[7]For simplicity, we ignore the fact that in certain years the central bank might produce more tokens of a certain denomination than it really needs (in order to build a buffer stock), while in other years it will draw on this stock.

[8]Sources : 2009 NCB.

"probably 70% fixed and 30% variable".[9] Finally, "currency production" - 46 percent of the total in 2009 - is mainly about variable costs (paper, ink, the cost of note packaging, etc.), but also includes depreciation for the printing/processing equipment.[10] Separately from the NCB, the BEP kindly provided us with detailed data on the manufacturing cost of U.S. banknotes, and this per denomination and per production facility (the BEP has two separate production facilities).[11] We have used the BEP data for September 2009 to calculate a weighted average of the share of depreciation in the manufacturing cost and arrived at 4.8 percent. In total, as a rough estimate, fixed costs would thus have accounted for 42.4 percent in 2009 (= 6% + 2% + 70%*46% + 4.8%*46%).[12]

As mentioned, Equation (4.1) not only distinguishes fixed from variable costs, but also makes a distinction between overall fixed costs and fixed costs per denomination. Where the latter are concerned, there is for instance, an "origination cost" attached to the creation of a new denomination. This relates to "the high-quality image preparation and the transformation of designs into production tools, such as printing plates and production forms for the different manufacturing steps" [14, p. 38].[13] Expenses on research and development - for example concerning security features - may constitute another type of fixed cost per denomination. In the U.S., the BEP for example "incurred significant fixed research and development costs in 2009 associated with testing the new-design $100 notes".[14] However, sometimes such R&D costs are not incurred for a single denomination but rather intended for and therefore amortized over multiple denominations - either the entire series or at least a "family" of denominations (say the higher-value notes). The data that we received from the BEP go some way to underpin the existence of fixed

[9]Source : Peterson, D., private e-mail, January 24, 2011 (on file with the authors).

[10]Source : Peterson, D., private e-mail, January 24, 2011 (on file with the authors).

[11]Source : BEP. (2009). Financial Report for Monthly Production and Cost Meeting - Month of September 2009.

[12]Note that this should be seen as a lower limit. Indeed, as Dan Peterson pointed out to us, it could be argued "that the direct labor [which is included in "currency production"] in the short term is fixed. The BEP would be reluctant to furlough its skilled workforce" (Source : Peterson, D., private e-mail, January 24, 2011 (on file with the authors)). It proved impossible to determine the relative importance of direct labor with the data that we have.

[13]The issuance of a new denomination may also require new public education materials, so that it could be argued that this item too is - in part - a per-denomination fixed cost.

[14]Source : 2010 NCB.

costs on the level of individual denominations. Indeed, in the BEP data for September 2009, the total manufacturing cost of a denomination per 1,000 notes contains an item called "depreciation", and for the BEP's Western Currency Facility this item varies between \$1.87 for the new-design \$100 note (which corresponds to 5% of its manufacturing cost) to \$4.94 for the next-generation \$20 note (or 10%).[15] As the manufacturing cost is expressed per 1,000 notes, part of this variation can be explained by the existence of fixed costs that are specific to the denomination in question.[16] However, depreciation of printing presses is often spread over more than one denomination.

Also, once a denomination is in production, the fixed costs are again *partly* situated on the level of individual denominations (as opposed to overall). Even though, as Dan Peterson of the BEP points out, "overall program volume is the primary cost driver, not the volume associated with each denomination"[17], for a central bank it does make a difference whether it prints, say, 2 million units of a single denomination or the same volume of two different denominations. For one, while the same printing machines may be used across denominations (with some denominations requiring additional machines because of additional security features), the machines do have to be halted and adjusted when switching to another denomination. Also, and related to this, while the chemical composition of the paper will be comparable, if not identical, the watermark will typically be different and so may the ink, the security features, etc.

Returning to our model, Equation (4.1) can be used to highlight the economies of scale on the different levels. The total production cost per denomination is :

$$C(j) = \frac{N_i(j)}{N_i} C_0^F + C_p^F(j) + c_p^V(j) \cdot N_i(j), \tag{4.2}$$

where $N_i = \sum_j N_i(j)$ is the total issuing volume and $\frac{N_i(j)}{N_i}$ the issuing ratio of denomina-

[15]Source : BEP. (2009). *Financial Report for Monthly Production and Cost Meeting - Month of September 2009.*

[16]Conversely, for the Eastern Currency Facility (ECF), at \$1.34 per 1,000 notes, the depreciation item is identical for all next-generation notes - that is, for the denominations of \$5, \$10, \$20, and \$50. This points towards a joint, multi-denomination fixed cost. For the \$1 note, depreciation runs at \$1.05 - again for September 2009 and the ECF.

[17]Source : Peterson, D., private e-mail, December 17, 2010 (on file with the authors).

tion j. Thus, the unit production cost of a denomination j is :

$$c_p(j) = \frac{C(j)}{N_i(j)} = \frac{C_0^F}{N_i} + \frac{C_p^F(j)}{N_i(j)} + c_p^V(j). \tag{4.3}$$

In Equation (4.3), the term $\frac{C_0^F}{N_i}$ highlights why, in case of the U.S., “the individual 2010 billing rates per thousand notes of comparable design are less than the 2009 billing rates per thousand notes. The individual 2010 billing rates are less, because *the fiscal year 2010 print order contains more notes across which to spread high [overall] fixed costs of production*”.[18] This said, there are also per-denomination economies of scale - which are captured by the term $\frac{C_p^F(j)}{N_i(j)}$ in Equation (4.3). In this respect, it is interesting - albeit not conclusive - to observe that, for denominations that are printed at both of the BEP’s production facilities, the level of the depreciation item discussed above differs between the WCF and the ECF, and that there is a negative relationship with the production volume of the denomination in the respective facility. For the case of the ECB too, there is some (anecdotal) evidence. Concerning the printing of the initial stock of euro banknotes, the ECB [14, p. 49] reports that “the national central banks could make bilateral pooling agreements with each other. This was especially useful for producing low volumes of the high denominations, *i.e.* €200 and €500. For some smaller countries in particular, it was more efficient not to produce these denominations but rather to order a proportion of a bigger country’s production volume and benefit from the economies of scale”.[19]

The term $\frac{C_p^F(j)}{N_i(j)}$ also captures the fact that adding a denomination might generate new, specific cost items - for instance, specific Optically Variable Ink. Also, as the new denomination will lower the demand for certain of the other denominations, the manufacturing cost of the latter might go up, at least where denomination-specific items are

[18] Source : NCB 2010, footnote 10 (emphasis added).

[19] Note that the economies of scale referred to in the quote are in all probability a mix of overall and per-denomination scale economies. Interestingly, after the change-over to the euro, the ECB decided to pool the printing of denominations, so that today the national central banks are each specialized in the printing of specific denominations. The 100, 200, and 500 euro denominations, for instance, are printed by a single central bank - that of Italy, Germany, and Austria, respectively (European Central Bank : http ://www.ecb.int/stats/euro/production/html/index.en.html (last visit 01/09/2010)).

concerned. Lower volumes of special ink, for example, might result in suppliers charging higher unit prices. The case of the U.S. shows that denomination-specific features can be manifold. A salient example is the fact that only the $100 notes include green taggant ink.

4.2.2 Currency usage and production costs

Now that we have our currency production cost function, we need to find a way to make explicit that the central bank's production costs are affected by the frequency of use of the tokens in circulation, and ultimately by the efficiency of its currency system.

We establish these links by showing, first, that the volume of new tokens of a specific denomination that are issued by the central bank, $N_i(j)$, depends, among other factors, on the frequency of use of that denomination, $a(j)$. To that end, let us start by pointing out that the average life span, $d(j)$, and the circulating volume of a denomination j, $N_c(j)$, together define the annual issuing volume, $N_i(j)$. Indeed, central banks monitor the quality of the tokens that are returned, and substandard specimens are withdrawn from circulation and replaced by new ones. The volume of new tokens that are released in this way each year is given by :

$$N_i(j) = \frac{N_c(j)}{d(j)}, \tag{4.4}$$

with $d(j)$ expressed in years.

The velocity of circulation, $q(j)$, of denomination j is the average number of times a token j is involved in cash transactions during a year. We express $q(j)$ as :

$$q(j) = \frac{N_T \cdot a(j)}{N_c(j)}, \tag{4.5}$$

In this Equation, $a(j)$ is the average frequency of use per transaction of denomination j and N_T the number of cash transactions observed in the economy during a year.[20]

[20]Note that we assume the number of cash transactions N_T to be given.

Hence, the numerator represents the number of times denomination j is involved in cash transactions during a year.

Rearranging the terms in (4.4), then replacing in (6.6), we have :

$$q(j) = \frac{N_T \cdot a(j)}{N_i(j) \cdot d(j)}. \tag{4.6}$$

Thus, the annual issuing volume is :

$$N_i(j) = \frac{N_T \cdot a(j)}{q(j) \cdot d(j)}. \tag{4.7}$$

We assume that the quality of a circulating token decreases after each use in transactions. Therefore, we assimilate the quality requirement of the central bank for a denomination j to a maximum number of uses $u_{max}(j)$ before the token is replaced. In addition, we assume that all denominations are made of the same manufacturing material and that the central bank imposes the same maximum number of uses for all denominations, denoted by u_{max} :

$$u_{max} = q(j) \cdot d(j). \tag{4.8}$$

Hence, replacing (4.8) in (4.7) we obtain :

$$N_i(j) = \frac{N_T \cdot a(j)}{u_{max}}. \tag{4.9}$$

If we replace this in the cost function introduced in Section 4.2.1, we obtain :

$$C = C_0^F + \sum_{j=1}^{j=J} \left(C_p^F(j) + c_p^V(j) \cdot \frac{N_T \cdot a(j)}{u_{max}} \right). \tag{4.10}$$

To simplify the latter equation, we denote by C_p^F the average fixed cost per denomination and by c_p^V the weighted average unit variable cost of the currency system. Thus, rearranging the terms of (4.10), we have :

$$C = C_0^F + J \cdot C_p^F + c_p^V \cdot \frac{N_T}{u_{max}} \cdot \sum_{j=1}^{j=J} a(j). \tag{4.11}$$

On the right-hand side of (4.11), the sum $\sum_j a(j)$ is obviously nothing else than the average number of tokens exchanged per transaction, denoted by $\bar{n}$:

$$\bar{n} = \sum_{j=1}^{j=J} a(j). \tag{4.12}$$

Hence we obtain :

$$C = C_0^F + J \cdot C_p^F + c_p^V \cdot \frac{N_T}{u_{max}} \cdot \bar{n}. \tag{4.13}$$

We can now use Equation (4.13) to compare the potential cost gaps between currency systems.

4.3 Comparing two currency systems

Let us consider two currency systems, $V^{(1)}$ and $V^{(2)}$, composed of $J^{(1)}$ and $J^{(2)}$ denominations of face value $v(j)$ with, respectively, $j \in [1, J^{(1)}]$ and $j \in [1, J^{(2)}]$. For all other variables, we build on the notations introduced in Section 4.2 and simply add superscripts $^{(1)}$ and $^{(2)}$. In order not to detract from our main point, we assume that denominations of both systems are made from the same material and that costs depend only on the quality of the manufacturing material. This implies that $V^{(1)}$ and $V^{(2)}$ have the same highest-level fixed costs, the same average per-denomination fixed costs, and the same weighted-average unit variable costs.

We now measure the gap in annual production costs between $V^{(1)}$ and $V^{(2)}$ as $\Delta C = C^{(2)} - C^{(1)}$. From Equation (4.13) we have :

$$\Delta C = \underbrace{\Delta J \cdot C_p^F}_{I} + \underbrace{c_p^V \cdot \frac{N_T}{u_{max}} \cdot \Delta \bar{n}}_{II}. \tag{4.14}$$

In Equation (4.14), the first term (I) indicates that the higher the number of denominations in a currency system, the higher the fixed production costs at the denomination level. The second term (II) compares the variable production costs and shows that the higher the average number of tokens per transaction, $\bar{n}$, the higher the variable production costs.

Crucially, the number of denominations and the efficiency of a currency system are not independent of one another. Suppose that $V^{(2)}$ has more denominations than $V^{(1)}$; that is, it is denser. In such a case, $V^{(2)}$ will typically allow transactors to economize on the number of tokens exchanged. Where modular currency systems are concerned, there is no doubt about this, at least not if the distribution of cash transactions is uniform. Indeed, in that case, we know from Van Hove and Heyndels [2] that the lower the spacing factor, the lower the average number of tokens exchanged in payments. For non-modular systems, this is less of a general rule, as the efficiency of a system not only depends on the number of denominations but also on their precise positions (vis-à-vis the amounts to be paid).[21] If $V^{(2)}$ effectively has a lower $\bar{n}$, it will have lower variable production costs. But, at the same time, since $V^{(2)}$ is denser, the per-denomination fixed production costs are greater than those of $V^{(1)}$.

Overall, system $V^{(2)}$ is less costly for the central bank when $\Delta C < 0$; it is more costly otherwise. Table 4.I provides an overview of the different possible cases. In case one, it is straightforward to see that $V^{(2)}$ costs more to produce than $V^{(1)}$. Since both the number of denominations, J, and the average number of tokens exchanged, $\bar{n}$, are higher for $V^{(2)}$, the same is true for both the fixed and variable production costs. Similarly, in case four, $V^{(2)}$ is definitely beneficial for the central bank because both conditions are reversed. In cases two and three, $V^{(2)}$ *can* benefit the central bank, but only if the stated conditions are satisfied.

Note that, in view of the relationship between J and $\bar{n}$ pointed out above, for modular systems (and uniform price distributions) cases one and four are actually not possible. In other words, in such a setting, a choice between two modular systems always implies a

[21] Modular currency systems by definition have a constant spacing factor. The positions of the denominations thus obey a certain rule, and it is precisely this which makes it possible to derive a formula for the average number of tokens exchanged - as it has been done by Caianiello *et al.* [1] and Van Hove and Heyndels [2].

	$\Delta J > 0$	$\Delta J < 0$
$\Delta\bar{n} > 0$	Case 1 : $\Delta C > 0$	Case 2 : $\Delta C < 0$ if and only if $\frac{c_p^V}{C_p^F} < -\frac{u_{max}}{N_T} \cdot \frac{\Delta J}{\Delta \bar{n}}$
$\Delta\bar{n} < 0$	Case 3 : $\Delta C < 0$ if and only if $\frac{c_p^V}{C_p^F} > -\frac{u_{max}}{N_T} \cdot \frac{\Delta J}{\Delta \bar{n}}$	Case 4 : $\Delta C < 0$

Tableau 4.I – Comparing the production costs between two currency systems.

trade-off between J and $\bar{n}$, and thus between fixed and variable costs. As explained, for non-modular systems it is more difficult to predict in which quadrant the central bank will end up, and cases one and four are real possibilities.[22] However, as we show in the next Section, the trade-off that is at the center of the present paper will also turn up often for non-modular systems.

4.4 An application to modular and real-life currency systems

In this section, we primarily compare the two currency systems that are most widely studied in the theoretical literature, namely the powers-of-two and the powers-of-three systems. The objective is to show that while a powers-of-two system uses fewer coins and notes in transactions, it can, under certain circumstances, be more costly for the central bank. However, in order to demonstrate that this assertion has broader validity, we have also included two non-modular real-life currency systems, namely those used, respectively, by the ECB and the Federal Reserve. The first is composed of the binary decimal triplets 1-2-5, 10-20-50, etc. ; the second has the same general structure but has a number of "gaps" (and also has a coin of 25 instead of 20 cents).

[22]This holds *a fortiori* when, on top of this, the assumption of a uniform price distribution is also abandoned.

4.4.1 Methodology and assumptions

To repeat, an X-modular currency system is a system in which each denomination is X times the one below it, X being a constant and called the module of the currency system. We denote the 2-modular (powers-of-two) and 3-modular (powers-of-three) currency systems by $V^{X=2}$ and $V^{X=3}$, respectively.

In practice, the size of the smallest cash transaction is obviously determined by the face value of the lowest coin. In our simulations, we express all cash transactions in cents and initially set the uniform distribution of transactions on which the comparison is performed at $I = [1,100]$, implying that $N_T = 100$. On this interval, we obtain the following modular currency systems : $V^{X=2} = \{1,2,4,8,16,32,64\}$ and $V^{X=3} = \{1,3,9,27,81\}$; hence, $J^{X=2} = 7$ and $J^{X=3} = 5$. Over the same interval, the currency system of the ECB is $V^{ECB} = \{1,2,5,10,20,50,100\}$ and that of the Federal Reserve is $V^{Fed} = \{1,5,10,25,100\}$, so that J^{ECB} equals 7 and J^{Fed} equals 5.[23]

In accordance with the standard assumptions in the literature, we assume that, when making cash payments, the public uses the principle of least effort, including overpayment and the return of change.[24] In other words, consumers and merchants use as few coins and notes as possible in transactions.[25]

4.4.2 Simulations and results

The objective now is to simulate the use of the principle of least effort on the interval I with different currency systems, and this in order to assess the average number of tokens exchanged per transaction, $\bar{n}$. Table 4.II summarizes the results of the simulations. If we first concentrate on the two modular currency systems, it can be seen that, as expected, $\bar{n}$ is lower for the powers-of-two than for the powers-of-three system. This is in line with the standard result of the literature, which states that lowering the spacing factor

[23]Following Telser [5], we omit the 50 cent piece, which does not circulate widely. In Section 3.3, we will also omit the $2 bill - for the same reason.

[24]See Cramer [10] for a formal description of the principle of least effort.

[25]Franses and Kippers [15] have shown, for the case of the Netherlands, that the observed cash payment behavior of the public approaches the principle of least effort.

improves the efficiency of the currency system [3].

$V^{X=2}$	$V^{X=3}$	V^{ECB}	V^{Fed}
a(1) = 0.50	a(1) = 0.80	a(1) = 0.40	a(1) = 1.25
a(2) = 0.32	a(3) = 0.66	a(2) = 0.52	a(5) = 0.37
a(4) = 0.34	a(9) = 0.70	a(5) = 0.38	a(10) = 0.54
a(8) = 0.32	a(27) = 0.59	a(10) = 0.30	a(25) = 0.85
a(16) = 0.33	a(81) = 0.47	a(20) = 0.48	a(100) = 0.38
a(32) = 0.37	-	a(50) = 0.40	-
a(64) = 0.53	-	a(100) = 0.26	-
$\bar{n}^{X=2}$ = 2.70	$\bar{n}^{X=3}$ = 3.22	$\bar{n}^{ECB}$ = 2.73	$\bar{n}^{Fed}$ = 3.38

Tableau 4.II – Average frequencies of use under different currency systems.

But what are the production costs incurred by the central bank ? Note that since $\Delta J = -2 < 0$ and $\Delta\bar{n} = 0.52 > 0$, we are in case two of Table 4.I. This means that the powers-of-three system is less costly for the central bank than the powers-of-two system if the ratio $\frac{c_p^V}{C_p^F}$ is below the threshold $\tau = -\frac{u_{max}}{N_T} \cdot \frac{\Delta J}{\Delta\bar{n}}$. Assuming that u_{max} equals unity (which implies that each token is used only once during its lifetime), the difference in annual production costs spelled out in Equation (4.14) is given by :

$$\Delta C = C^{X=3} - C^{X=2} = -2 \cdot C_p^F + 52 \cdot c_p^V, \tag{4.15}$$

and the 3-modular system is less costly than the 2-modular when :

$$\frac{c_p^V}{C_p^F} < 0.0385(3.85\%). \tag{4.16}$$

If this condition is fulfilled, then the weighted average unit variable cost c_p^V is sufficiently low to offset the higher volume of tokens required under $V^{X=3}$ and/or the average fixed cost per denomination C_p^F is sufficiently high to let the lower number of denominations of $V^{X=3}$ play to the fullest. The simulation results thus confirm that increasing the spacing factor - or, put differently, lowering the density/efficiency of a currency system

- does not systematically increase the production costs incurred by the central bank.

Turning to the real-life currency systems in Table 4.II, it can be seen that the choice between the currency systems of the ECB and the Fed is in fact not fundamentally different from the choice that we have just discussed. Indeed, since $\Delta J = -2 < 0$ and $\Delta\bar{n} = 0.65 > 0$, we are again in case two of Table 4.I, and the same considerations apply. As an aside, it can be noted that the ECB series is almost as efficient as a pure powers-of-two system - with $\bar{n}$ equal to 2.73 and 2.70, respectively. This should not come as a surprise since the average multiple of a series composed of complete binary-decimal triplets equals 2.2, and is thus much closer to 2 than 3. In addition, over the interval [1, 100], the ECB series and the powers-of-two system have the same number of denominations. Given this, Equation (4.14) indicates that the difference in production costs between the two systems will only come from the variable costs, which will be slightly higher for the ECB series as $\bar{n}$ is somewhat higher.

4.4.3 Sensitivity analysis and discussion

To strengthen our results, we now perform additional simulations so as to analyze the influence of the number of transactions N_T on the relative cost efficiency of currency systems. Concretely, we start with the interval studied in the previous Section - namely I = [1,100] - and gradually increase N_T all the way to 10,000. The results of the simulations are summarized in Figures 4.1, 4.2 and 4.3.

Again we focus first on the two modular systems. Figure 4.1 confirms that, although the difference in the number of denominations, ΔJ, fluctuates somewhat depending on the interval considered, $V^{X=2}$ is, by definition, always denser than $V^{X=3}$. In line with Van Hove and Heyndels [2], Figure 4.2 confirms the superiority of $V^{X=2}$ in terms of (theoretical) efficiency.[26] Bringing the insights of Figures 4.1 and 4.2 together, it is clear that across all intervals we remain in case two of Table 4.I, as $\Delta J < 0$ and $\Delta\bar{n} > 0$. Finally, and most importantly, Figure 4.3 shows that for $V^{X=3}$ to maintain its cost advantage

[26] "Theoretical" in the sense that we work with uniform price distributions, whereas in reality the distribution of cash transactions is skewed.

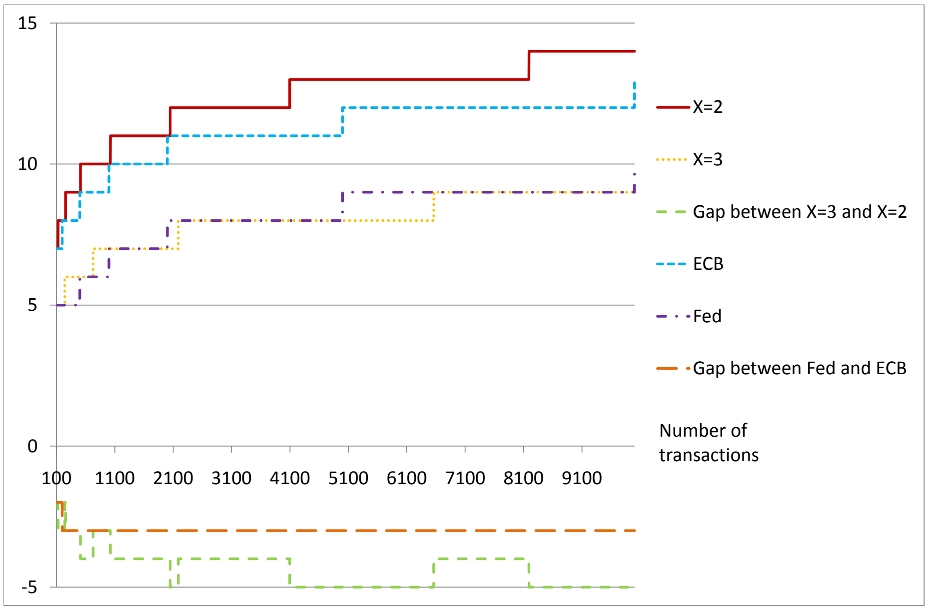

Figure 4.1 – Number of denominations in interval, J : comparison of selected currency systems.

over $V^{X=2}$, the ratio $\frac{c_p^V}{C_p^F}$ must be low when the number of transactions N_T is high. The rationale behind this result can be explained as follows. From Equation (4.9) we know that the total issuing volume, across denominations, of a given currency system is given by :

$$\sum_j N_i(j) = \frac{N_T}{u_{max}} \cdot \bar{n}.$$

That is, if we assume that the central bank's quality requirements, which we capture by u_{max}, are not influenced by a change in the number of transactions, then the total issuing volume depends on N_T and on the system's efficiency. Building on this, the total variable production costs, C_T^V, of a currency system can be written as :

$$C_T^V = c_p^V \cdot \sum_j N_i(j) = c_p^V \cdot \frac{N_T}{u_{max}} \cdot \bar{n}. \tag{4.17}$$

That is, C_T^V depends on N_T, $\bar{n}$, and c_p^V. Taking this one step further, the difference in total variable production costs between $V^{X=3}$ and $V^{X=2}$ is given by :

$$\Delta C_T^V = c_p^V \cdot \sum_j \Delta N_i(j) = c_p^V \cdot \frac{N_T}{u_{max}} \cdot (\bar{n}^{X=3} - \bar{n}^{X=2}). \tag{4.18}$$

The above equation shows that the variable cost gap increases proportionally to N_T, even if $\Delta\bar{n}$ were invariant. However, the latter is not invariant : as N_T increases, $\Delta\bar{n}$ actually widens (to the advantage of $V^{X=2}$; see Figure 4.2), thus increasing the cost gap even further. Hence, if $V^{X=3}$ is to maintain its cost advantage as N_T increases, the unit variable cost c_p^V must be low to compensate for the increase in the issuing volume just explained ; see Figure 4.3. At the same time, this effect is tempered somewhat by the fact that as N_T increases, ΔJ widens (to the advantage of $V^{X=3}$; see Figure 4.1), which impacts ΔC_p^F positively.[27]

Turning to our two non-modular systems, the observations proffered in Section 5.3.2 remain largely valid. While the difference in efficiency between the ECB series and a powers-of-two system initially increases somewhat when the interval is widened,[28] it eventually stabilizes, and it is obvious that the ECB series substantially outperforms both a powers-of-three system and the currency system of the Federal Reserve (which, over the full interval, happens to have an average spacing factor of exactly 3).[29] As for the number of denominations, the gap between the ECB and Fed series quickly stabilizes at 3, with the first being the most dense.[30] Again Equation (4.14) can be used to demonstrate the trade-off between fixed costs (the ECB series has a higher J) and variable costs (the ECB series is more efficient).

[27] C_0^F is, by definition, not affected by an increase in N_T.

[28] It increases, though not monotonically, from -1.11% over the interval [1,100] to -4.3% over [1,10000].

[29] Over the interval [1,10000] the gap in efficiency amounts to 16.3% and 16%, respectively.

[30] Since the upper limit of our interval is 100 dollars/euro, the banknotes of €200 and €500 do not enter into play.

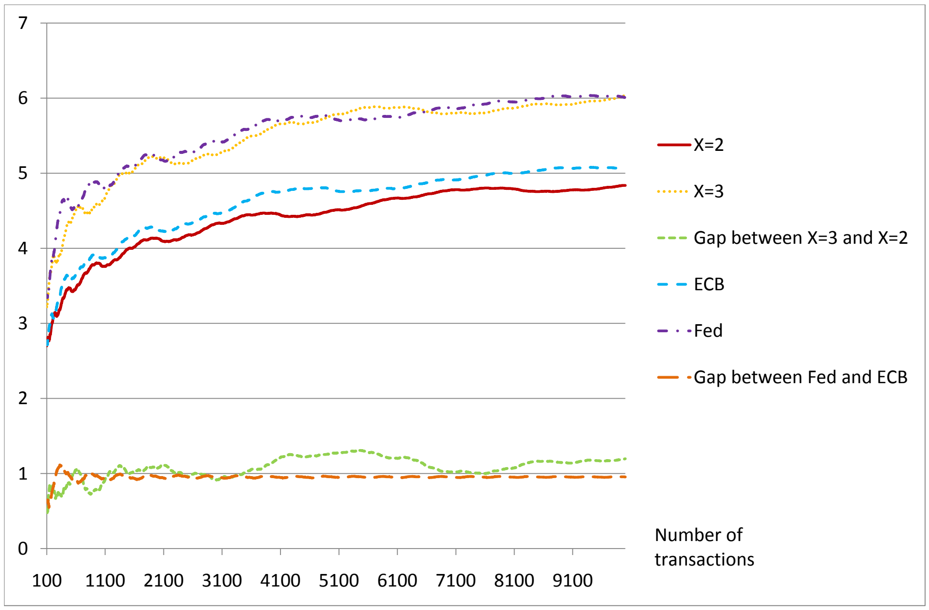

Figure 4.2 – Average number of tokens per transaction as a function of N_T.

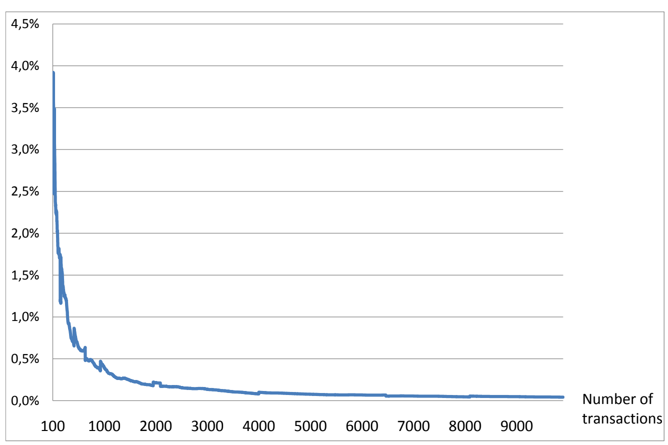

Figure 4.3 – Profitability of $V^{X=3}$ vs. $V^{X=2}$: threshold for $\frac{c_p^V}{C_p^F}$ as a function of N_T, with $u_{max} = 1$.

4.5 Conclusion

The main objective of this paper was to challenge one of the most important results of the theoretical research on currency systems. This result holds that, for modular currency systems, lowering the spacing factor between denominations not only improves the efficiency of the system but also reduces the production costs of the central bank. In order to debunk this finding, we compare the two most popular hypothetical currency systems, namely the powers-of-two and powers-of-three systems. We first develop a general model of the currency production costs incurred by the central bank and, using numerical simulations, we then show that while a powers-of-two system has a lower average number of tokens exchanged in transactions (and is thus indeed more efficient), it can be more costly for a central bank than a powers-of-three system. The intuition behind this result is that, for modular currency systems, a lower spacing factor goes hand in hand with a higher number of denominations, which tends to increase the production costs because of the origination costs attached to additional denominations and because the economies of scale are smaller. We also show that the validity of this result is not limited to hypothetical currency systems. Indeed, the choice between an ECB - or Fed - style series proves to be governed by the same efficiency-cost trade-off.

This finding raises the question how central banks across the world have responded to the trade-off. Interestingly, two earlier studies have computed average spacing factors across a large number of countries. Wynne [13] analyzes data on 156 countries and finds that the arithmetic mean of the spacing factors is ... exactly 3. For the subset of OECD countries, the mean equals 2.8. Tschoegl [12] performs a similar exercise for 50 countries, and comes up with slightly lower values, namely 2.60 for coins and 2.62 for notes. Still, both studies find that the mean average spacing factor is closer to 3 than to 2.

Prima facie, this would seem to indicate that central banks opt not to maximize user convenience at all costs (pun intended), and do economize on the number of denominations. However, several qualifications are in order here. First, to ease mental arithmetic, currency systems have to be compatible with the decimal system. As a result, even in the case of central banks that opt for complete binary-decimal triplets, such as the ECB, the average spacing factor will be higher than 2 (in casu 2.2). Second, Wynne in his paper finds that the distribution of the average multiples is not concentrated around the mean, but rather appears to be bi-modal, with peaks at 2.2 and 2.7. This would seem to indicate that some central banks handle the trade-off differently than others. Third, the average spacing factor says nothing about the number of denominations[31] nor about where the "gaps" in the denominational structure are, if any. Fourth, and related to this, in deciding on their denominational structure, central banks also take into account other criteria besides the principle of least effort. There is, for example, the issue of surveyability : the larger the variety of denominations, the harder it becomes for the public to recognize the different coins and notes, to sort and store them, etc. Also, in reality the distribution of cash payments is not uniform ; small payments occur more frequently. As a result, a low density in the upper regions of a denominational structure - where the store-of-value function of currency overwhelms its function as a medium of exchange - need not greatly affect the public's ability to handle cash payments efficiently, even though it raises the average multiple.

In short, as acknowledged by the existing literature, multiple criteria play a role in the determination of the optimal number and spacing of currency denominations. Our contribution highlights that it is not only a multi-criteria, but also a multi-actor problem. We show that central banks face an efficiency-cost trade-off and thus have to weigh the benefits for transactors against those for the central bank itself.[32] Extending our approach so that it can incorporate additional stakeholders and their activities - such as commercial banks and their ATM operations - would seem a fruitful avenue for further research.

[31] Real-life currency systems differ substantially in this respect : while the number of denominations in the U.S. and Japan amounts to 10, it equals 12 in Sweden and the UK, 15 in the eurozone, and no less than 18 in Singapore.

[32] This is not to say that the interests of central banks and the public cannot partly coincide. As pointed out, the public may also wish to conserve on denominations.

4.6 Literature cited

[1] Caianiello E.R., Scarpetta G. and Simoncelli G. (1982). A Systemic Study of Monetary Systems. *International Journal of General Systems*, 8, 81-92.

[2] Van Hove L. and Heyndels B. (1996). On the Optimal Spacing of Currency Denominations. *European Journal of Operational Research*, 90, 547-552.

[3] 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.

[4] Sumner S. (1993). Privatizing the Mint. *Journal of Money, Credit, and Banking*, 25, 13-29.

[5] Telser L.G. (1995). Optimal Denominations for Coins and Currency. *Economics Letters*, 49, 425-427.

[6] Boeschoten W.C. and Fase M.M.G. (1989). The Way We Pay with Money. *Journal of Business and Economic Statistics*, 7, 319-26.

[7] Eriksson S. and Kokkola T. (1993). *Cash*. In Payment and Settlement Systems in Finland, Publications of the Bank of Finland, Series A, No. 88, 21-57.

[8] Abrams R.K. (1995). *The Design and Printing of Bank Notes : Considerations When Introducing a New Currency*. Working Paper 95-26, International Monetary Fund, March.

[9] Pedersen E.H. and Wagener T. (1996). *Circulation of Notes and Coins in Denmark*. Danmarks Nationalbank Monetary Review, 35, 29-45.

[10] Cramer J.S. (1983). Currency by Denomination. *Economics Letters*, 12, 299-303.

[11] Tschoegl A.E. (1997). The Optimal Denomination of Currency. *Journal of Money, Credit, and Banking*, 9(4), 546-554.

[12] Wynne M.A. (1997). More on Optimal Denominations for Coins and Banknotes. *Economics Letters*, 55, 221-225.

[13] Bouhdaoui Y. and Bounie D. (2010). *Optimal Denominations for Coins and Banknotes : Reconciling the Principle of Least Effort with Cost Considerations*. Telecom ParisTech Working Paper
(http ://ssrn.com/abstract=1629192).

[14] European Central Bank. (2007). How the Euro Became Our Money - A Short History of the Euro Banknotes and Coins. (http ://www.ecb.int/pub/
pdf/other/euro_became_our_moneyen.pdf)

[15] Franses P.H. and Kippers J. (2007). An Empirical Analysis of Euro Cash Payments. *European Economic Review*, 51, 1985-1997.

CHAPITRE 5

PRODUCTION COSTS, SEIGNIORAGE AND COUNTERFEITING : CENTRAL BANKS' INCENTIVES FOR IMPROVING THEIR BANKNOTE TECHNOLOGY

Bouhdaoui Y.[1] , Bounie D.[2] and Van Hove L.[3]

5.1 Introduction

Despite the rapid uptake of alternative electronic payment instruments such as the debit card (Kiser *et al.*, 2008), recent empirical studies confirm that in most countries cash is still today the dominant form of payment in retail transactions. At the same time, it is also the most costly payment instrument for society (Garcia Swartz *et al.*, 2006 ; Guibourg and Segendorff, 2007). In order to reduce the cost of cash payments, central banks have adopted two main strategies, both geared toward lengthening the life span of the denominations in circulation and, in this way, lowering production costs.[4]

The first strategy consists in replacing low-value banknotes by coins - the rationale being that while coins may be more costly to produce, they last substantially longer. The central banks of France, the United Kingdom and Canada - to name but a few - have in the 70s and 80s all introduced coins to replace existing notes. Likewise, in the U.S., the Federal Reserve in 2000 launched a one dollar coin named the "golden dollar".[5] Unfortunately, for business and psychological reasons anticipated in a pre-launch study (Wynne, 1997), the "Sacagawea" is little used in retail transactions.

[1]Telecom ParisTech, Economics and Social Sciences, 46 rue Barrault, 75634, Paris Cedex 13 ; tel : +33 1 45 81 82 95 ; e-mail : yassine.bouhdaoui@telecom-paristech.fr

[2]Telecom ParisTech, Economics and Social Sciences, 46 rue Barrault, 75634, Paris Cedex 13 ; tel : +33 1 45 81 73 32 ; e-mail : david.bounie@telecom-paristech.fr.

[3]ECON - 2C147, Vrije Universiteit Brussel, Pleinlaan 2 ; B-1050 Brussels, Belgium ; tel : +32 2 629 21 25 ; e-mail : Leo.Van.Hove@vub.ac.be ; Corresponding author.

[4]Many central banks have also endeavored to make the distribution of coins and banknotes more efficient (de Meijer, 2010), but such measures fall outside the scope of this paper.

[5]The golden dollar aims at replacing the Susan B. Anthony dollar introduced in 1979.

The second strategy concerns the adoption of "plastic notes" (made from polymer). One of the main advantages of plastic notes is that their life span is, reportedly, around four times that of their paper counterparts, which yields substantial savings in ordering, processing, withdrawal, destruction and redistribution costs (Coventry, 2001a). The first plastic banknote was the $10 commemorative note printed in 1988 by Note Printing Australia, a subsidiary of the Reserve Bank of Australia.[6] Today, more than 30 countries around the world - such as New Zealand, Romania, Mexico, Singapore, and Hong Kong - have partially or fully adopted plastic notes. Recent examples include India and especially Canada. The Reserve Bank of India (RBI) announced in March 2010 that it will introduce one billion 10 rupee banknotes on polymer substrate in a field trial in five cities. Governor Duvvuri Subbarao stressed that the RBI "will embark on polymer notes on a long term basis only if the cost-benefit calculus is decidedly positive in all dimensions" (Subbarao, 2010). One of these dimensions is the fight against counterfeiting of banknotes. Indeed, discouraging counterfeiting is a priority for central banks since it causes losses for the public and may even destabilize the monetary system. Polymer banknotes include novel security features that make them much more difficult and time consuming to reproduce. They are thus a deterrent to counterfeiters. The fight against counterfeiting was also the primary driver behind the Bank of Canada's (BoC) recent conversion to polymer. The BoC unveiled its new C$100 and C$50 notes to the public on 20 June 2011 (Carney, 2011). The first polymer C$100 bills will be issued in November 2011 and the C$50 notes will follow in March 2012. The other denominations will be replaced by the end of 2013 (ibidem).

While the above examples may suggest otherwise, calculating the potential gains of the adoption of a new banknote technology is not that simple for central banks : they have to disentangle the cross-effects between production costs, durability, and counterfeiting, all the while minding the side effects on seigniorage. For instance, a more costly new technology can increase the life span of a denomination and *a priori* look beneficial to a central bank, but the central bank also has to take into account how the new technology

[6]The very first plastic banknote was in fact issued in 1974 in Haiti, but the ink failed to adhere during circulation (de Heij, 2002).

affects its seigniorage revenue and the incidence of counterfeiting, both of which are closely related to the face value of a denomination (albeit not in the same way). Central banks could thus use a formal framework that integrates all these dimensions, the more so since not all of them seem convinced of the benefits of polymer banknotes, the technology that we study in more detail in this paper. The Banco Central do Brasil (BCB) is a case in point. In 2000, it launched a large-scale, four-year circulation trial with a commemorative polymer note of 10 real. In the end, the BCB concluded that there was "insufficient evidence" of the benefits of polymer, and decided to gradually withdraw the notes (Sidney, 2009 : 11).

The objective of this paper is twofold. The first is, as explained, to provide a general framework that allows central banks to assess the potential cost savings of a change in manufacturing material of their coins and/or banknotes. The second is to provide estimates of these savings for the (complete or partial) introduction of plastic banknotes by the U.S. Federal Reserve. We have chosen the case of the U.S. because the Federal Reserve is a rare example of a central bank that releases information on the production costs of its banknotes and on the production volumes per denomination. We show that while a complete adoption of plastic notes would entail a drop in monetary seigniorage revenue of roughly 0.2% (because of the higher initial production costs), it would cut by more than half the annual replacement cost of banknotes, resulting in net savings of $138.4 million per year. We also find that the $1 banknote would generate the biggest cost saving.

In doing the above, the present paper contributes to a relatively under-researched field on two main points. For one, there is the theoretical framework. To the best of our knowledge, the only academic paper that is somewhat similar is Menzies (2004). Menzies shows, for the case of Australia, that the net present value of the printing costs is lower for polymer than for paper banknotes. Our paper is more comprehensive and is the first to fully integrate the essential value drivers of the provision of currency by a central bank - namely production costs, seigniorage, and counterfeiting - in a single model. Second, we exploit a variety of statistics from reports by central banks and other institutions to calibrate our model. Within the framework of a sensitivity analysis, we compare and confront statistics from different countries on durability, production costs, and counterfeiting of plastic banknotes. This allows us to present a range of scenarios as to the size of the potential operating gains for the U.S. Federal Reserve instead of just a point estimate.

The remainder of the paper is structured as follows. In Section 2, we present a general framework that allows central banks to map the cost drivers as well as the revenue streams that are affected by a change in minting or printing technology. In Section 3, we describe the data of our case study and simulate how the production costs and seigniorage revenue of the U.S. Federal Reserve would be impacted by the introduction of plastic banknotes. In Section 4, we try to improve upon these estimates by also taking into the impact on counterfeiting. Section 5.5 concludes.

5.2 General Framework

In this Section, we present a theoretical framework that makes explicit the links between the manufacturing technology used by the central bank on the one hand and the currency production costs and revenues of the bank on the other. This framework is subsequently used to assess, still on a purely theoretical level, the impact of alternative manufacturing technologies.

In a given economy, let us consider a currency system composed of J denominations of face values $v(j)$, with $j \in [\![1,J]\!]$, and let us denote by $N_c^t(j)$ the volume of (genuine) tokens of denomination j in circulation at the end of year t. The central bank that issues the coins and/or banknotes (hereafter : tokens) obviously incurs costs (as it needs to maintain the quality of the tokens that are in circulation) but also derives revenue from its issuing activity.

5.2.1 The Currency Production Costs of the Central Bank

In modeling the cost side, we focus on those costs that are potentially affected by a change in manufacturing technology, namely the processing and replacement costs.[7] Indeed, central banks monitor the quality of the tokens that are returned, and substandard specimens are withdrawn from circulation and replaced by new ones. We denote by $n^t_{pr}(j)$ the average number of times per year a token of denomination j is processed by the central bank. The annual volume of processed notes of denomination j is then given by :

$$N^t_{pr}(j) = n^t_{pr}(j) \cdot N^t_c(j).$$

Importantly, $n^t_{pr}(j)$ is not necessarily equal to the average frequency of return of the denomination. Just how close these two indicators are to one another depends on the intensity with which the central bank wishes to verify the fitness for circulation of its tokens.[8] Turning to the costs involved, if we denote by $c^t_{pr}(j)$ the unit processing cost incurred by the central bank, then we can write the total annual processing cost, on the level of an individual denomination, as :[9]

$$C^t_{pr}(j) = N^t_{pr}(j) \cdot c^t_{pr}(j), \tag{5.1}$$

with $C^t_{pr}(j)$ and $N^t_{pr}(j)$ respectively the total processing cost and the volume of processed notes for denomination j.

As mentioned, the central bank replaces the notes that are torn, soiled or otherwise deemed unfit for re-issue. On the level of individual denominations, the volume of new tokens that are released in this way in year t, $N^t_r(j)$, is determined by the circulating

[7]The production costs of additional tokens that the central bank may need to put into circulation are not considered here, but obviously do impact the bank's seigniorage revenue modeled below.

[8]Cf. Langwasser (2010 : 60) on the case of New Zealand : "The Reserve Bank received 67.7 million bank notes in repatriations in 2009. Of these notes, 23 million were processed by the Reserve Bank's note-processing machine". Conversely, the National Bank of Romania reports that, in 2009, "out of 991 million banknotes destined for processing [...], some 95 percent were subject to processing" (Source : Annual Report 2009, p. 96).

[9]We use lower case c for unit costs and capital C for total costs.

volume of the denomination together with its average life span, $d(j)$:

$$N_r^t(j) = \frac{N_c^t(j)}{d(j)}. \tag{5.2}$$

If we then denote by $c_p^t(j)$ the unit production cost of denomination j, the total annual replacement cost incurred by the central bank for that denomination is :[10]

$$C_r^t(j) = N_r^t(j) \cdot c_p^t(j). \tag{5.3}$$

Replacing (5.2) in (5.3) we obtain :

$$C_r^t(j) = N_c^t(j) \cdot \frac{c_p^t(j)}{d(j)}. \tag{5.4}$$

5.2.2 The Seigniorage Revenue of the Central Bank

Let us now turn to the revenue side of our model. Revenue from money creation is typically called "seigniorage". However, as Neumann (1992, p. 29) complains, "unfortunately, this terms has been subject to a variety of interpretations in the literature". We thus have to specify what our understanding of seigniorage is.

For one, Neumann (1992, p. 30) points out that "seigniorage has been used interchangeably for either the total revenue or the profit derived from money production and maintenance. Of course, revenues and profits are identical only if costs are zero. Although theoretical analysis can be simplified by assuming that costs are zero, this assumption cannot be maintained in empirical applications". Given that the present paper is all about the manufacturing technology of coins and banknotes, and that a change in this technology mainly alters costs - and, as we will show, only has a marginal impact

[10] We assume that, as is common with central banks, the amortization of the fixed costs related to the printing and/or minting equipment is included in the unit production cost (Galán and Sarmineto, 2007 : 14, note 21). We have taken care to double-check this for the banknotes issued by the Federal Reserve, the central bank that we scrutinize in our case study. The 2008 "New Currency Budget" of the Federal Reserve Board confirms that capital investment costs are included in the billing rate that the Bureau of Engraving and Printing, the government agency responsible for the production of banknotes, charges to the Federal Reserve ; see http ://www.federalreserve.gov/generalinfo/foia/2008newcurrency.htm (last visit on 15/07/2010).

on gross revenues - we clearly have to include costs. In other words, we examine profit from money creation, rather than just revenue, and calculate *net* seigniorage.

Second, seigniorage usually refers to profit (or revenue) from the creation of base money, which consists of *two* components, namely currency in circulation and reserves of depository institutions. We focus on profit from the creation of currency alone. The reason is straightforward : reserves of depository institutions are not affected by a change in the printing or minting technology used by the central bank.

Finally, similar to what we do on the cost side of our model, here too we only consider those revenues that are potentially affected by a change in technology. In particular, we concentrate on what Neumann (1992) terms "monetary seigniorage" and Buiters (2007) simply calls "seigniorage".[11] Indeed, if the demand for currency grows, a central bank not only needs to replace sub-standard tokens already in circulation (as explained above), it also needs to put in circulation additional tokens in order to accommodate the increase in demand. These additional tokens constitute a source of seigniorage revenue for the central bank, amounting to the difference between their nominal value and unit production cost.[12] On the level of a denomination j, this (net) seigniorage revenue (S) is thus determined by :

$$S^t(j) = N_i^t(j) \cdot (v(j) - c_p^t(j)), \tag{5.5}$$

with $N_i^t(j)$ the annual *net* emission volume of a denomination j ($N_i^t(j) = N_c^t(j) - N_c^{t-1}(j)$).

All in all, our seigniorage concept is a narrow one. Neumann (1992, p. 31) is correct in arguing that the "traditional concept of monetary seigniorage does not provide a complete account of [...] revenue from base money provision". He therefore develops a new concept, called 'extended monetary seigniorage', which also includes the interest the central bank earns by investing the resources obtained through the past issuance of

[11] Neumann also distinguishes "opportunity cost seigniorage", but explains that this concept "does not provide a measure of the gains to the monetary authority from money creation and maintenance" (1992, p. 30).

[12] We assume that the central bank does not derive any revenue other than seigniorage from its banknote activity. In other words, we assume that it does not charge commercial banks for the processing of banknotes.

base money in interest-bearing assets.[13] However, for the purposes of the present paper there is no need to include this second component. The reason is that it is unlikely that the outstanding stock of currency is affected to any substantial extent by a change in the manufacturing technology.

Returning to our model, what is crucial is that seigniorage, as we have defined it here, is sensitive to the manufacturing technology that the central bank uses because the technology affects the unit production cost. Hence, central banks might be confronted with a trade-off to the extent that the lower maintenance costs brought about by a change in technology are offset by a decrease in seigniorage revenue. In order to be able to highlight this trade-off, we define the annual net monetary revenues from currency production for a denomination j, $R(j)$, as :

$$R^t(j) = S^t(j) - C^t_r(j) - C^t_{pr}(j) - C^t_{other}(j), \tag{5.6}$$

with $C^t_{other}(j)$ including handling costs other than the processing costs, the costs of marketing, a portion of the overhead, and other costs.

Counterfeiting also enters into play here. In our model, counterfeiting does not so much affect the central bank's net monetary revenues via an increase in $C^t_{other}(j)$, but rather manifests itself via a reduction in the demand for genuine currency, and thus in the central bank's seigniorage revenue. Importantly, counterfeits are either seized, *i.e.* detected prior to having been put in circulation, or passed, *i.e.* identified as fake after having entered circulation. Following Sumner (1993 : 26), we assume below that "successful counterfeiting results in a one-for-one reduction in demand for legitimate currency". In other words, we focus on the passed counterfeits.

Let us denote the annual volume of passed counterfeits by $N_f(j)$, with the subscript f referring to 'fake'. For reasons of comparability (across denominations and across countries), when reporting about counterfeiting central banks tend to work with counterfeit rates, and typically with the number of passed counterfeits divided (in one way or

[13] In Buiter's (2007) terminology, this is "central bank revenue".

another) by total circulation :[14]

$$r_f(j) = \frac{N_f(j)}{N_c(j)}. \tag{5.7}$$

We assume that counterfeits are detected by the authorities at worst during the sorting operations. If we denote by $d_f(j)$ the average life span of a counterfeit of denomination j, then the volume of circulating counterfeits, $N_{cf}(j)$, is :

$$N_{cf}(j) = d_f(j) \cdot N_f(j). \tag{5.8}$$

Let us stress that the concept of 'life span' of a counterfeit note differs from that of a genuine note. For counterfeit notes, 'life span' should mainly be read as 'detection time', as the amount of time a counterfeit note spends in circulation is determined by the speed of detection, not by wear and tear - except in the rather extreme case that counterfeit notes wear out before they are detected.[15] Replacing (5.7) we obtain :

$$N_{cf}(j) = d_f(j) \cdot r_f(j) \cdot N_c(j). \tag{5.9}$$

The total circulating volume (genuine *plus* fake) of denomination j is given by :

$$N_{cT}(j) = N_c(j) + N_{cf}(j). \tag{5.10}$$

This reflects the public demand for denomination j. Replacing (5.9) and rearranging the terms, we obtain :

$$N_{cT}(j) = N_c(j) \cdot (1 \ + \ d_f(j) \cdot r_f(j)). \tag{5.11}$$

Equation (5.11) shows clearly that, in our model, counterfeit notes substitute for legitimate currency. In the absence of counterfeiting, the volume of central bank notes in

[14]In practice, total circulation is usually expressed in millions of notes ; see Section 5.4.

[15]Judging from conversations with central bankers, when cash processing is centralized at the central bank, the sorting operations at the central bank constitute the ultimate (also in the sense of infallible) point of detection so that the speed of detection of counterfeits depends (in part) on the frequency with which banknotes return to the central bank. This frequency varies across countries and from one denomination to another but will typically be higher than once per year. A similar reasoning applies to situations in which part of the processing and distribution of cash is done by commercial banks.

circulation would be $d_f(j) \cdot r_f(j) \cdot N_c(j)$ higher and the central bank would earn more seigniorage.

We now use our model to study the impact of the introduction of a new manufacturing technology on the operating result of the central bank.

5.2.3 Comparing Two Manufacturing Technologies

Let us now consider the case of a central bank that migrates in year t from its existing technology $tech_1$ to a new technology $tech_2$. In our comparison, we initially assume that $tech_2$ impacts neither the annual net increase in public demand for currency, $N_d(j)$, nor the counterfeiting rate, $r_f(j)$, thus leaving unchanged the net issuing volumes, $N_i(j)$, as well as $N_c(j)$ and $N_{pr}(j)$.[16] The restriction on $r_f(j)$ is relaxed in Section 5.4. For the sake of simplicity, we also assume that the introduction of the new technology does not affect the unit cost of processing $c_{pr}(j)$ nor $C_{other}(j)$. Where notations are concerned, in this section we drop the superscripts t as the time dimension is superfluous here.

Given our assumptions and building on (5.6), the potential gains for the central bank, $\Delta R(j)$, are given by :[17]

$$\Delta R(j) = R^{tech_2}(j) - R^{tech_1}(j) = \Delta S(j) - \Delta C_r(j).^{18} \tag{5.12}$$

The first term on the right-hand side of (5.12) is the change in seigniorage revenue. Making use of (5.5), we obtain $\Delta S(j) = -N_i(j) \cdot \Delta c_p(j)$.

[16]The assumption that the new technology does not affect $N_i(j)$ nor $N_c(j)$ - which at first sight seems to imply that gross monetary seigniorage is fixed - may raise the question why we have a revenue side in our model at all, and why we have not modeled the decision as a straightforward cost-minimizing issue. One reason is that we have modeled counterfeiting as reducing the circulation of genuine currency ; cf. supra. Secondly, the explicit presence of seigniorage makes the model more general, and allows for extensions. For instance, the model can accommodate a decentralization of the cash distribution process, as done in Australia and New Zealand (Coventry, 2001b ; Lang, 2001, 2002). Unlike in the scenario considered in the main text, such a change does have an impact on N_c. We leave this issue aside for further research since the adoption of plastic banknotes does not necessarily imply a reorganization of the cash cycle. Similarly, the processed volume, $N_{pr}(j)$, is not *automatically* affected by a change in technology but requires a decision by the central bank to increase or lower its processing intensity.

[17]We build on the notations introduced in the general framework and simply add subscripts ^{tech_1} and ^{tech_2} when needed.

[18]$\Delta C_{pr}(j) = 0$ since the processed volume of banknotes and the unit processing cost are both unchanged.

If we define the relative cost-effectiveness of the two technologies as $r_p(j) = \frac{c_p^{tech_2}(j)}{c_p^{tech_1}(j)}$, we can also write this as :

$$\Delta S(j) = -N_i(j) \cdot c_p^{tech_1}(j) \cdot (r_p(j) - 1). \tag{5.13}$$

Equation (5.13) indicates that the annual seigniorage revenue generated by denomination j increases if and only if (and to the extent that) $r_p(j) < 1$; that is, if the new technology $tech_2$ is more cost-effective.

Returning to Equation (5.12), the second term on the right-hand side is the change in the annual replacement cost. Again assuming that the demand for currency is not affected by the change in manufacturing technology, we obtain from (5.4) :

$$\Delta C_r = N_c(j) \cdot \Delta \frac{c_p(j)}{d(j)}. \tag{5.14}$$

Equation (5.14) can be rewritten by introducing - similar to what we did for the unit production cost - a measure of the relative improvement in life span brought about by the new technology, namely $r_d(j) = \frac{d^{tech_2}(j)}{d^{tech_1}(j)}$:

$$\Delta C_r = N_c(j) \cdot \frac{c_p^{tech_1}(j)}{d^{tech_2}(j)} \cdot (r_p(j) - r_d(j)). \tag{5.15}$$

Hence, the annual replacement cost of denomination j decreases when $r_p(j) < r_d(j)$; that is, when the improvement in the average life span of the denomination exceeds the relative increase in unit production cost. To summarize, by replacing (5.13) and (5.15) in (5.12) we have :

$$\Delta R(j) = -N_i(j) \cdot c_p^{tech_1}(j) \cdot (r_p(j) - 1) - N_c(j) \cdot \frac{c_p^{tech_1}(j)}{d^{tech_2}(j)} \cdot (r_p(j) - r_d(j)). \tag{5.16}$$

Overall, the introduction of technology $tech_2$ will benefit the central bank if $\Delta R(j) >$

0. By rearranging the terms in (5.16) we obtain that $\Delta R(j) > 0$ if and only if :

$$r_p(j) < \frac{r_d(j) + r_i(j) \cdot d^{tech_2}(j)}{1 + r_i(j) \cdot d^{tech_2}(j)}, \tag{5.17}$$

with $r_i(j)$ the ratio of net issuing of denomination j : $r_i(j) = \frac{N_i(j)}{N_c(j)}$.

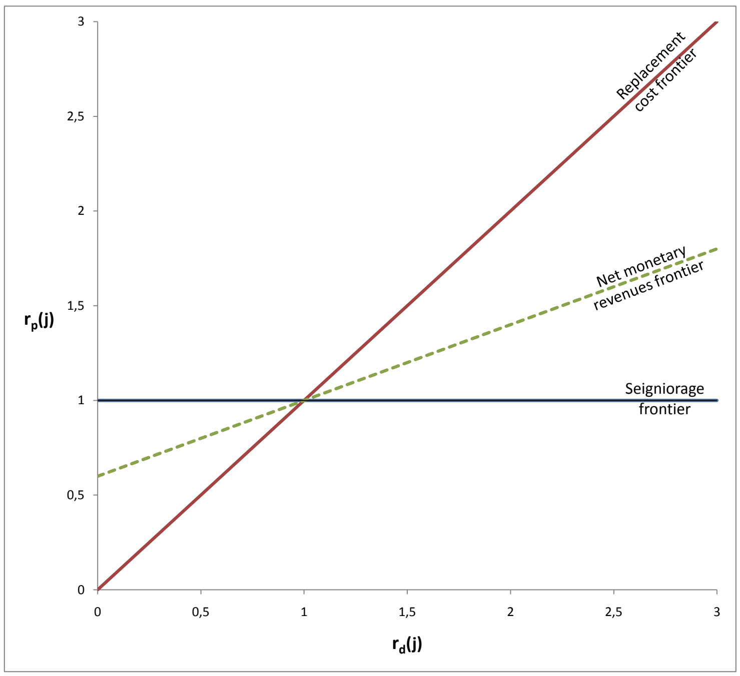

Figure 5.1 – Central Bank's Benefits or Losses of a Change in Technology.

Figure 5.1 provides a graphical analysis of the problem. The horizontal line at $r_p = 1$ captures the observation that for the seigniorage revenue to increase, r_p must be below one ; see our analysis of equation (5.13). The 45° line, for its part, separates the region where the annual replacement cost increases ($r_p(j) > r_d(j)$; to the left of the line) from that where it decreases ($r_p(j) < r_d(j)$; to the right) ; see our analysis of equation (5.15). Combining these two partial break-even lines, it is obvious that in area (A) the introduction of the new technology does not benefit the central bank at all. Indeed, in this area, the new technology $tech_2$ both increases the replacement cost and lowers seigniorage revenue. In area (B) both conditions are reversed, and the technology switch is clearly beneficial for the central bank. In the areas in between the replacement cost and seigniorage frontiers, the situation is less clear, since the central bank faces a trade-off. This is where the third and most important line comes in. The net monetary revenues frontier originates from condition (7.18), and can be thought of as the overall break-even line. In conformity with (7.18), the new technology $tech_2$ only benefits the central bank in the region below the frontier.

Note that the net revenues frontier drawn in Figure 5.1 is but an illustration. However, in drawing it we did take into account that *i.* the net revenues frontier obviously has to go through the point (1,1), where the two partial break-even lines intersect, *ii.* it has a slope that is positive but smaller than one ($\frac{1}{1+r_i(j)\cdot d^{tech_2}(j)}$), and *iii.* it has an Y-intercept ($\frac{r_i(j)\cdot d^{tech_2}(j)}{1+r_i(j)\cdot d^{tech_2}(j)}$) between zero and one. As can be inferred from the previous sentence, the exact position of the net revenues frontier - which is obviously bounded by the two partial break-even lines - is determined by the product $(r_i(j)\cdot d^{tech_2}(j))$. When this product increases, the slope of the net revenues frontier becomes smaller and the frontier rotates clockwise around the point (1,1). The explanation is as follows. To the left of (1,1), we have $r_d < r_p < 1$. Given (5.13) and (5.15), this means that both ΔS and ΔC_r are positive - implying gains in seigniorage revenue but increases in replacement costs. When the product $r_i(j)\cdot d^{tech_2}(j)$ increases, ΔS increases (thus improving seigniorage revenue) and/or ΔC_r decreases (which limits the increase in replacement costs). Hence, the "beneficial area" - the area below the net revenues frontier - becomes larger. The exact opposite happens to the right of (1,1) since the conditions are reversed ($1 < r_p < r_d$).

5.3 An Application to the U.S. Currency System

This Section aims at estimating the potential savings of the adoption by the U.S. Federal Reserve of "plastic" notes in place of conventional paper notes. We first present the data that we use for the two technologies - in Sections 5.3.1 and 5.3.2, respectively - and subsequently present and comment on the results, which relate to the period 2004-2008.

5.3.1 Production Cost and Circulating Volume of Paper Notes

In order to be able to gauge the gains, for the Federal Reserve, of adopting plastic notes, we first need statistics on the circulating and net issuing volumes per denomination, as well as on the production costs of the Fed's paper banknotes. Since the latter two are subject to important fluctuations, we opted to examine an 'average year' within the five year period 2004-2008. The Federal Reserve kindly provided us with detailed yearly information on the volume of banknotes in circulation per denomination (Table 5.I).[19] We denote by $\bar{N}_c(j)$ the average circulating volume of denomination j over the period that we examine. Next, by comparing the circulating volumes over successive years we computed the annual net issuing volumes per denomination as well as the average, $\bar{N}_i(j)$; see Table 5.II. Finally, the average ratios of net issuing are calculated as $\bar{r}_i(j) = \frac{\bar{N}_i(j)}{\bar{N}_c(j)}$.

$v(j)$ / t	\$1	\$5	\$10	\$20	\$50	\$100
2004	8,291,269	1,965,371	1,510,301	5,382,026	1,212,962	5,166,961
2005	8,756,319	2,054,448	1,553,386	5,770,322	1,242,642	5,449,550
2006	8,905,530	2,071,207	1,596,555	5,959,970	1,255,583	5,640,797
2007	9,159,021	2,126,212	1,613,477	6,088,155	1,259,350	5,692,249
2008	9,333,994	2,178,814	1,626,618	6,254,960	1,294,315	6,250,058
Average circulating volume ($\bar{N}_c(j)$)	8,889,227	2,079,210	1,580,068	5,891,087	1,252,970	5,639,923

Tableau 5.I – Annual circulation volumes (thousand of pieces) of U.S. banknotes, per denomination, over the period 2004-2008.

Note that whereas the Bureau of Engraving and Printing (BEP) reports that "95% of the notes printed each year are used to replace notes already in, or taken out of circulation" - which implies that the average ratio of net issuing amounts to 5% - Table 5.II indicates that the ratio varies substantially between denominations, and ranges from 1.6% for the \$50 note to 4.8% for the \$100 bill, with a weighted average of 3.4%.[20] The

[19] Ferrari, S., personal e-mail, April 27, 2011.

[20] The Bureau of Engraving and Printing (BEP) aims at designing, manufacturing and printing U.S. paper banknotes for delivery to the Federal Reserve System. The BEP does not produce coins ; all U.S. coinage is minted by the United States Mint. Information is available at the following address : http ://www.moneyfactory.gov/uscurrency/annualproductionfigures.html (last visit on 15/07/2010).

$v(j)$ / t	\$1	\$5	\$10	\$20	\$50	\$100
2005	465,050	89,077	43,085	388,295	29,679	282,589
2006	149,210	16,759	43,169	189,649	12,941	191,247
2007	253,491	55,005	16,922	128,184	3,767	51,452
2008	174,973	52,602	13,141	166,805	34,965	557,810
Average annual net issuing volume ($\bar{N}_i(j)$)	260,681	53,361	29,079	218,233	20,338	270,774
Average ratio of annual net issuing ($\bar{r}_i(j)$)	2.9%	2.6%	1.8%	3.7%	1.6%	4.8%

Tableau 5.II – Average net issuing volumes (thousand of pieces) and ratios of net issuing of U.S. banknotes, per denomination, over the period 2005-2008.

net issuing of large banknotes is greater in part because large banknotes are used as a means of hoarding in the U.S., but in particular because of the foreign demand for dollar banknotes (Boeschoten and Fase, 1992).

Now we need to calculate, for all denominations, the annual replacement cost, $\bar{C}_r^{paper}(j)$, and seigniorage revenue, $\bar{S}^{paper}(j)$ - again considering an 'average year' of the period 2004-2008.[21] For this, we need data on replacement volumes and unit production costs per denomination. For the first we use data provided by the Federal Reserve on the annual volumes of destroyed banknotes (Table 5.III).[22] These volumes comprise mainly worn banknotes that are replaced with new ones. For our simulation, we calculate and use the per-denomination averages over 2004-2008, which we denote by $\bar{N}_r^{paper}(j)$. We also use these figures to compute the average life spans of denominations by means of Equation (5.2).[23]

[21] $\bar{C}_r(j) = \bar{c_p}(j) \cdot \bar{N}_r(j)$ and $\bar{S}(j) = \bar{N}_i(j) \cdot (v(j) - \bar{c_p}(j))$.

[22] Source : Hodges, J., personal e-mail, April 7, 2011.

[23] In earlier versions of the paper we relied on life-span statistics published on the website of the Federal Reserve Bank of New York, and which apparently related to the year 2008 (see http ://www.newyorkfed.org/aboutthefed/fedpoint/fed01.html, last visit on 15/07/2010). However, we later found out that these figures have not been updated since the BEP conducted a formal study in the early 1990s (Ferrari, S., personal e-mail, July 12, 2011). Our new life-span estimates based on the up-to-date figures received from the Board of Governors of the Federal Reserve System are on average 1.5 times higher than the old figures, which explains why our estimates of the savings for the Fed are significantly lower in this version of the paper. Another explanation is that we now use average figures over 2004-2008 rather than just figures for 2008. Note that the primary reason for the longer note life reportedly lies

$v(j)$ / t	$1	$5	$10	$20	$50	$100
2004	3,276,709	616,288	491,846	1,840,539	179,676	338,640
2005	3,485,573	675,413	443,787	1,439,497	144,853	358,474
2006	3,615,050	712,929	486,624	1,435,361	145,470	367,082
2007	3,365,159	678,033	467,836	1,249,557	135,745	608,629
2008	2,849,728	686,107	493,047	1,311,334	122,365	1,048,947
Average annual replacement volume ($\bar{N}_r^{paper}(j)$)	3,318,444	673,754	476,628	1,455,258	145,622	544,354
Life span ($d^{paper}(j)$) (years)	2.7	3.1	3.3	4	8.6	10.4

Tableau 5.III – Annual replacement volumes (thousand pieces) and life spans of U.S. banknotes, per denomination, over the period 2004-2008.

Secondly, unit production costs can be found in the “New Currency Budgets” of the Federal Reserve Board. Note that some denominations, such as the $100 banknote, are produced in different series with different unit production costs. For instance, in 2008, the $100 banknote was produced either in Series-1996 at a unit production cost of $0.079 and for a total number of units of 854.4 million, or in Series-2004 at $0.125 apiece - for a total number of 817.7 million. To account for this, we calculated a weighted average unit production cost, which, in the case of the $100 note, amounts to $0.1015. Table 5.IV summarizes unit production costs of denominations over the period 2004-2008.

$v(j)$ / t	$1	$5	$10	$20	$50	$100
2004	0.0393	0.0521	0.0643	0.0759	0.0759	0.0643
2005	0.0401	0.0529	0.0708	0.0763	0.0763	0.0651
2006	0.0430	0.0573	0.0806	0.0806	0.0806	0.0701
2007	0.0445	0.0619	0.0835	0.0835	0.0835	0.0785
2008	0.0471	0.0742	0.0765	0.0916	0.0916	0.1016
Average unit production cost ($\bar{c}_p^{paper}(j)$)	0.0428	0.0597	0.0751	0.0816	0.0816	0.0759

Tableau 5.IV – Unit production cost of U.S. banknotes, per denomination, over the period 2004-2008 ($)

with improvements in “the ways that Reserve Banks process currency in terms of sensor technology and inventory management” (Ferrari, S., personal e-mail, July 12, 2011).

Face value	Average annual replacement cost	Average annual seigniorage revenues
$v(j)$	$\bar{C}_r^{paper}(j)$	$\bar{S}^{paper}(j)$
$1	142.0	249.5
$5	40.2	263.6
$10	35.8	288.6
$20	118.7	4,346.9
$50	11.9	1,015.3
$100	41.3	27,056.9
Total	390.0	33,220.7

Tableau 5.V – Average annual replacement cost and average annual seigniorage revenues of U.S. banknotes, per denomination, over the period 2004-2008 (million $).

Table 5.V shows that the annual replacement cost is the largest for the $1 banknote and that the annual seigniorage revenue is the largest for the $100 banknote. [24] The intuition behind the two results is straightforward. Regarding the case of the $100 banknote, we know from (5.5) that the difference between the face value and the unit production cost is largest for the $100 bill. Moreover, Table 5.II shows that the net issuing volume of $100 banknotes is the highest (at 270.8 million units per year). As a result, the seigniorage revenue is without any doubt highest for the $100 banknote. Where the $1 note is concerned, (5.4) shows that the annual replacement cost increases with the circulating volume as well as with the ratio of the unit production cost over the lifespan of a denomination. Table 5.I indicates that the circulating volume of the $1 note is by far the highest one. At the same time, it can be inferred from Tables 5.III and 5.IV that, at 0.0160, the cost/lifespan-ratio of the $1 note is also high, together with those of the $10 (0.0227), $20 (0.0202) and $5 banknotes (0.0193).

[24] Note that the total estimated annual replacement cost cannot be directly compared with the currency production cost that is reported in the Federal Reserve Annual Report (and which, for example, amounts to $500 million for the year 2008). One reason is that the latter also includes the cost of new currency. A second and more important reason is that the volume of notes that the Fed has printed by the BEP in a given year does not necessarily correspond to the volume of additional banknotes injected into circulation in that year.

5.3.2 Production Cost of Plastic Notes : An International Comparison

As mentioned in the Introduction, the Reserve Bank of Australia (RBA) was the first to successfully introduce a polymer banknote (in January 1988) and the first to use polymer for all its banknotes (by 1996). Today, more than thirty central banks throughout the world have adopted the technology. It is generally accepted that plastic notes last longer than conventional paper notes. There is also little doubt that the production of polymer banknotes is more costly. However, for our simulations we need to put figures on both $r_d(j)$ and $r_p(j)$. To start with the first, Note Printing Australia, on its website, claims that the circulation life of a polymer banknote is "around four times that of paper". The subsidiary of the RBA refers to the fact that the "paper $10 note had an average life in circulation of eight months. Today, by contrast the polymer $10 lasts at least 30 months".[25] This quadrupling of the average life span was already reported by Less Coventry (2001a, 2001b), who was at the time Head of Note Issue at the RBA. The experience is similar in Zambia (Mulomba, 2007).

However, there are reports of both lower and higher durability. In an early survey, de Heij (2002 : 6) of De Nederlandsche Bank cites "different reports" - for several countries - to conclude that the switch to polymer increases the life of banknotes "by 200-400 %". In Brazil, the trial with the commemorative polymer note of 10 real showed that its lifetime was 3.2 times that of its paper counterpart (Sidney, 2009 : 10). In Mexico, the mean lifetime of the 20 peso note rose from 8.3 to 28.8 months ; that is, by a factor of 3.5 (Galán, 2009 : 6). Where Romania is concerned, a country that has converted fully to polymer, Dumitriu (2009 : 15) talks about an increase in duration of "at least three times". Leonard Wilson Kamit, Deputy Governor, Bank of Papua New Guinea, the second country outside Australia to adopt the new technology, reports in a speech that the average life span of the 2 Kina polymer note - which is a low denomination - is two years compared to the paper note's four to six months (Wilson Kamit, s.a.). This would suggest that $r_d(j)$ equals 4 to 6. Finally, there is the experience of New Zealand. The Reserve Bank of New Zealand (RBNZ) put its first polymer banknote into circulation in

[25] Note Printing Australia, "The world standard in durability and cost effectiveness," webpage, no date, <http ://www.noteprinting.com/banknotes_durability.html> (last visit on 26/09/2010).

May 1999 and the final denomination was converted to polymer in March 2000. Early on, Brian Lang (2001), Chief Manager, Currency and Building Services, of the RBNZ reported that "the greater durability of polymer is already clearly apparent with the $20 denomination, which has been in circulation for 22 months. Our paper $20 note had an average life of 19/20 months and our current estimates show that the average life of our polymer $20 will likely be five times that figure". In a more recent article, Boaden (2009 : 13) is more conservative : "It is normal when a country issues a new series of notes for them to have a "honeymoon period" when attrition rates are very low. But at some point numbers of worn notes rise to a rather higher level. New Zealand may be experiencing that transition now. Nevertheless, the average life of the three lowest denomination polymer notes, namely $5, $10, and $20, is about 5 to 6 years, compared with 1.6 years when all notes were paper, *i.e.* about 3 to 4 times the earlier paper note life".[26]

This short round-up of estimates suggests a range for $r_d(j)$ between roughly 3 and 5. Obviously, a lot depends on the quality of polymer that is chosen and on the precise security features. But, crucially, because $r_d(j)$ is a relative measure, the quality of the paper notes prior to their displacement also matters. Indeed, for the case of the U.S., the Committee on Technologies to Deter Currency Counterfeiting of the National Research Council stresses that while "it has been reported that the lifetime of plastic banknotes can be as great as four times that of paper" (2007 : 74), a 2002 Federal Reserve Board study highlighted "the fact that the life of the $1 note was more than 20 months in 2001, which is a longer duration than the life span of comparable denominations in several other countries" (ibidem). The Committee concludes that "these data indicate that a plastic substrate would provide about a *twofold* lifetime increase for U.S. currency" (ibidem ;

[26]Two recent articles in the RBNZ Bulletin present detailed numbers on the "destruction rates" of polymer vs. paper notes. In 2008, across *all* denominations, 16.5% of all polymer notes in circulation were destroyed. In 1998, when all banknotes were still made from cotton-based paper, this figure was 59.5% (Boaden and Langwasser, 2009, Table 3 : 26) ; that is, 3.6 times higher. An update for the year 2009 makes mention of a destruction rate of 11.4% for polymer notes (Langwasser, 2010, Table 3 : 61), which implies that the relative improvement would be as high as 5.3. Boaden (2009 : 13), for his part, works with *average* destruction rates over 1995-1998 and 2006-2008, respectively. These rates are almost exactly 5 times lower for polymer than for paper.

emphasis added), rather than the generally projected fourfold increase. Interestingly, the Bank of Canada, in its preliminary calculations, uses a figure that comes close to the U.S. figure just mentioned, namely 2.5. However, the BoC considers this figure "conservative according to other countries' statistics" (PE Americas and Tryskele, 2011 : 11) ; see also Spencer (2011 : 5).

In view of all this, we decided to conduct a sensitivity analysis with the values for $r_d(j)$ ranging from 2 to 5, but with $r_d(j) = 4$ as the benchmark scenario, as this is the value that is mentioned most often.[27] We will come back later to the question of whether this is realistic in the specific case of the U.S. Where the relative cost-effectiveness of the two technologies, $r_p(j)$, is concerned, estimates are far more scarce. In fact, just about all evidence relates to the Australian case. Coventry (2001a), in his simulation, assumes that polymer notes cost on average twice as much as paper notes. This is also the figure used by Menzies (2004 : 362), who provides more details : "It is assumed that paper notes have been sold to the Reserve Bank at around 6-7¢, while polymer notes have been sold to the RBA at around 12-14¢."[28] This factor of two is confirmed by Spencer (2011 : 6) for the case of Canada : "The initial cost of the polymer notes is about twice that of paper notes". Note that Canada's new notes are printed on a substrate of Australian making (Spencer, 2011 : 2). However, in our sensitivity analysis we also consider the situation where $r_p(j)$ would be 3. A final remark is that in our "preferred" scenario - where $r_d(j) = 4$ and $r_p(j) = 2$ - the polymer technology would, in terms of

[27]One could raise the question whether we should not allow for variation, across denominations, in the relative improvement in life span, rather than simply assume that $r_d(j)$ is identical. In this respect, the following statement by Coventry, for the case of Australia, is interesting : "In Australia, we have experienced a quadrupling of the average life of our low denomination notes, ... Our higher denomination polymer notes have not been in circulation long enough to be precise about their longevity, but indications are that we will see a *similarly* impressive performance" (Coventry, 2001b ; emphasis added). Where New Zealand is concerned, the destruction rates mentioned earlier do vary across denominations, but with no clear pattern. For the year 2008, the rates vary between 3.0 and 4.5, with an average of 3.6, as mentioned in footnote 26. For 2009, the range is 4.5 to 7.3, with an average of 5.3. In any case, even with unambiguous per-denomination information, a transfer to another country would not be straightforward, especially since much depends on the actual choices made concerning the polymer notes.

[28]This is in line with a statement by Lang (2002 : 55) concerning New Zealand : "The issue cost per note in circulation has decreased by 58 per cent between 1997/98 (when all notes issued were paper) and 2000/01 (when all notes issued were polymer)." However, Lang's cost concept is broader than just production, and also includes transportation costs and even local taxes (ibidem).

Figure 5.1, lie in the ambiguous area to the right of the intersection of the replacement cost and seigniorage frontier and in between the two, where there is a trade-off between lower replacement costs and lower seigniorage revenue.

5.3.3 Gains for the Federal Reserve

By bringing together the results of the two previous sections, we can now estimate the potential gains, for the Federal Reserve, of the introduction of the polymer technology. We start by studying in detail our benchmark scenario (where $r_d(j) = 4$ and $r_p(j) = 2$). Afterwards, we present the aggregate results of the alternative scenarios in order to check just how sensitive our results are to changes in the assumptions.

As set out in Section 5.2.3, plastic notes will benefit a central bank for a denomination j when

$$r_p(j) < \alpha(j),$$

with $\alpha(j) = \frac{r_d(j)+r_i(j)\cdot d^{polymer}(j)}{1+r_i(j)\cdot d^{polymer}(j)}$. Table 5.VI indicates that this profitability condition is satisfied for all U.S. denominations, albeit only by a narrow margin for the \$100 bill. (The 2.0 in the Table is in fact 2.0034.)

Going into more detail, Table 5.VII presents estimates of the change in annual replacement cost and seigniorage revenue and reveals why the replacement of the \$100 banknote would be just about neutral for the Federal Reserve : the lower replacement costs only just offset the decrease in seigniorage revenue. This said, the simulation does not take into account the gains related to a decrease in counterfeiting ; see, however,

$v(j)$	$r_p(j)$	$r_d(j)$	$r_i(j)$	$d^{paper}(j)$	$d^{polymer}(j)$	$\alpha(j)$
\$1	2	4	2.9%	2.7	10.7	3.3
\$5	2	4	2.6%	3.1	12.3	3.3
\$10	2	4	1.8%	3.3	13.3	3.4
\$20	2	4	3.7%	4.0	16.2	2.9
\$50	2	4	1.6%	8.6	34.4	2.9
\$100	2	4	4.8%	10.4	41.4	2.0

Tableau 5.VI – Testing the Profitability of a Change in Technology, per denomination.

Section 5.4. [29]

Denomination (Banknote)	Change in annual replacement costs	Change in annual seigniorage revenue	Potential gains
$v(j)$	$\Delta\bar{C}_r(j)$	$\Delta\bar{S}(j)$	$\Delta\bar{R}(j)$
\$1	-71.0	-11.2	59.9
\$5	-20.1	-3.2	16.9
\$10	-17.9	-2.2	15.7
\$20	-59.4	-17.8	41.6
\$50	-5.9	-1.7	4.3
\$100	-20.7	-20.6	0.1
Total	-195.0	-56.5	138.4

Tableau 5.VII – Potential Gains of the Introduction of Plastic Notes, in million \$.

Aggregating the per-denomination changes in seigniorage revenue (cf. (5.13)) :

$$\Delta S = -\sum_{j \in Banknotes} N_i(j) \cdot \Delta c_p(j) = -\$56.5 \text{ million},$$

and replacement cost (cf. (5.14)) :

$$\Delta C_r = \sum_{j \in Banknotes} N_c(j) \cdot \Delta \frac{c_p(j)}{d(j)} = -\$195.0 \text{ million},$$

the overall potential gains, excluding counterfeiting, amount to :

$$\Delta R = \Delta S - \Delta C_r = \$138.4 \text{ million}. \tag{5.18}$$

The results of the other scenarios are summarized in Table 5.VIII. It can be seen that the gains are sensitive to the parameters $r_p(j)$ and $r_d(j)$: a unit variation of one or both

[29] As a reminder, the estimates also do not take into account possible cost savings due to lower processing costs ; see equation (5.12). As pointed out by Coventry (2001b), the switch to polymer enabled the RBA to significantly scale back its note processing activities. The rationale was straightforward : “Because of the greater durability and confidence in the security of polymer notes, the RBA believed that it did not need to check notes for authenticity and fitness as frequently as it had done in the past with paper notes” (ibidem). However, as already explained in footnote 16, this parameter is at the discretion of the central bank and therefore difficult to model. Moreover, we lacked data on the processing costs of the Federal Reserve.

$r_d(j)$ / $r_p(j)$	2	3	4	5
2	-56.5 (0)	73.4 (80.2)	**138.4** (138.4)	177.4 (177.4)
3	-266.9 (0)	-113.1 (0)	-15.6 (21.5)	42.9 (67.5)

Tableau 5.VIII – Sensitivity Analysis of the Potential Gains of the Introduction of Plastic Notes, in million \$ - Partial implementation results between brackets (aggregating only positive results).

of them has a substantial impact. This is especially the case for changes in $r_p(j)$. Indeed, if we move from our benchmark scenario (2,4) to scenario (3,4), the gains turn into losses, which are accentuated if, on top of this, $r_d(j)$ is lowered by one unit (cf. scenario (3,3)). This said, a unit change in $r_p(j)$ from 2 to 3 is quite substantial in percentage terms. Moreover, in the literature we did not find claims that $r_p(j)$ would effectively be as high as 3 ; cf. Section 3.2. For $r_d(j)$, however, there is substantial variation in the reported estimates ; see again Section 3.2. Crucially, if the National Research Council is correct in claiming that, in the case of the U.S., $r_d(j)$ would only be 2, a switch to polymer would - provided that $r_p(j)$ equals 2 - bring no gains at all, for none of the denominations ; cf. scenario (2,2). However, if the longer duration of the U.S. paper banknotes is due to paper of a better quality that is more expensive, a low $r_d(j)$ might go hand in hand with a low $r_p(j)$ - *i.e.*, lower than 2 - so that the results might be radically different. This said, using $r_d(j) = 4$ does imply lifespans of the polymer notes that are rather long, as can be seen in Table 5.VI. For example, the lifespan of the \$100 polymer banknote would be more than 40 years. Clearly, in reality central banks will typically introduce a new generation of banknotes more frequently than this. On the other hand, Menzies (2004) in his calculations for Australia uses 32 years as the average lifespan of the A\$100 banknote.

It should also be stressed that the aggregate results in Table 5.VIII hide disparities between denominations. Indeed, depending on the scenario, the introduction of plastic notes can generate losses for one or several denominations. As the figures between brackets indicate, in certain scenarios the central bank then has reason to be selective in its choice of denominations. Losses occur especially for high-value denominations.[30] This is related to the high seigniorage revenues that they generate : compared to lower-value denominations, an identical percentage increase in production costs results in a larger absolute loss of seigniorage revenues. More generally, the potential gains are negatively correlated to the face value of denominations (for our benchmark scenario the correlation coefficient amounts to -0.67).[31]

5.4 The impact of counterfeiting

In the previous Sections, we assumed that the change in technology does not affect the counterfeiting rate and thus does not affect $N_i(j)_j$. In this Section, we relax this assumption. This makes sense because the polymer technology was initially developed to increase the security of banknotes and fight against counterfeiting.[32] Counterfeiting is a major issue for central banks in general and for the U.S. Federal Reserve in particular. In a detailed analysis for the year 2002, Judson and Porter (2003, 2010) estimate the value of the counterfeits passed at about $40 million. The second column in Table 5.IX provides details as to the number of counterfeits passed per denomination. Note that the $100 bill accounts for 39% of the total number (and no less than 78% of the total value). In the third column, we express these figures in terms of numbers of counterfeits per

[30]As an aside, where the euro-zone is concerned, several authors have argued that the 200 and 500 euro banknotes are not needed for legal, everyday transactions (Van Hove and Vuchelen, 1996 ; Rogoff, 1998 ; Buiter, 2005). If the ECB were to introduce polymer notes, it could thus take advantage of the switch to remove these high-denomination notes from circulation.

[31]This result is in line with one of the outcomes of Menzies' model : "the benefits of polymer appear quickly for low denomination notes. Given this, it is easy to mount a case for converting low denomination notes first. Since they disintegrate most rapidly, they account for a large bulk of printing costs. ... Converting low denomination notes is therefore likely to speed up cost recovery, and fund the conversion of higher denomination notes" (2004 : 358-359). In their overview of real-life experiences, Galán Camacho and Sarmineto Paipilla (2007 : 12) observe that central banks effectively adopt polymer mainly for their low-value denominations. According to Kristin Langwasser of the RBNZ, part of the explanation for the fact that some countries mix their denominations - using polymer or a hybrid substrate for their lower denominations and paper for the higher-denomination banknotes - might be that sometimes polymer notes are too durable compared to the lifecycle of the note series : "a note lasting 20 years might not be needed if a central bank upgrades and replaces the entire series every ten years" (Langwasser, K., personal e-mail, October 11, 2010).

[32]According to Coventry (2001a), the Reserve Bank of Australia took the decision to move to a polymer substrate mainly to improve the overall security of its banknotes. See also Rankin (2009 : 8).

million notes in circulation, in order to make them comparable with figures for other countries below.

Denomination (Banknote)	Volume of paper counterfeits passed in (2002)	Counterfeit ratio of paper notes in (2002) (units per million notes in circulation)
$v(j)$	$N_f^{paper}(j)$	$r_f^{paper}(j) \cdot 10^6$
$1	14,960	1.6
$5	32,643	14.8
$10	106,221	65.2
$20	284,357	45.5
$50	41,848	32.3
$100	312,591	50.0
Total	792,620	29.2*

Tableau 5.IX – Counterfeits Passed in the U.S. by Denomination (* : weighted average).

In order to add the impact on counterfeiting to the estimates, reported in Section 5.3, of the gains of the polymer technology for the Fed, we mainly build on the experience of New Zealand, Romania and Australia, three countries that have migrated fully to polymer and for which we have detailed data.

Initially, we simply intended to *i.* compare counterfeiting levels - in each of the three countries mentioned - prior to and after the transition to polymer, *ii.* express the impact on counterfeiting as a percentage (or as a factor), and - provided that the estimates converged - *iii.* apply the obtained percentage/factor to observed counterfeiting figures for the U.S. This approach proved to have several shortcomings. For one, the choice of the two benchmark years is not always so straightforward because the polymer notes are typically introduced gradually. In the case of Australia, a plastic $5 banknote was issued into general circulation as early as 1992. Hence, using 1995-96 - the period when the migration to plastic was completed - as the first benchmark might underestimate the true impact on counterfeiting. Also, even when all denominations have moved to polymer, some paper banknotes may still circulate and, if so, there will always be some counterfeiting of the paper denominations. This argues against using a year immediately after the completion of the migration to plastic as the second benchmark year. Another argument in this direction is that right after the introduction of polymer, counterfeiters may not yet have adjusted so that the estimates give an incorrect picture of the longer-run impact on counterfeiting (Barry and Lang, 1999 : 45 ; Lang, 2002 : 53). All this explains why we decided to not follow this method.

Instead, we decided to focus on *current*, "sustainable" post-polymer counterfeiting levels in countries that have fully adopted plastic banknotes and use those levels as proxies for what might also be obtainable for the U.S. if it were to introduce polymer banknotes (of similar quality). For instance, looking at data for New Zealand, the figures presented in Langwasser (2010, Figure 6 : 62) show that over the period 2001-2009 the number of counterfeits found per million notes in circulation has moved in the range 0.4-1.7. Concerning Romania, Dumitriu (2009 : 15) reports that in 2007 the counterfeit rate was 2.23 per million banknotes in circulation - a figure that is only marginally higher than the rates observed in New Zealand. However, by 2009 the Romanian counterfeit rate had increased to 4.55 (National Bank of Romania, Annual Report 2009, p. 99). In Australia, this figure was about 7 over 2007/08-2009/10, and around 6 in 2006/07.[33] Taken together, this would suggest a range of 2 (cf. New Zealand) to 7 (cf. Australia), which - when confronted with the figure of 29 for the U.S. - would imply that counterfeiting of U.S. dollar notes could be reduced by a factor of 4 to 14.[34]

Before we discuss our results, it is worth noting that this second method also has its drawbacks. For example, it results in identical counterfeit ratios for all denominations - which is contradicted by real-life experience. Likewise, it could be argued that, in view of its international acceptance, counterfeiters will always be more interested in the U.S. dollar than in other currencies. The lower end of the 2-7 range derived above is therefore probably not realistic. We thus decided to adopt $r_f^{polymer}(j) = 7$ as the benchmark scenario. However, it will become clear below that whatever the scenario considered, the estimates do not differ much from the benchmark scenario.

[33]Source : RBA Annual Reports for 2007/08, 2008/09 and 2009/10. Rankin (2009 : 8) presents counterfeit rates per calendar year. Figure 1 in Rankin (2009 : 8) confirms that counterfeit rates have been very stable since 2003.

[34]The experience in Mexico, another country for which we have detailed counterfeit data, points in the same direction. The Banco de Mexico introduced a 20 peso polymer banknote in September 2002. Today, the current counterfeit ratio for the 20 peso note is 3.2 counterfeits per million notes in circulation (Galán, 2009).

As explained in Section 5.2.2, in our model counterfeiting reduces the demand for genuine currency and, in this way, also reduces the central bank's seigniorage revenue. Hence, if and when the central bank migrates to a manufacturing technology that is less prone to counterfeiting, as we now assume in this Section, its seigniorage revenue will increase. Crucially, this positive impact of a reduction in counterfeiting on seigniorage is two-fold : there is not only a (one-off) impact on $N_c^t(j)$ - the volume of genuine banknotes already in circulation in year t - in later years, starting from year $t+1$, it will also affect $N_i(j)$, the net emissions of genuine notes.[35]

Indeed, once the new polymer notes are launched, counterfeiters will find it progressively harder to peddle their current-generation counterfeits, and over time the incidence of counterfeits in $N_c(j)$ will drop from the current rate, r_f^{paper}, to $r_f^{polymer}$.[36] Just how long this will take depends on the speed with which the paper banknotes are withdrawn. Below we assume, for the sake of simplicity, that the complete migration to polymer takes place in one single year (the year of introduction), and this for all denominations. To the extent that these conditions are not fulfilled, what we deem a one-off effect below will in practice be spread over multiple years.

Apart from this 'one-off' effect, the reduction in counterfeiting also positively affects what we have called the central bank's annual seigniorage revenue - the seigniorage revenue that stems from the additional tokens the central bank puts into circulation each year - provided, that is, that the volume of currency in circulation effectively grows, $N_i^t(j) > 0$. Indeed, such net emissions will now contain a larger proportion of genuine notes.

In order to quantify these two effects, we first assume that neither the public demand for currency, $N_{cT}(j)$, nor the growth in this demand, $\Delta N_{cT}(j)$, are affected by

[35] For the sake of completeness, it should be pointed out that the reduction in counterfeiting will also have a positive impact on the second component of Neumann's extended monetary seigniorage - a component which we disregard, as explained in 5.2.2 - namely, the interest revenues of the central bank from the issuance of currency. The increase in these revenues will, however, be minor.

[36] If $r_f^{polymer}$ is a 'sustainable' counterfeiting level - as it is in our calculations ; cf. supra - the observed counterfeiting level may initially fall *below* $r_f^{polymer}$, as counterfeiters may need some time to adjust to the new technology. The resulting additional seigniorage revenues - 'additional' compared to our estimate below - will, however, be temporary. In this sense, our estimates of the impact of counterfeiting are long-run.

the technology or by the extent of counterfeiting (that is, under the current technology counterfeiting is not so rampant that it has undermined public confidence in banknotes). Therefore :

$$\Delta N_{cT}^{polymer}(j) = \Delta N_{cT}^{paper}(j).$$

If we then start with the one-off effect on $N_c^t(j)$, it is useful to repeat - as indicated earlier in Equation (5.10) and (5.9) - that the total circulating volume (genuine *plus* fake) of denomination j is given by

$$N_{cT}^t(j) = N_c^t(j) + N_{cf}^t(j), \tag{5.19}$$

with

$$N_{cf}^t(j) = d_f(j) \cdot r_f(j) \cdot N_c^t(j). \tag{5.20}$$

Given this, a decrease in $r_f(j)$ generates, as explained, an increase in $N_c^t(j)$:

$$\Delta N_c^t(j) = -\Delta N_{cf}^t(j) = -d_f(j) \cdot N_c^t(j) \cdot \Delta r_f(j),$$

which, in turn, results in the following additional seigniorage revenue :

$$S^m(j) = \Delta N_c^t(j) \cdot (v(j) - c_p(j)),$$

with the superscript m referring to 'migration'.

Summing up across all denominations, for the case of the U.S. - assuming $r_f^{polymer} = 7$ as announced above - we find the total effect to be in the range of \$31 million ; cf. Table 5.X.[37] As can be seen, this overall figure hides large disparities between denominations, with the bulk of the benefit unsurprisingly coming from the \$100 note. Table 5.XI shows that the magnitude of the total effect is not very sensitive to the choice of the benchmark $r_f^{polymer}$.

[37]Due to the lack of public data, we assume that the passed counterfeits in 2008 are equivalent to those in 2002. Note also that the current counterfeit ratio of the paper \$1 denomination is already below our sensitivity analysis interval (2-7). There are good reasons to believe that this level would not increase once the paper banknote is replaced by a polymer banknote. We therefore left it unchanged in our simulations.

$v(j)$	$r_f^{paper}(j) \cdot 10^6$	$r_f^{polymer}(j) \cdot 10^6$	$S^m(j)$ (thousand $)
$1	1.6	1.6	0
$5	14.8	7	79
$10	65.2	7	906
$20	45.5	7	4,499
$50	32.3	7	1,580
$100	50.0	7	24,215
Total	-	-	31,279

Tableau 5.X – Seigniorage Revenue per Denomination from the Adjustment of the Circulating Volume of Genuine Banknotes.

$r_f^{polymer}(j) \cdot 10^6$	$\sum_j S^m(j)$ (thousand $)
2	35,119
3	34,351
4	33,583
5	32,815
6	32,047
7	31,279

Tableau 5.XI – Impact of the First Effect of the Reduction in Counterfeiting : Sensitivity Analysis.

Let us now turn to the second, annual effect. In order to highlight the effect of counterfeiting on the net emissions of the central bank, we compare two consecutive years t and $t+1$. For a given manufacturing technology, we assume that the counterfeit rate, r_f, and the detection time of counterfeits, d_f, are stable over time. Therefore, building on (5.11) we have :

$$\Delta N_{cT}(j) = N_{cT}^{t+1}(j) - N_{cT}^{t}(j) = N_i^{t+1}(j) \cdot (1 \; + \; d_f(j) \cdot r_f(j)). \tag{5.21}$$

Combining (5.21) with the assumption that $\Delta N_{cT}^{polymer}(j) = \Delta N_{cT}^{paper}(j)$ and rearranging the terms, we obtain :

$$N_i^{polymer}(j) = N_i^{paper}(j) \cdot \beta(j),^{38}$$

[38] We assume that counterfeits of plastic and paper notes have the same 'life span', $d_f(j)$, even though it could be argued that since plastic notes are more difficult to counterfeit, the quality of the counterfeits will most likely be lower so that the public is quicker to detect them. It is, however, difficult if not impossible to quantify this effect.

with $\beta(j) = \frac{1+d_f(j)\cdot r_f^{paper}(j)}{1+d_f(j)\cdot r_f^{polymer}(j)}$. That is, the net issuing volume is higher under the new technology ($\beta(j) > 1$) if and to the extent that $r_f^{polymer}(j) < r_f^{paper}(j)$. In words, to the extent that the new manufacturing technology is more resistant to counterfeiting.

Making use of (5.5), the additional annual seigniorage revenue brought about by a reduction of counterfeiting, denoted by $\Delta S^f(j)$, can then be written as :

$$\Delta S^f(j) = N_i^{polymer}(j) \cdot (v(j) - c_p^{polymer}(j)) - N_i^{paper}(j) \cdot (v(j) - c_p^{paper}(j)).$$

Rewriting this as a function of $\beta(j)$ and $r_p(j)$, we obtain :

$$\Delta S^f(j) = -N_i^{paper}(j) \cdot [c_p^{paper}(j) \cdot (\beta(j) r_p(j) - 1) + v(j) \cdot (\beta(j) - 1)]. \qquad (5.22)$$

Note that when $\beta(j) \backsimeq 1$, we obtain $\Delta S^f(j) = \Delta S(j)$, defined in Equation (5.13).

Now making the calculations for the U.S. data, one finds $\beta(j) \backsimeq 1.0001$ for all j. In other words, the annual impact of the decrease in counterfeiting, resulting from the introduction of the polymer technology, on the central bank's seigniorage revenue is insignificant.

Let us mention, before concluding this Section, that our simulations do not take into account a possible longer-term benefit of a reduction in counterfeiting, in that the central bank may not need to replace its banknote series as often anymore. Indeed, Robert Rankin, Assistant Governor (Currency) of the RBA and Chairman of Note Printing Australia, points out that "one benefit of having very low counterfeit rates [...] is that there has not been the same pressure as in other countries to upgrade the banknote series with enhanced security features or to introduce a new banknote series" (Rankin, 2009 : 9). This can, however, not be incorporated in a one-period model like ours. We also lacked data on the magnitude of such potential savings.

5.5 Conclusion

This paper develops an original, general framework that allows central banks to assess the potential cost savings of a change in manufacturing material of their coins and/or banknotes. The model is comprehensive in that it integrates the essential value drivers of the provision of currency by a central bank - namely production costs, seigniorage, as well as the impact on counterfeiting for which it distinguishes a dual impact - in a single model. Unlike Menzies (2004), our model is not multi-period but single-year. It does, however, allow to take into account both one-off and recurring effects, and in our calculations for the U.S. we work with averages over a five-year period so as to simulate a 'typical' year.

Specifically, we use the model to assess whether the migration from paper to plastic banknotes would be beneficial for the U.S. Federal Reserve. Exploiting data on the production cost of U.S. paper notes for the period 2004-2008, the simulations of our benchmark scenario show that while a complete adoption of plastic notes would entail a drop in seigniorage revenue of roughly 0.2% (because of the higher initial production costs), it would cut by more than half the annual replacement cost of banknotes, resulting in net savings of $138.4 million per year. This overall estimate does not include the impact from the decrease in counterfeiting. As explained in Section 4, the recurrent gains are negligible.[39] There is, however, a one-off windfall which was estimated at $31.3 million.

On the level of individual denominations, we find that the $1 banknote would generate the biggest cost saving. This finding is interesting because, as mentioned in the Introduction, the U.S. Federal Reserve has attempted twice to replace the one dollar bill by a coin. This raises the question whether a one dollar plastic note would not be more beneficial for the Fed. Reportedly, the "golden dollar" costs 5.6 times more to produce than the $1 paper note, but lasts much longer (about 25 years compared to 2.7 years for the $1 banknote).[40] In the case of a complete adoption of the golden dollar, the savings

[39] This is in line with the findings of Menzies (2004) : "Given the small magnitudes, the direct seigniorage benefit [from the decrease in counterfeiting] is not added to the calculations".

[40] The unit production cost of the $1 coin is provided in the 2009 Annual Report of the United States

in replacement costs would amount to \$56.8 million, and the drop in seigniorage revenue to \$51.3 million.[41] The potential savings are thus only about \$5.5 million, which is marginal compared to those of the \$1 plastic note estimated at \$59.9 million in Table 5.VII. Hence, from this perspective, the introduction of a one dollar plastic note is recommended for the Fed.

Returning to our overall result, the two most obvious points of comparison are Australia and Canada. Concerning Australia, Menzies (2004 : 365) mentions that "the net benefit [for the central bank] after 10 years is just over A\$100 million" (roughly \$72 million at 2004 exchange rates). However, Menzies' estimates are difficult to compare to ours because he assumes that after the migration the paper notes continue to circulate and are not withdrawn until they wear out. Moreover, he also considers a scenario of gradual, per-denomination migration. That leaves the case of Canada. Spencer (2011 : 6) states that "[the longer note life] will result in a saving of at least [C]\$200 million, or more than 25 per cent of the total production costs, over the assumed eight-year life of the series, when compared with the option of achieving a similar level of resistance to counterfeiting by adding new security features to paper-based notes". This would imply an annual saving of C\$25 million and roughly the same in US dollars. Crucially, the savings compared to the current generation of paper notes are smaller, but Spencer does not reveal the exact amount. On the other hand, one also has to factor in that the number of Canadian banknotes in circulation is 15 times smaller than the number of US banknotes, which would put the Canadian estimate at C\$375 million in equivalent terms. Then everything depends on the relative magnitude of the savings compared to the current generation of paper notes and those compared to paper notes that offer more protection against counterfeiting. If the former are only one third of the latter, we arrive at C\$125 million - a figure which is in the same ballpark as our estimate for the US. But if ratio is more one-to-two, then we arrive at C\$188 million, suggesting that the savings

Mint, available at : *http://www.usmint.gov/about_the_mint/?action=annual_report* (last visit on 15/07/2010). Likewise, the U.S. Mint's website indicates that the life span of U.S. coins in general is 25 years (*http://www.usmint.gov/faqs/circulating_coins/* ; last visit on 15/07/2010).

[41] This result does not include the impact of counterfeiting on seigniorage revenue as we have no statistics on counterfeiting for low-value denominations.

are more substantial for the Bank of Canada than for the Fed.

Regardless of how the savings compare to those for other central banks, the estimated absolute magnitude of the potential cost savings for the Fed raises the question why it has not already adopted the polymer technology. There are a number of elements that might help explain this. For one, there are transition costs in that the central bank's sorting and detecting devices, for example, will during a certain period have to handle the newly introduced plastic notes in parallel with the old paper notes.[42] Second, one should realize that certain central banks have made considerable investments in the development and production of special paper for their banknotes. The economic lifetime of such investments is quite long. A third element is specific to the U.S. and is related to the circumstance that over half of the value of U.S. currency is thought to be held outside the country (Judson and Porter, 2003). Some experts are convinced that the U.S. Treasury is worried that the introduction of new notes with a radically different look and feel would cause foreign holders to think that the old notes are no longer "good". This could trigger a sizeable "redemption run", and thus a (minor) hit to the government budget). Finally, perhaps our benchmark scenario is somewhat optimistic, especially where the lifespan of the polymer notes is concerned.

Broadening the scope, it is worth noting that the adoption of plastic notes may well be beneficial for a central bank but might be negative for other parties, namely banks and merchants who have to adapt ATMs, vending machines, etc. For Canada, Spencer (2011 : 6) estimates this conversion cost at as much as C$75 to C$100 million, but is convinced that "the cash system will benefit in the long run from the change to a polymer substrate". Where New Zealand is concerned, Lang (2002 : 55) points out that because the RBNZ retained the same size notes, the country's entire ATM network was converted in a very short period (about a month), as the change-over only required a software adjustment. Coventry (2001b) makes mentions of an RBA survey among users and suppliers of machines that process notes (such as ATMs and note counters). The

[42]However, for the case of New Zealand, Lang (2002 : 55) reports that, because the same machine-readable covert features that were incorporated in the paper notes were implanted into their polymer replacements, an inexpensive software upgrade enabled the RBNZ's processing machines to identify paper notes for destruction and to band polymer notes for re-issue.

survey revealed that polymer notes are actually better for machine processing than paper notes : there are fewer jams, fewer service call outs, and maintenance staff can be reduced.[43] In addition, Coventry (2001b) points out that, because of their higher quality and durability, polymer notes allow increased recirculation of currency between the various participants, leading to “a reduction in the socially wasteful excessive amount of churn and cross-shipping of notes”. It would thus be interesting in future research to extend our approach to a calculation of the (net) social cost of plastic notes. Coventry (2001b), for one, is convinced that the move to polymer has not only been “very good for the RBA” but also “for the community generally”.

5.6 Literature cited

Barry J. and Lang B., 1999, “Polymer Bank Notes,” *Bulletin*, Reserve Bank of New Zealand, (62)2 : 44-46.

Boaden A., 2009, “New Zealand’s Experience with Polymer Banknotes,” *Billetaria - International Review on Cash Management*, Banco de Espana, (3)5 : 12-13.

Boaden A. and Langwasser K., 2009, “Recent Trends and Developments in Currency,” *Bulletin*, Reserve Bank of New Zealand, (72)1 : 23-30.

Boeschoten W. and Fase M., 1992, “The Demand for Large Bank Notes,” *Journal of Money, Credit, and Banking*, 24(3) : 319-337.

Buiter W.H., 2005, “New Developments in Monetary Economics : Two Ghosts, Two Eccentricities, a Fallacy, a Mirage and a Mythos”, *Economic Journal*, 115(502) : C1-C31.

Buiter W.H., 2007, “Seigniorage”, *Economics : The Open-Access, Open-Assessment E-Journal*, 1(2007-10) (http ://www.economics-ejournal.org/econ omics/discussionpapers/2007-8).

[43]The experience in New Zealand is similar. Lang (2002 : 55) reports that one of the main ATM servicing companies had seen an average decline of 50 per cent in fault call-outs. In Brazil, however, professional cash handlers strongly rejected the 10 real polymer note, with 95% of the banks and cash-in-transit companies preferring the paper note. There was even resistance on the part of the banks to use the note in ATMs (Sidney, 2009).

Carney M., 2011, “Canada’s New Polymer Bank Notes - Celebrating Canada’s Achievements at the Frontiers of Innovation,” speech, Bank of Canada, Ottawa, Ontario, June 20 ;
http ://www.bankofcanada.ca/2011/06/speeches/canada-new-polymer-bank-notes/ (last visit on 14/07/2011).

Committee on Technologies to Deter Currency Counterfeiting, National Research Council, 2007, “*A Path to the Next Generation of U.S. Banknotes : Keeping Them Real*,” The National Academies Press, Washington, D.C.

Coventry L., 2001a, “Polymer Notes and the Meaning of Life,” Currency Conference, Barcelona, Spain, April.

Coventry L., 2001b, “Cost-Effectiveness of Polymer Currency Notes - Australia’s Experience,” XV Pacific Rim Banknote Printers’ Conference, Thailand, November.

de Heij H., 2002, “Durable Banknotes : an Overview,” Presentation of the BPC Paper Committee to the BPC General Meeting, Prague, Czech Republic, May 27-30.

de Meijer C.R.W., 2010, “The Single Euro Cash Area : Towards a More Efficient European Cash Society,” *Journal of Payments Strategy & Systems*, (4)3.

Dumitriu R., 2009, “Polymer Banknotes in Romania,” *Billetaria - International Review on Cash Management*, Banco de Espana, (3)5 : 14-15.

Galán M., 2009, “The Use of Polymer in the Banknotes of Banco de Mexico,” *Billetaria - International Review on Cash Management*, Banco de Espana, (3)5 : 6-7.

Galán J.E. and Sarmineto M., 2007, “Banknote Printing at Modern Central Banking : Trends, Costs, and Efficiency,” Borradores de Economíca No. 476, Banco de la Republica, Colombia.

Garcia Swartz D.D., Hahn R.W. and Layne-Farrar A., 2006, “The Move Toward a Cashless Society : A Closer Look at Payment Instrument Economics,” *Review of Network Economics*, 5(2) : 175-198.

Guibourg G. and Segendorff B., 2007, “A Note on the Price and Cost Structure of

Retail Payment Services in the Swedish Banking Sector 2002," *Journal of Banking and Finance*, 31 : 2817-2827.

Judson R. and Porter R., 2003, "Estimating the Worldwide Volume of Counterfeit U.S. Currency : Data and Extrapolation,"*Finance and Economics Discussion* Series 2003-52, Board of Governors of the Federal Reserve System (U.S.).

Judson R. and Porter R., 2010, "Estimating the Volume of Counterfeit U.S. Currency in Circulation Worldwide : Data and Extrapolation," forthcoming in Bliss, R. and G. Kaufman (eds.), "The Financial Crisis - An Early Retrospective," *Financial Institutions and Markets*, Vol. 3, Palgrave Macmillan, New York, 2010.

Kiser E., Borzekowski R. and Ahmed S., 2008, "Consumers' Use of Debit Cards : Patterns, Preferences, and Price Response," *Journal of Money, Credit, and Banking*, 40 : 149-172.

Lang B., 2001, "A Central Bank Cash Wholesaling Model," speech,
<http ://www.noteprinting.com/report_012.html> (last visit on 28/09/2010).

Lang B., 2002, "Polymer Bank Notes - the New Zealand experience," *Bulletin*, Reserve Bank of New Zealand, (65)1 : 53-57.

Langwasser K., 2010, "Recent Trends and Developments in Currency - 2009," *Bulletin*, Reserve Bank of New Zealand, (73)1 : 58-65.

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

Mulomba M., 2007, "Bank of Zambia Experience with Polymer Banknotes," presentation to the 2007 Currency Conference, Bangkok, Thailand, May 6-9
<http ://www.securency.com.au/en/presentations> (last visit on 28/09/2010).

Neumann M.J.M., 1992, "Seigniorage in the United States : How Much Does the U.S. Government Make from Money Production ?," *Review*, Federal Reserve Bank of St. Louis, (74)2 : 29-40.

Nguyen C.T., 2009, "Polymer Banknotes in Vietnam," *Billetaria - International Review on Cash Management*, Banco de Espana, (3)5 : 17.

PE Americas and Tryskele, 2011, “Life Cycle Assessment of Canada’s Polymer Bank Notes and Cotton-Paper Bank Notes - Final report,” report prepared for the Bank of Canada, May 27 (http ://www.bankofcanada.ca/wp-content/uploads/2011/06/Life-Cycle-Assessment-of-Polymer-and-Cotton-Paper-Bank-Notes_opt.pdf).

Rankin R., 2009, “Australia’s Experience with Polymer Banknotes,” *Billetaria - International Review on Cash Management*, Banco de Espana, (3)5 : 8-9.

Rogoff K.S., 1998, “Blessing or Curse ? Foreign and Underground Demand for Euro Notes,” *Economic Policy*, 13(26) : 263-303.

Sidney J., 2009, “Results of the Circulation Trial of a R$10 Polymer Note Commemorating the Fifth Centenary of the Discovery of Brazil,” *Billetaria - International Review on Cash Management*, Banco de Espana, (3)5 : 10-11.

Spencer C., 2011, “Paying with Polymer : Developing Canada’s New Bank Notes,” *Bank of Canada Review*, Supplement, June 20, 1-7.

Subbarao D., 2010, “Some Issues in Currency Management,” speech at the Foundation Stone laying function for the Bank Note Paper Mill at Mysore, India, March 22.

Sumner S., 1993, “Privatizing the Mint,” *Journal of Money, Credit, and Banking*, 25 : 13-29.

Van Hove L. and Vuchelen J., 1996, “Who Needs High-Denomination Euro Banknotes ? A Note on the Proposed Denominational Structure of the Euro”, *Rivista Internazionale di Scienze Economiche e Commerciali*, 43(4) : 791-803.

Wilson Kamit, L., s.a., “The Polymer Experience in Papua New Guinea,” speech <http ://www.polymernotes.org/other_country/PNG_experience.ht- -m> (last visit on 28/09/2010).

Wynne M., 1997, “The Economics of One Dollar,” *Southwest Economy*, Federal Reserve Bank of Dallas, issue 4 : 1-5.

CHAPITRE 6

EFFICIENT PAYMENTS : HOW MUCH DOES IT COST FOR THE CENTRAL BANK ?

Bouhdaoui Y.[1] and Bounie D.[2]

6.1 Introduction

In recent years, abundant research has been devoted to the study of denominational structures of currency systems (Caianiello *et al.* (1982), Sumner (1993), Telser (1995), Van Hove and Heyndels (1996), Tschoegl (1997), Wynne (1997), Van Hove (2001), Lee *et al.* (2005), Franses and Kippers (2007), etc.). Among the multiple properties of a currency system, the principle of least effort (PLE) is considered the most important.[3] This principle states that the settlement of cash transactions should involve as few coins and notes as possible.

The preeminence of this principle, supported by many economists such as Boeschoten and Fase (1989), Eriksson and Kokkola (1993), Abrams (1995), Pedersen and Wagener (1996), Van Hove and Heyndels (1996) and Van Hove (2001), is justified by two main arguments. First, the principle states that it is more convenient for transactors given that it reduces the bulk and weight carried around by the cash-using public in turn limiting handling costs. Second, it keeps down the number of coins and notes in circulation and thus, so the reasoning goes, the production costs incurred by the central bank. Therefore, it is preferable from a social cost viewpoint to opt for a currency system that limits the number of coins and notes used in transactions.

In this article, we demonstrate that the second argument is biased and that the efficient payments increase the production costs incurred by the central bank. As a result,

[1]Telecom ParisTech, Department of Economics and Social Sciences.
[2]Telecom ParisTech, Department of Economics and Social Sciences.
[3]Caianiello *et al.* (1982) point out two other elements, namely the surveyability (variety of denominations) and the compatibility with the decimal system.

the private benefits of transactors can be undermined by the private costs of central banks and, thus, *ceteris paribus*, efficient payments do not necessarily reduce the social cost of cash payments. To demonstrate this, we proceed in three stages.

Firstly, we propose a general framework that links the production costs of the central bank to the usage cost of cash transactions. Secondly, we compare the costs of cash transactions using the principle of least effort and a hypothetical cost-minimizing payment behavior named the "Principle of least cost" (PLC). This latter minimizes the costs of cash transactions without considering the number of tokens exchanged in transactions ; it is only used to identify inefficient payments from the point of view of the principle of least effort. Thirdly, we perform simulations on a set of cash transactions using production cost data for the U.S. currency system for the year 2010. The simulation results show that while the number of notes and coins used in transactions is certainly efficient (minimum) with the principle of least effort, the costs of cash transactions are on average 24.2% greater than those obtained with the principle of least cost. Hence, while the PLE keeps down the total number of coins and notes in circulation it can also contribute to an increase in the costs of cash transactions and thus in the production costs of denominations incurred by the central bank. We precisely estimate the increase in the annual cost incurred by the Federal Reserve to $156 million.

The remainder of the paper is structured as follows. In Section 2, we present the general framework and the payment behavior models. In Section 3, we describe the data used to perform simulations and comment on the results obtained. Finally, in Section 4, we discuss the implications of our results and some extensions.

6.2 Model

In this section, we first present a general framework that links the usage cost of cash in transactions to the production costs of denominations incurred by the central bank. Next, we present two payment behaviors with cash, namely the principle of least effort and the principle of least cost. Finally, we describe our comparison approach.

6.2.1 General framework

In a given economy, let set $\mathscr{D}$ be a distribution composed of N_T cash transactions. The distribution $\mathscr{D}$ represents the cash transactions made by the public during an observation period of one year. To carry out the transaction, the agents are using a currency system composed of J tokens of face values $v(j)$ with $j \in [\![1, J]\!]$. The payment behavior with cash represents the criteria that determine the way in which consumers and merchants use denominations in transactions. Those criteria can lead to several possible combinations to carry out a transaction. Let set k be one of the $K(x)$ solutions to pay an amount x. Following to this combination the amount is paid by exchanging $n_k(x, j)$ token(s) for each denomination j such that :

$$x = \sum_j n_k(x, j) \cdot v(j). \tag{6.1}$$

The integer $n_k(x, j)$ is set positive when the money is given by the consumer to the merchant and negative when it is a return of change.

Considering all the possible combinations, the average number of times the denomination j is involved in a transaction x is denoted by $a(x, j)$ and called the frequency of use of the denomination j for the amount x :

$$a(x, j) = \frac{1}{K(x)} \cdot \sum_k |n_k(x, j)| \,. \tag{6.2}$$

Finally, we measure the average frequency of use of a denomination j over the distribution $\mathscr{D}$:

$$a(j) = \frac{1}{N_T} \cdot \sum_{x \in \mathscr{D}} a(x, j). \tag{6.3}$$

Turning to the cost of currency, let $c_p(j)$ denotes the unit production cost of denomination j including all the costs and expenses for producing, marketing and distributing coins and notes. By combining the unit production cost of a denomination and its life span $d(j)$, we obtain the annual unit production cost, $c_a(j)$, that captures the annual cost of maintaining each circulating denomination j :

$$c_a(j) = \frac{c_p(j)}{d(j)}. \tag{6.4}$$

In addition, each circulating token j is involved in several transactions per year. Let $q_a(j)$ denotes the average number of times per year a circulating token j is involved in transactions. Hence, the usage cost of a token, *i.e.* the cost of a single use, corresponds to the ratio of the annual unit production cost and the annual velocity of circulation :

$$c_u(j) = \frac{c_a(j)}{q_a(j)}. \tag{6.5}$$

Formally, we can write the annual velocity of circulation as the ratio of the number of uses per year of all the circulating tokens j and their circulating volume, $N_c(j)$:

$$q_a(j) = \frac{N_T(j) \cdot a(j)}{N_c(j)}. \tag{6.6}$$

Therefore, replacing (6.4) and (6.6) in (6.5), we obtain :

$$c_u(j) = \frac{c_p(j) \cdot N_c(j)}{d(j) \cdot a(j) \cdot N_T}. \tag{6.7}$$

The central bank is generally responsible of processing the currency in circulation. During this operation, the substandard tokens are withdrawn and replaced by new ones. The volume, $N_r(j)$, of tokens of denomination j replaced each year is determined by $N_c(j)$ and $d(j)$ as follows :

$$N_r(j) = \frac{N_c(j)}{d(j)}. \tag{6.8}$$

Therefore, replacing in (6.7), one can write :

$$c_u(j) = \frac{c_p(j) \cdot N_r(j)}{a(j) \cdot N_T}. \tag{6.9}$$

Finally, we measure the cost of a cash transaction by aggregating the usage costs of denominations exchanged during the payment. The cost of a transaction x is written as a function of the frequencies of use of denominations as follows :

$$c_u(x) = \sum_j a(x,j) \cdot c_u(j). \tag{6.10}$$

We show that this cost directly impact the annual production cost of currency, C_r, incurred by the central bank. By definition, we have :

$$C_r = \sum_j N_r(j) \cdot c_p(j). \tag{6.11}$$

Therefore, rearranging (6.9) and replacing in (6.11), one can write :

$$C_r = N_T \cdot \sum_j a(j) \cdot c_u(j). \tag{6.12}$$

Thus, replacing (6.3) then (6.10) in (6.12), we obtain :

$$C_r = \sum_x c_u(x). \tag{6.13}$$

This equation shows that some combinations of denominations reduce the cost of cash transactions and, thus, the production costs of currency.

6.2.2 Models of cash payment behavior

In this section, we formalize two models of cash payment behavior. The first is the "Principle of least effort" that we extend to account for the usage costs of denominations ; and the second one is a hypothetical cost-minimizing model called the "Principle of least cost" that we use as a benchmark to assess the cost efficiency of the Principle of least effort.

The principle of least effort

The principle of least effort (PLE) was introduced by Caianiello *et al.* (1982) and subsequently refined by Cramer (1983). Following this principle, a consumer and a merchant use the minimum number of coins and notes to pay a given amount of cash. More formally, an amount x is paid efficiently by exchanging $n^{PLE}(x,j)_j$ tokens of each denomination j :

$$x = \sum_j n^{PLE}(x,j) \cdot v(j),$$

such that the number of coins and notes exchanged $n^{PLE}(x)$ is minimum, with :

$$n^{PLE}(x) = \sum_j \left| n^{PLE}(x,j) \right|. \tag{6.14}$$

Absolute values in Equation (6.14) indicate that overpayment and the return of change are allowed.

According to this model of payment behavior, certain transactions admit several possible combinations of denominations. The cost of the transaction x following the combination indexed by k is written :

$$c_u^{PLE}(x)_k = \sum_j \left| n_k^{PLE}(x,j) \right| \cdot c_u^{PLE}(j). \tag{6.15}$$

It is worth noting that the payment combinations can lead to different costs of transaction. Being not equivalent from a cost perspective, we denote respectively by $c_u^{PLE}(x)_{min}$ and $c_u^{PLE}(x)_{max}$ the minimum and the maximum costs for the transaction x. Likewise, the average cost of a cash transaction can be written as follows (cf. Equation (6.10)) :

$$c_u^{PLE}(x) = \sum_j a^{PLE}(x,j) \cdot c_u^{PLE}(j). \tag{6.16}$$

The principle of least cost

By analogy to the principle of least effort, we propose an alternative model called the "Principle of least effort" (PLC). This model disregards the number of tokens used in the transaction and combines a number of coins and notes that minimize the cost of a cash transaction. The PLC model constitutes thus what would be optimal from the central bank's viewpoint in terms of reducing the costs of cash transactions and consequently its production costs.

More formally, according to the PLC an amount x is paid by exchanging $n^{PLC}(x, j)$ token(s) of each denomination j :

$$x = \sum_j n^{PLC}(x, j) \cdot v(j),$$

such that the cost of the cash transaction, $c_u^{PLC}(x)$, is minimum :

$$c_u^{PLC}(x) = \sum_j \left| n^{PLC}(x, j) \right| \cdot c_u^{PLC}(j).$$

It is worth noting that several optimal combinations are also possible with the PLC. However, on the contrary to the PLE, they all have the same cost by definition.

6.2.3 Comparing the PLE and PLC Costs of Cash Payments

This part aims at comparing the costs of cash payments obtained with the PLE and the PLC models. For that, we need to measure the usage costs of denominations in both models. We assume that the annual velocities of circulation of denominations are exogenous and not influenced by the cash payment behavior. Given this assumption, we can write :

$$q_a^{PLE}(j) = q_a^{PLC}(j). \tag{6.17}$$

Turning to the life span of the circulating currency, we consider that the lifespan of tokens is mainly influenced by the resistance of the manufacturing technology of currency and by the intensity of use. When comparing the PLE to the PLC, the first factor does not play since we keep the same manufacturing technology as well as the second factor since the annual velocity of circulation, that captures the intensity of use, is assumed to be invariant. Then, we deduce that the cash payment behavior does not affect the life span of a circulating denomination j :

$$d^{PLE}(j) = d^{PLC}(j). \tag{6.18}$$

Finally, the usage cost of a denomination j is written as :

$$c_u(j) = \frac{c_p(j)}{d(j) \cdot q_a(j)}.$$

Therefore, from (6.17) and (6.18), we obtain the same usage cost of denominations for the PLE and the PLC :

$$c_u^{PLE}(j) = c_u^{PLC}(j). \tag{6.19}$$

This, however, does not mean that the PLE and the PLC will generate identical costs of transaction since they opt for different combinations according to distinct criteria. The resulting costs of transactions will therefore be different, the PLC combinations being by definition less costly if not equal to those of the PLE.

Hereafter, given (6.19), we no longer mention the superscript related to the payment behavior and simply denote the usage cost of denomination j by $c_u(j)$.

6.3 Simulation : the U.S. currency system

Based on the production volumes and costs data provided by the Federal Reserve, we measure the usage costs of denominations. In so doing, we simulate a series of payments with the U.S. currency system following the PLE and then the PLC. Finally, we compare the costs of transactions generated by the models and aggregate them to provide an estimate of the total cost of paying efficiently.

6.3.1 The usage cost of U.S. denominations

According to Equation (6.9), measuring the usage cost of denominations requires to gather the unit production costs, the annual replacement volumes, the average frequencies of use and an estimation of the number of cash transactions performed in one year.

On the production cost side, the United States Mint and the annual "New Currency Budgets" of the Federal Reserve Board are publishing annual reports that include the unit production costs of currency.[4] We observe that some denominations are produced in different series such as the 100$ banknote with different unit production costs. For instance, in 2008, the 100$ banknote was produced either in Series-1996 at a unit production cost of 0.079$ and for a total number of units of about 854.4 millions or in Series-2004 at 0.125$ for a total number of 817.7. We get around this issue by considering the weighted average production cost which, in the case of the 100$ denomination, amounts to 0.1015$. Table 6.I summarizes the average unit production costs per denomination measured for the year 2010.

Next, the Federal Reserve kindly provided us with detailed information on the volume of banknotes replaced in year 2010. For the coins, we rely on the annual production volume provided by the Mint for the year 2010 assuming that all coins produced this year are put into circulation. The volumes obtained are summarized in Table 6.I.[5]

Turning to the annual number of cash transactions, the data of such figure in the U.S. are very scarce. The study of Gerdes *et al.* (2005) is, to the best of our knowledge, the only study that provides such an estimation. According to the authors, the total number of cash transactions in the U.S. amounts to approximately 100 billion. We use this reference as a rough approximation.

Finally, the remaining data required to calculate the usage cost of denominations are the average frequencies of use $a(j)$ (cf. Equation (6.9)). In this context, there is, first, a lack of precise data on the payment behavior of the public. To tackle this issue, we rely

[4] Accrding to our model, the costs we are considering include the costs and expenses for producing, marketing, and distributing circulating denominations.

[5] We do not include the 50¢denomination and the 1$ coin because they do not circulate widely.

Denomination	Unit production cost (in $)	Annual Replacement volume (in millions)
$v(j)$	$c_p(j)$	$N_r(j)$
1¢	0.0179	4010.83
5¢	0.0922	490.56
10¢	0.0569	1,119.00
25¢	0.1278	347.00
1$	0.0481	2,622.04
5$	0.0776	673.01
10$	0.0789	484.26
20$	0.0851	1270.22
50$	0.0851	104.92
100$	0.1100	785.54

Tableau 6.I – The unit production costs and the annual replacement volumes of U.S. denominations in the 2010.

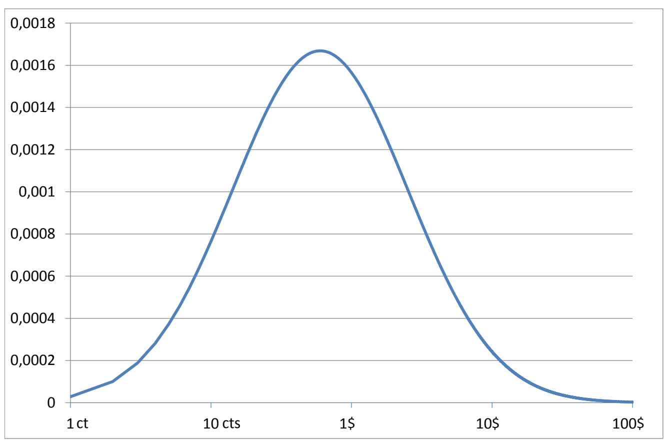

Figure 6.1 – Log-normal distribution of cash transactions : occurrence frequency as a function of the amounts. (Logarithmic scale for the x-axis)

on the survey of Franses and Kippers (2006) and consider that the PLE constitutes a reasonable approximation of this payment behavior. Second, we approximate the pattern of cash transactions in the economy with a log-normal distribution calibrated to the average value of the cash transactions observed in the U.S.[6] that amounts to 11.52$ according

[6]Boeschoten (1989) shows that the log-normal distribution fits better to the pattern of the observed cash transactions.

to Garcia-Swartz *et al.* (2006). Figure 6.1, illustrates the distribution considered with a logarithmic scale on the x-axis. We limit our distribution to the payments below 100$ since the daily cash transactions are generally below this threshold. We simulate then the PLE on this series of payments and calculate the average frequencies of use. Finally, we deduce the usage costs of denominations (cf. Table 6.II).

Denomination	Average Frequency of use with the PLE	Usage cost (in $)
$v(j)$	$a^{PLE}(j)$	$c_u(j)$
1¢	1.24	0.00058
5¢	0.36	0.00124
10¢	0.54	0.00117
25¢	0.86	0.00051
1$	1.15	0.00110
5$	0.38	0.00139
10$	0.20	0.00194
20$	0.17	0.00648
50$	0.06	0.00143
100$	0.02	0.03704

Tableau 6.II – The average frequencies of use according to the PLE and an estimate of the usage costs of U.S. denominations in 2010.

6.3.2 Results

First, we comment on the results of the PLE payments. We observe that the PLE generates more than one efficient combination for 48.8% of the transactions of the distribution. Indeed, several amounts can be paid in multiple efficient ways. For instance, the amount 26.20$ is payable with two efficient combinations both exchanging five tokens, namely 20$ + 5$ + 1$ + 25¢ - 5¢ and 20$ + 5$ + 1$ + (2 * 10¢). The efficient combinations are not equivalent from a cost perspective because the usage costs of denominations are not homogeneous (cf. Table 6.II) and some PLE combinations increase the cost of cash transactions by using costly denominations. For instance, the costs of the two combinations mentioned in the example above are respectively 0.0107$ and 0.0113$. Over the distribution studied, the average gap between the most and the least costly PLE combinations, $(c_u^{PLE}(x)_{max} - c_u^{PLE}(x)_{min})$, amounts to 0.0016$ per transaction.

Second, we simulate the same pattern of cash payments following this time the PLC and compare the results obtained with those of the PLE.

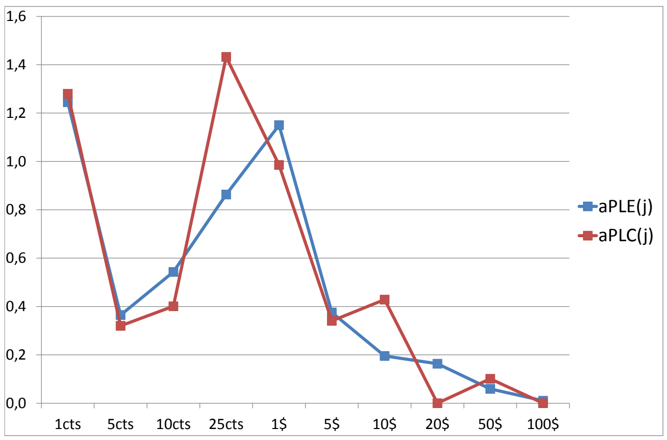

Figure 6.2 – Average frequencies of use of denominations with the PLE and the PLC.

Moreover, calculating the frequencies of use of the denominations in the case of the principle of least cost (Figure 6.2), we find that some of the average frequencies equal zero (20$ and 100$). This finding is due to the trade-off operated by the principle of least cost. As an illustration, the PLC prefers to use 2*10$ instead of 20$ banknote although it involves an additional token because it reduces the usage cost ($c_u(20\$) > 2 * c_u(10\$)$). Overall, compared to the PLC, the PLE reduces the use of tokens in transactions by 6.1%, but costs in turn almost a quarter more(cf. Table 6.III).

	PLE	PLC	Gap
Average number of tokens exchanged per transaction	4.97	5.29	6.1%
Average cost per transaction	0.0065$	0.0049$	-24.2%

Tableau 6.III – The payment efficiency and the cost of transactions with the PLE and the PLC.

Going into details on the cost gap between the PLE and the PLC, it turns out that 61.7% of the PLE payments are not cost efficient. Indeed, as commented on, when comparing the frequencies of use, the combinations that minimize the cost include more tokens than the optimum. As a consequence, they are not considered by the PLE. If we build on the previous example related to the amount 26.20$, the PLC generates the following combination (2 * 10$) + 5$ + 1$ + 25¢ - 5¢ that involves six tokens and costs about 0.0081$.

Finally, the impact of the payment behavior on the production costs of currency incurred by the central bank can be measured using Equation (6.12). We find that efficient cash payments induce an annual cost of about 156 million $ for the Federal Reserve.

6.4 Discussion

The studies related to the optimal structure of a currency system consider that efficient payments, *i.e.* following the principle of least effort, reduce the production costs incurred by the central bank because they minimize the number of coins and notes used in transactions. Our results show however that they increase the production costs at least for the U.S. currency system. The main reason is that the usage costs of denominations do not have a regular pattern in the sense that some denominations have a significantly high usage cost that makes them interchangeable by less costly combinations.

An extension of this work would be to improve the production cost function of the central bank in order to provide a more accurate estimate of the cost of efficient payments. In the same line as Bouhdaoui *et al.* (2011), we could separate the fixed and the variable costs to account for the economies of scale on the production volumes.

As a final point, when we broaden the analysis to the social cost of cash, we note that paying efficiently can have a positive effect on the duration of the transactions, the processing costs of the merchants, the carrying costs of money and/or also on the computational effort to determine the efficient combinations. All these effects lead to reduce the social cost. Therefore, the conclusions of this study remain restricted to the production costs of currency incurred by the central bank and cannot be generalized to the social cost.

At the same time, it raises a larger issue for monetary authority : making a denominational mix more convenient for transactors while reducing the resulting production costs. This question of the cost efficiency of a denominational structure has gone largely unresearched (Massoud, 2005) even though recent studies confirm that the social cost of cash payments amounts to several billion dollars or euros (Garcia Swartz *et al.*, 2006 ; Guibourg and Segendorff, 2007).

6.5 Literature cited

Abrams R.K., 1995, "The Design and Printing of Bank Notes : Considerations When Introducing a New Currency," Working Paper 95-26, International Monetary Fund, March.

Boeschoten W.C. and Fase M.M.G., 1989, "The Way We Pay with Money," *Journal of Business and Economic Statistics*, 7 : 319-26.

Bouhdaoui Y., Bounie D. and Van Hove L., 2011, "Central banks and their banknote series : The efficiency-cost trade-off," *Economic Modelling*, 28(4) : 1482-1488.

Caianiello E.R., Scarpetta G. and Simoncelli G., 1982, "A Systemic Study of Monetary Systems," *International Journal of General Systems*, 8 : 81-92.

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

Eriksson S. and Kokkola T., 1993, "Cash," In Payment and Settlement Systems in Finland, Publications of the Bank of Finland, Series A, No. 88 : 21-57.

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

Garcia Swartz D.D., Hahn R.W. and Layne-Farrar A., 2006, "The Move Toward a Cashless Society : A Closer Look at Payment Instrument Economics," *Review of Network Economics*, 5(2) : 175-198.

Gerdes G.R., Walton II J.K., Liu M.X. and Parke D.W., 2005, "Trends in the Use of Payment Instruments in the United States," Federal Reserve Bulletin, Board of Governors of the Federal Reserve System (U.S.) : 180-201.

Guibourg G. and Segendorff B., 2007, "A Note on the Price and Cost Structure of Retail Payment Services in the Swedish Banking Sector 2002," *Journal of Banking and Finance*, 31 : 2817-2827.

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

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

Pedersen E.H. and Wagener T., 1996, "Circulation of Notes and Coins in Denmark," Danmarks Nationalbank Monetary Review, 35 : 29-45.

Sumner S., 1993, "Privatizing the Mint," *Journal of Money, Credit and Banking*, 25 : 13-29.

Telser L.G., 1995, "Optimal Denominations for Coins and Currency," *Economics Letters*, 49 : 425-427.

Tschoegl A.E., 1997, "The Optimal Denomination of Currency : A Conjecture," *Journal of Money, Credit and Banking*, 29 : 546-554.

Van Hove L. and Heyndels B., 1996, "On the Optimal Spacing of Currency Denominations," *European Journal of Operational Research*, 90 : 547-552.

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 : 1015-1021.

Wynne M.A., 1997, "More on Optimal Denominations for Coins and Currency," *Economics Letters*, 55 : 221-225.

CHAPITRE 7

THE SOCIAL COST OF DENOMINATION STRUCTURES

Bouhdaoui Y.[1]

7.1 Introduction

In the last years, a growing number of empirical studies focused on the social cost of payment instruments (Bergman et *al.*, 2007 ; Garcia Swartz, Hahn and Layne-Farrar, 2006a, 2006b). All these studies confirm that cash is the most costly payment instrument and that the society will benefit to foster cheaper alternative electronic payment instruments. Surprisingly however, there is little theoretical research that focuses on the costs of cash payments even though there is a wide agreement that they affect the social welfare. For instance Lee and Wallace (2006) have shown that the maintenance cost of money and the initial production cost are crucial to explain the optimal divisibility of a *single* denomination currency system. Likewise, Lee et *al.* (2005) have introduced the first general equilibrium model of denomination structures that endogenizes the carrying cost of money. This model have been used by Lee (2010) in a context of monetary expansion to discuss the real effects of denomination structures.

The main objective of this paper is precisely to include all the costs of money separately examined in the aforementioned studies to measure and compare the social costs of multiple denomination currency systems. In so doing, we set up a search-based model build on Lee et *al.* (2005) that includes the carrying cost of money and we incorporate three further costs financed by lump-sum taxes : a maintenance cost, a cost of providing the initial stock of money (Lee and Wallace, 2006), and a fixed cost per denomination as suggested by Lee et *al.* (2005).

We start with a description of the equilibrium for a version of the model of Lee et *al.*

[1]Telecom ParisTech, Economics and Social Sciences, 46 rue Barrault, 75634, Paris Cedex 13 ; tel : +33 1 45 81 75 68 ; e-mail : yassine.bouhdaoui@telecom-paristech.fr

(2005) without lotteries. We enumerate the different steps to compute the steady state, we formally define the different components of the costs of money and we aggregate them to calculate the ex-ante welfare of an agent. Using a parameterized version of our model, we express the gap of the ex-ante welfare between two specific denomination structures as a function of the model parameters. As an application, we compute the steady states for the powers-of-two and the powers-of-three denomination structures.[2] The study of these two specific denomination structures is of interest as Van Hove and Heyndels (1996) and Van Hove (2001) have shown that a powers-of-two structure can result in a lower average number of tokens used in transactions. In other words, lowering the spacing factor between denominations such as in the case of a powers-of-two structure is supposed to improve the efficiency of payments (compared to a powers-of-three one).

In a specific set of parameters, we show that the social cost of a powers-of-two-based structure is higher than that of a powers-of-three when the fixed cost per denomination exceeds a threshold value. In other words, when we account for different costs of money, a powers-of-two denomination structure which implies a lower spacing factor between denominations can be detrimental for the society. Several intuitions explain this result. To begin with, a powers-of-two denomination structure implies more denominations and thus higher fixed cost. Likewise, the number of tokens exchanged between agents is surprisingly slightly higher when they use a powers-of-two denomination structure resulting in higher costs of maintenance. As a result, even though we find a drop in the carrying and the initial productions costs when people use of a powers-of-two denomination structure, the simulation results confirm that a powers-of-three denomination structure outperforms a powers-of-two one.

The remainder of the paper is structured as follows. In Section 2, we describe the equilibrium, the computation steps and the different cost items, then we express the ex-ante welfare. In Section 3, we calibrate our model, compute the steady states with

[2]Powers-of-two and powers-of-three systems refer to modular currency systems with the modules 2 and 3 respectively. In the modular denomination structures, each denomination is X times the one below it, X called the 'module' and being a constant. They have been largely studies in the theoretical literature (cf. Caianiello et *al.* (1982), Van Hove and Heyndels (1996), Telser (1995) and others).

the powers-of-two and the powers-of-three denomination systems, then we discuss the results. In Section 4, we conclude.

7.2 The model

7.2.1 Environment

We introduce the matching model studied by Zhu (2003) with divisible goods, indivisible money and a large upper bound on the agents money holdings. Time is discrete, there are N distinct produced and perishable types of divisible goods and a $[0,1]$ continuum of each $N \geq 3$ types of infinitely lived agents. Given $n \in \{1,2,...,N\}$, a type n agent produces only type n good and consumes only type $n+1$ (modulo N). Each agent maximizes expected discounted utility with a discount factor $\beta \in (0,1)$.

We define a denomination structure (that we call currency system hereafter) $s = \{s(1), s(2), ..., s(J)\}$, where $s(k)$ is the face value of the kth denomination. We assume that $s(k+1) > s(k)$, with $s(k)/s(1)$ an integer and $1/s_1$ a positive integer. Let the upper bound on individual holdings be denoted by B and the average money holding by $\bar{m}$. The set of possible individual holdings is written $B_s = \{s(1), 2s(1), ..., B\}$. We let π_t denote a probability measure on B_s, such that $\pi_t(z)$ refers to the fraction of type n with wealth z at the beginning of period t. Agents hold portfolios that are denoted by $y = \{y(1), y(2), ..., y(J)\}$, where $y(k) \geq 0$ is an integer that refers to the quantity of the kth denomination held. We let $\mathbf{Y} = \{y \in \mathbb{Z}_+^J : y.s \leq B\}$ be the set of feasible individual portfolios and θ_t be a probability measure over $\mathbf{Y}$ where $\theta_t(y)$ is the fraction of each type with a portfolio y.

The period sequencing is that of Lee *et al.* (2005). In accordance with their wealth, agents choose a portfolio at the beginning of the period. Then each person meets randomly another person. Since $N \geq 3$, the only relevant kind of meetings is the single-coincidence meeting where a type n agent (the buyer) makes a take-it-of-leave-it offer to a type $n+1$ one (the seller). This offer takes into account the portfolio composition of the seller and can therefore involve a return of change.

The realized utility for a type n agent that has chosen a portfolio y is $u(q_{n+1}) - q_n - \gamma.sum(y)$, where q_n and q_{n+1} refer respectively to the production of good n and the consumption of good $n+1$ (both $\in \mathbb{R}^+$), γ the utility cost of carrying a token and $sum(y) = \sum_k y(k)$ is the number of monetary items in the portfolio. The utility function $u : \mathbb{R}^+ \to \mathbb{R}$ is strictly increasing and concave, continuously differentiable, and satisfies $u(0) = 0$, $u'(\infty) = 0$ and $u'(0)$ sufficiently large.

7.2.2 Equilibrium

In this part, we present a non lottery version of the equilibrium.

Given two portfolios y_c and y_p, as a reference to a consumer's and producer's portfolios, we denote by $T(y_c, y_p)$ the set of positive feasible transfers from y_c to y_p :

$$T(y_c,y_p) = \{(v_c - v_p).s \in [0, B - y_p.s] : \forall k,\ \ 0 \leq v_c(k) \leq y_c(k), 0 \leq v_p(k) \leq y_p(k)\}. \tag{7.1}$$

For a given w_{t+1}, the consumer problem is :

$$\max_{v \in T(y_c,y_p)} \{u(q) + \beta w_{t+1}((y_c - v).s)\} \tag{7.2}$$

s.t.

$$\beta w_{t+1}(y_p.s) \leq \beta w_{t+1}((y_p + v).s) - q.$$

Since w_{t+1} is strictly increasing, the constraint is satisfied at equality. Therefore the consumer problem is written :

$$\max_{v \in T(y_c,y_p)} \{u(\beta w_{t+1}((y_p + v).s) - \beta w_{t+1}(y_p.s)) + \beta w_{t+1}((y_c - v).s)\}. \tag{7.3}$$

The optimized value of (7.3) is denoted by $f(y_c, y_p, w_{t+1})$ and the set of optimum wealth transfer by $optT(y_c, y_p, w_{t+1})$.

We consider a probability measure $\lambda(\cdot, y_c, y_p, w_{t+1})$ on $\mathbf{Y}$ that satisfies :

$$\lambda(y_c', y_c, y_p, w_{t+1}) = 0 \ \text{ if } \ (y_c - y_c') \notin optT(y_c, y_p, w_{t+1}). \tag{7.4}$$

$\lambda(y'_c, y_c, y_p, w_{t+1})$ can be interpreted as the probability that a consumer with a portfolio y_c ends up with a portfolio y'_c, after a meeting with a consumer with portfolio y_p. Thus, the wealth distribution at $t+1$ is written :

$$\begin{aligned}\pi_{t+1}(z) = \frac{N-2}{N}.\pi_t(z) + \frac{1}{N}. \sum_{(y_c,y_p)\in\mathbf{Y}^2} \theta_t(y_c).\theta_t(y_p). \sum_{y\in\mathbf{Y}:\ y.s=z} \lambda(y, y_c, y_p, w_{t+1}) \\ + \frac{1}{N}. \sum_{(y_c,y_p)\in\mathbf{Y}^2} \theta_t(y_c).\theta_t(y_p). \sum_{y\in\mathbf{Y}:\ y.s=z} \lambda(y_c + y_p - y, y_c, y_p, w_{t+1})\end{aligned} \tag{7.5}$$

The agents who were in no coincidence meeting are represented by the first term. The second and the third terms respectively represent the fraction of the consumers and the producers that end up with a portfolio of wealth z. The sum of the three terms provide the fraction of agents with wealth z at $t+1$.

On the other hand, we can deduce the expected value from holding a portfolio y :

$$h_t(y, w_{t+1}) = -\gamma.sum(y) + \frac{N-1}{N}.\beta.w_{t+1}(y.s) + \frac{1}{N}. \sum_{y_p\in\mathbf{Y}} \theta_t(y_p).f(y, y_p, w_{t+1}). \tag{7.6}$$

Then the expected value from detaining a wealth z :

$$w_t(z) = \max_{y\in\mathbf{Y}:\ y.s=z}\{h_t(y, w_{t+1})\}. \tag{7.7}$$

For each wealth z, we denote by $optY(z, w_{t+1})$ the set maximizing portfolios. Finally, given a probability measure $\phi(\cdot, z, w_{t+1})$ on $\mathbf{Y}$ satisfying :

$$\phi(y, z, w_{t+1}) = 0 \text{ if } y \notin optY(z, w_{t+1}), \tag{7.8}$$

we deduce the portfolios distribution at $t+1$:

$$\theta_{t+1}(y) = \sum_{z\in B_s} \pi_{t+1}(z).\phi(y, z, w_{t+1}). \tag{7.9}$$

Now the definition of equilibrium is stated.

Definition 1 *Given* π_0*, a sequence* $\{w_t, \pi_t, \theta_{t+1}\}$ *is an equilibrium if it satisfies* (7.3) *to* (7.9)*. A monetary equilibrium is an equilibrium with positive consumption and production. A collection* $\{w, \pi, \theta\}$ *is a steady state if* $\{w_t, \pi_t, \theta_{t+1}\}_{t=0}^{\infty}$ *with* $w_t = w$, $\pi_t = \pi$ *and* $\theta_{t+1} = \theta$ *for all t is an equilibrium for* $\pi_0 = \pi$.

7.2.3 Computing the model

Since analytical characterizations of the steady state do not exist, we perform numerical simulations using an initial calibration of the model. The objective is to find a fixed point for the value function, the distribution of portfolios and wealth distribution.[3]

The different steps of the algorithm are the following :

Step 1 : Set initial assumptions on s, $\bar{m}$, B, N, β and γ, and specify a functional form for the utility function.

Step 2 : Set an initial guess, w_0, for the value function, w_{t+1}, for each value on the set of possible individual holdings B_s. Next, solve (7.3) and deduce $f(y_c, y_p, w_{t+1})$ and $optT(y_c, y_p, w_{t+1})$.

Step 3 : Build a probability measure on $\mathbf{Y}$ that satisfies (7.4).[4]

Step 4 : Set an initial guess, π_0 and θ_0, for respectively the wealth distribution, π_t, and the portfolios distribution, θ_t. Using (7.5), compute the new wealth distribution, π_{t+1}.

Step 5 : Calculate the expected discounted utility after a portfolio is chosen $h(y, w_{t+1})$ (cf. (7.6)) then update the value function, w_t, (cf. (7.7)) and tcompute he set of optimal portfolios, $optY(z, w_{t+1})$.

Step 6 : Build a probability measure on $\mathbf{Y}$ that satisfies (7.8). Update the portfolios distribution, θ_{t+1}, Using ϕ and π_{t+1}, equation (7.9).[5]

---

[3]For the proof of the existence of a steady state on the lottery version, see Lee et *al.* (2005). This latter also holds for the non-lottery version.

[4]For instance, the probability measure $\lambda(\cdot, y_c, y_p, w_{t+1})$ can be uniform on the subset $optT(y_c, y_p, w_{t+1})$.

[5]As we suggested for $\lambda(\cdot, y_c, y_p, w_{t+1})$, the restriction of $\phi \cdot, z, w_{t+1}$ on the subset $optY(z, w_{t+1})$ can be uniform.

Step 7 : Compute the absolute distances $w_t - w_{t+1}$, $\pi_{t+1} - \pi_t$ and $\theta_{t+1} - \theta_t$. Break the loop if they all fall below a threshold $\varepsilon \ll 1$. Otherwise, initialize w_t, π_{t+1} and θ_{t+1} as the new guesses for the next iteration, and repeat the steps 2 to 7 until reaching convergence.

7.2.4 Cost of money

In this part, we present the per-capita resource cost needed to produce and maintain the stock of money. As in Lee and Wallace (2006), we assume that money depreciates with use. The maintenance cost, aiming at maintaining the money stock, and the initial production cost are assumed to be financed by lump-sum taxes. As a consequence, it does not affect the trades in pairwise meetings.

More formally, when a money transfer v occurs in a pairwise meeting, this latter involves the exchange of $n(v) = \sum_{k=1}^{J} \mid v(k) \mid$ tokens. We assume that the loss generated by this exchange is $\mu.n(v)$, where $\mu \in \mathbb{R}_+$ is a parameter.

In the steady state (w^s, π^s, θ^s) associated with the currency system s, let $a^s(k)$ denote the frequency of use of denomination k in transactions *i.e.* the average number of denominations $s(k)$ involved in each cash transaction. This latter can be written :

$$a^s(k) = \sum_{(y_c,y_p)\in \mathbf{Y}^2} \theta^s(y_c).\theta^s(y_p). \sum_{v\in optT(y_c,y_p,w)} \lambda(y_c - v, y_c, y_p, w). \mid v(k) \mid . \tag{7.10}$$

Therefore, one deduces the average number of tokens exchanged per transaction $\bar{n_t^s} = \sum_{k=0}^{J} a^s(k)$ and the corresponding loss $\mu.\bar{n_t^s}$.

On the other hand, we take $\frac{\bar{q^s}}{N.r}$ to be the cost of producing a monetary item. The parameter r refers to the number of monetary items that a person could produce in a period and $\bar{q^s}$ to the average output produced in a single-coincidence meeting :

$$\bar{q^s} = \sum_{(y_c,y_p)\in \mathbf{Y}^2} \theta^s(y_c).\theta^s(y_p). \sum_{v\in optT(y_c,y_p,w)} \lambda(y_c - v, y_c, y_p, w).q(v, y_p, w), \tag{7.11}$$

with $q(v, y_p, w) = \beta.(w^s((y_p + v).s) - w^s(y_p.s))$ (cf. the constraint in equation (7.2)). The idea being to compensate the agent of what he sacrifices when employed in producing money instead of being matched.[6]

Thus, the ex-ante maintenance cost associated with the steady state is :

$$c_m^s = \frac{1}{1-\beta}.\frac{\bar{q^s}}{N.r}.\frac{\mu.\bar{n_t^s}}{N}. \tag{7.12}$$

Next, we define the fixed cost per denomination as an equivalent output, q_F, the fixed cost can be written :

$$c_F^s = J^s.q_F.^7 \tag{7.13}$$

The initial production cost corresponds to the product of the average number of monetary items in the portfolio, that we denote by $\bar{n_p^s}$, and the unit production cost :

$$c_i^s = \frac{\bar{q^s}}{N.r}.\bar{n_p^s}.^8 \tag{7.14}$$

The term, $\bar{n_p^s}$, is measured as follows :

$$\bar{n_p^s} = \sum_{z \in B_s} \pi^s(z). \sum_{y \in optY(z,w)} \phi(y,z,w).sum(y), \tag{7.15}$$

where $\sum_{y \in optY(z,w)} \phi(y,z,w).sum(y)$ corresponds to the average number of items in the agent portfolios with wealth z at the beginning of the period.

Finally, aggregating all the cost items and the expected value, we obtain the ex-ante welfare of the currency system s associated with the steady state (w^s, π^s, θ^s) :

$$\mathscr{W}^s = \pi^s.w^s \; - \; (c_i^s + c_F^s + c_m^s). \tag{7.16}$$

[6] For the sake of simplicity, we assume the same labor cost for all denominations.

[7] J^s refers to the number of denominations in the currency system s.

[8] This cost is assumed by the agents only once.

7.2.5 Comparison of two currency systems

In this part, we express the gap of the ex-ante welfare between two denomination structures as a function of the parameters of the model .

Considering two currency systems s_1 and s_2, the variation of the ex-ante welfare, $\Delta\mathscr{W}^s = \mathscr{W}_2^s - \mathscr{W}_1^s$, is written :

$$\Delta\mathscr{W}^s = \Delta(\pi^s.w^s) \; - \; \Delta(c_i^s + c_F^s + c_m^s), \tag{7.17}$$

where :

$$\Delta(c_i^s) = \frac{1}{N.r} \; . \; \Delta(\bar{q^s}.\bar{n_p^s}),$$

$$\Delta(c_F^s) = q_F \; . \; \Delta(J^s),$$

and

$$\Delta(c_m^s) = \frac{1}{1-\beta} \cdot \frac{\mu}{N^2.r} \; . \; \Delta(\bar{q^s}.\bar{n_t^s}).$$

We state the condition on which the social cost of s_2 is higher than that of s_1, in other words $\Delta\mathscr{W}^s > 0$, as follows :

$$\mu \; . \; \frac{1}{1-\beta} \cdot \frac{\Delta(\bar{q^s}.\bar{n_t^s})}{N^2.r} < -q_F.\Delta J + \Delta(\pi^s.w^s) - \frac{\Delta(\bar{q^s}.\bar{n_p^s})}{N.r}. \tag{7.18}$$

Depending on the sign of $\frac{\Delta(\bar{q^s}.\bar{n_t^s})}{N^2.r}$ and ΔJ, there is a linear condition on μ and q_F that determines which of the two systems generates the highest ex-ante welfare. This condition depends also on the term $\Delta(\pi^s.w^s)$ because of the endogenization of the carrying cost of money and the term $\frac{\Delta(\bar{q^s}.\bar{n_p^s})}{N.r}$ that refers to the potential gain on the initial production cost.

7.3 Simulation with the powers-of-two and the powers-of-three currency systems

7.3.1 Calibrating the model

In this section, we calibrate the model described above before applying it to the powers-of-two and the powers-of-three denomination structures.

The size of each currency system is defined such that the highest denomination remains below the upper bound on individual holdings B. This approach implies that there is no restriction on the choice of the portfolio and *a fortiori* no restriction on the size of monetary items used to perform transactions. The calibration of the parameters used is provided in Table 7.I.

We set a low upper bound on money holdings ($B = 10$) because the computational capabilities required to reach the steady state increase exponentially with the number of feasible portfolios and this number is itself very sensitive to B. [9] As a consequence, the powers-of-two system includes only the first four denominations $\{1,2,4,8\}$ ($J^{X2} = 4$) and the first three ones $\{1,3,9\}$ ($J^{X3} = 3$) for the powers-of-three. Turning to the utility function, we take $\alpha = 0.5$, and concerning the calibration of the utility cost of carrying money, γ, we obtain a convergence to a 'nice' steady state only when this latter is very low. Instability or degenerate steady state is generated otherwise.[10] This is in line with *Proposition 1* in Lee et *al.* (2005) which states that for sufficiently large B, $u'(0)$ and $1/s_1$, there exists $\gamma_s > 0$ such that for all $\gamma \in [0, \gamma_s]$ there exists a 'nice' steady state.

7.3.2 Results

The values obtained for the main parameters with each denomination structure are presented in Table 7.II.

[9] The computation time can take many hours before knowing whether the scenario tested is reaching a 'nice' steady state. In the calibration scenario presented, we enumerate 60 feasible portfolios under the powers-of-two structure and 28 under the powers-of-three one, and almost 170 iterations were needed to reach the steady state for each of them.

[10] In the literature of the search models of money, 'nice' steady state means a steady state with w strictly increasing and with π having full support.

β	0.97
B	10
$\bar{m}$	3
N	3
γ	0.005
r	1
$u(q)$	q^{α}

Tableau 7.I – Summary of parameterizations.

The upper bound on agent money holdings has been set at $B = 10$. In the corresponding set of possible individual holdings, we count four denominations for the powers-of-two structure (denoted $X2$ hereafter) and three for the powers-of-three one (denoted $X3$). This implies a higher fixed cost, c_F^s, for the first structure.

The portfolios carry 15% more tokens on average under the $X3$ structure compared to the $X2$ one, $\bar{n_p^{X3}} = 2.03$ versus 1.76. In addition, the average output, $\bar{q^s}$, that reflects the level of the cost of labor does not vary significantly. Therefore, the initial production cost, c_i^s, is higher under the $X3$ structure (cf. Equation (7.14)).

The average expected value, $\pi^s.w^s$, observed for the $X3$ structure is lower for two reasons. First, because the carrying costs of money are higher since the agents carry more tokens on average under the $X3$ structure. And second, because we observe more frequent frictions on the transfers of wealth : the composition of the portfolios of the consumers and the producers reduces the number of possible wealth transfers in a significant number of single coincidence meetings.[11] In the appendix, we present the solution of the consumer's wealth transfer problem under the $X3$ structure.

The frictions on the transfer of wealth impact the average number of tokens exchanged in a single coincidence meeting, $\bar{n_t^s}$. In comparison with $X2$, this latter is slightly lower with the $X3$ structure : 0.96 versus 0.98. Given Equation (7.12), this results in a 4% lower ex-ante cost of maintenance, c_m^s, expressed as a function of the depreciation rate per use μ, for the $X3$ structure (cf. Table 7.II). This finding contrasts with Van

[11] For instance, when a consumer with a wealth $z = 10$, holding a portfolio $\{1,0,1\}^{X3}$ meets in a single coincidence meeting a producer with a portfolio $\{1,0,0\}^{X3}$, only three amounts can be transfered : 1, 8 (with a return of change) and 9.

Hove and Heyndels (1996) (and their references) since they consider that the $X2$ structure is supposed to have a higher payment efficiency because of a higher 'density' of denominations. Remind that these research assume an exogenous distribution of cash transactions, no portfolio restrictions and the use of the principle of least effort.[12] The counter-intuitive result that we found points out the limitations of those studies.

	Powers-of-two system ($X2$)	Powers-of-three system ($X3$)
J^s	4	3
$\pi^s.w^s$	1.56	1.53
$\bar{n}_p^s$	1.76	2.03
$\bar{n}_t^s$	0.98	0.96
$\bar{q}^s$	1.08	1.07
c_i^s	0.63	0.72
c_F^s	$4 \cdot q_F$	$3 \cdot q_F$
c_m^s	$3.93 \cdot \mu$	$3.79 \cdot \mu$

Tableau 7.II – Simulation results for the main parameters under the powers-of-two and the powers-of-three structures.

If we express numerically the condition stated in (7.18), so that the ex-ante welfare generated with the powers-of-three structure is higher than that generated with the powers-of-two one, we obtain the condition :

$$-0.14 \cdot \mu < q_F \; - \; 0.12.$$

This means that, regardless of μ, the $X3$ structure has a lower social cost provided that the fixed cost per denomination, q_F, is not below 0.12.

The Figures 7.1 and 7.2 show respectively the wealth distribution and the value function defined on money holdings. The value function is strictly increasing and concave and we note a slight but increasing gap between the two denomination structures that is positively correlated to money holdings. On the other hand, we only observe a slight distortionary effect on the wealth distribution.

[12]According to the principle of least effort, transactions are carried out by exchanging the minimum number of tokens. It has been formally defined by Cramer (1983).

The frequencies of use of denominations defined in (7.10) are presented in Table 7.III. The highest denominations - $s^{X2}(4)$ and $s^{X3}(3)$ - are not used in transactions because they are not needed : in the steady state, more than 90% of the amounts transfered do not exceed 1, and their maximum value is 4 with the $X2$ structure and 3 with the $X3$ structure.

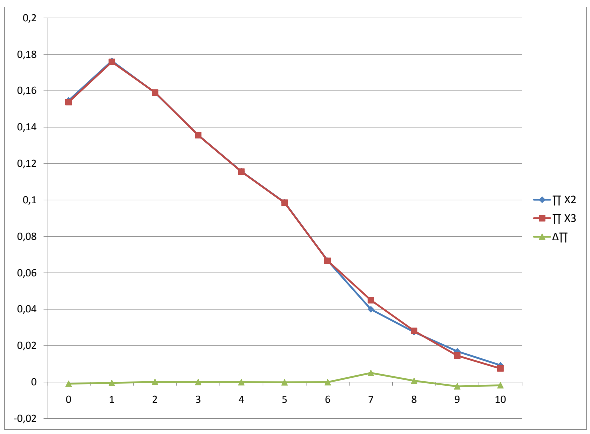

Figure 7.1 – Wealth distribution under the powers-of-two and the powers-of-three structures.

Currency System	powers-of-two system	powers-of-three system
k	$a^{X=2}(k)$	$a^{X=3}(k)$
1	0.78	0.88
2	0.17	0.08
3	0.03	0
4	0	-

Tableau 7.III – Average frequencies of use of denominations under the powers-of-two and the powers-of-three structures.

Concerning the portfolios, we can note, from tables 7.IV and 7.V, that the agents opt for a unique optimal portfolio for each value. In general, we observe that this choice does not minimize the number of tokens to be transported for the corresponding value because it reduces the number of possible wealth transfers and therefore the expected

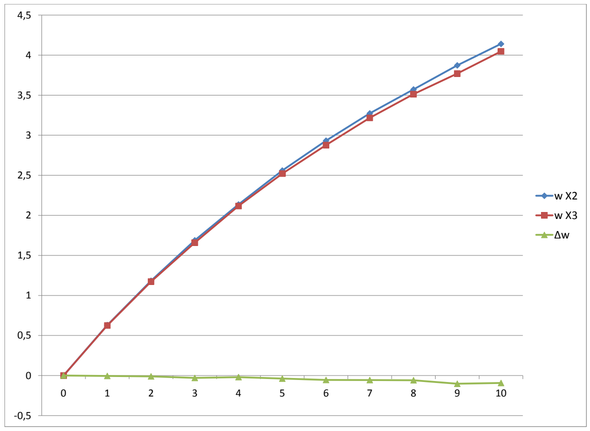

Figure 7.2 – Expected value as a function of money holdings under the powers-of-two and the powers-of-three structures.

Powers-of-two optimal portfolios	Value	Fraction with portfolio y
$y = \{y(1), y(2), y(3), y(4)\}$	$s.y$	$\theta^{X2}(y)$
$\{0,0,0,0\}$	0	0.15
$\{1,0,0,0\}$	1	0.18
$\{2,0,0,0\}$	2	0.16
$\{1,1,0,0\}$	3	0.14
$\{2,1,0,0\}$	4	0.12
$\{1,0,1,0\}$	5	0.10
$\{2,0,1,0\}$	6	0.07
$\{1,1,1,0\}$	7	0.04
$\{0,2,1,0\}$	8	0.03
$\{1,0,0,1\}$	9	0.02
$\{0,1,0,1\}$	10	0.01

Tableau 7.IV – Optimal portfolios with the powers-of-two structure sorted by increasing value.

value. Likewise, this choice does not tend to maximize the number of tokens to be transported because it increases the carrying cost of money. The final choice is then a kind of compromise between both options. They also always carry at least one token of the lowest denomination when their wealth allows that.

Powers-of-three optimal portfolios	Value	Fraction with portfolio y
$y = \{y(1), y(2), y(3)\}$	$s.y$	$\theta^{X3}(y)$
$\{0,0,0\}$	0	0.15
$\{1,0,0\}$	1	0.18
$\{2,0,0\}$	2	0.16
$\{3,0,0\}$	3	0.14
$\{1,1,0\}$	4	0.12
$\{2,1,0\}$	5	0.10
$\{3,1,0\}$	6	0.07
$\{1,2,0\}$	7	0.04
$\{2,2,0\}$	8	0.03
$\{3,2,0\}$	9	0.01
$\{1,0,1\}$	10	0.01

Tableau 7.V – Optimal portfolios with the powers-of-three structure sorted by increasing value.

7.4 Conclusion

This study aims at studying the social cost of denomination structures. In so doing, we proposed a framework that extends the Lee *et al.*'s model and that integrates several components of the costs of money : a carrying cost, a fixed cost per denomination, a maintenance cost and an initial production cost. The Lee *et al.*'s model is particularly well adapted since the denomination structure affects not only the payment efficiency but also the nominal output. Integrating the three other cost components makes the comparison of the social cost of currency systems more realistic. We used this framework to compare two specific modular denomination structures : a powers-of-two and a powers-of-three denomination structure. As an application, the parameterized version of our model shows that the social cost of powers-of-two system is higher than that of the powers-of-three when the fixed cost per denomination exceeds a threshold value.

Although innovative, our study has several limitations that should be raised in future works. First, the upper bound on money holdings and the average money detention can be raised to obtain larger transactions and denomination structures. This will help understand the effect of the carrying cost of money on the choice of portfolio and the payment behavior.[13] Another possibility is to raise the carrying cost, but in this case there is a convergence issue. According to our tests, the model converges to a degenerate steady state ($\pi = 0$, $w = 0$ ...) when this cost is set at a high value whereas for the intermediate values we obtain dynamics cycles - no steady state -, a 'nice' steady state is obtained only for the very low values.

Second, there are multiple optimal possibilities to transfer the wealth in a significant number of single coincidence meetings notified with (*) in Table 7.VII (Appendix). Those transfer solutions are selected by the agents regardless of the number of tokens exchanged during the transfer because this does not affect their expected value. Now given that the use of each token generates a depreciation (cf. Section 7.2.4), those solutions are therefore not equivalent in terms of the cost of maintenance. It is thus interesting to study to what extent the endogenisation of the cost of maintenance will affect the payment behavior of the agents and, more broadly, the ex-ante welfare. However, these extensions would need to redefine the equilibrium and thus to prove the existence of a steady state.

[13]Note that, with a wider set of possible individual money holdings, the computational capabilities required to compute the steady state will be more important.

7.5 Literature cited

Bergman, M., Guibourg, G., Segendorff, B., 2007, "The cost of paying : Private and social costs of cash and card," Working paper, Sveriges Riksbank.

Bouhdaoui Y., Bounie D. and Van Hove L., 2011, "Central banks and their banknote series : The efficiency-cost trade-off," *Economic Modelling*, 28(4) : 1482-1488.

Caianiello E.R., Scarpetta G. and Simoncelli G., 1982, "A systemic study of monetary systems," *International Journal of General Systems*, 8(2) : 81-92.

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

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

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

Garcia Swartz, D., Hahn, R., Layne-Farrar, A., 2006b, "The Move Toward a Cashless Society : Calculating the Costs and Benefits," *Review of Network Economics*, 5 : 199-228.

Lee M. and Wallace N., 2006, "Optimal divisibility of money when money is costly to produce," *Review of Economic Dynamics*, 9(3) : 541-556.

Lee M., 2010, "Carrying cost of money and real effects of denomination structure," *Journal of Macroeconomics*, 32(1) : 326-337.

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

Telser L.G. (1995). "Optimal Denominations for Coins and Currency," *Economics Letters*, 49 : 425-427.

Van Hove L. and Heyndels B., 1996, "On the Optimal Spacing of Currency Denominations," *European Journal of Operational Research*, 90 : 547-552.

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.

Zhu T., 2003, “Existence of a monetary steady state in a matching model : indivisible money,” *Journal of Economic Theory*, 112(2) : 307-324.

7.6 Appendix

In this section, we present the solution of the consumer's wealth transfer problem under the $X3$ structure.

Given our parameterizations, we enumerate 28 possible portfolios in the set of possible individual money holdings under the powers-of-three structure. Those portfolios are presented in Table 7.VI, they are sorted by ascending value then by ascending number of tokens, an index is associated to each portfolio. The Table 7.VII illustrates all the possible single coincidence meetings according to the indexes of the portfolios.

First, we observe that the agents tend to choose the portfolios that enhance their chances to carry out transactions and gain utility of consumption when possible. For instance, the agents with a wealth $z = 3$ have to choose between two portfolios, indexes 3 and 4 (cf. Table 7.VI). In the steady state, all those agents choose the portfolio of index 3 and no one that of index 4 because holding it implies that 75% of the possible meetings do not generate any transfer of wealth and therefore no consumption (cf. Table 7.VII).

Second, in a number of single coincidence meetings notified with "O", a transfer of wealth is feasible but does not take place. The reason behind this choice is that the potential gains from the utility of consumption do not compensate the loss in the expected value after the transfer of wealth.

Portfolio index	Composition	Value
1	$\{0,0,0\}$	0
2	$\{1,0,0\}$	1
3	$\{2,0,0\}$	2
4	$\{0,1,0\}$	3
5	$\{3,0,0\}$	3
6	$\{1,1,0\}$	4
7	$\{4,0,0\}$	4
8	$\{2,1,0\}$	5
9	$\{5,0,0\}$	5
10	$\{0,2,0\}$	6
11	$\{3,1,0\}$	6
12	$\{6,0,0\}$	6
13	$\{1,2,0\}$	7
14	$\{4,1,0\}$	7
15	$\{7,0,0\}$	7
16	$\{2,2,0\}$	8
17	$\{5,1,0\}$	8
18	$\{8,0,0\}$	8
19	$\{0,0,1\}$	9
20	$\{0,3,0\}$	9
21	$\{3,2,0\}$	9
22	$\{6,1,0\}$	9
23	$\{9,0,0\}$	9
24	$\{1,0,1\}$	10
25	$\{1,3,0\}$	10
26	$\{7,1,0\}$	10
27	$\{4,2,0\}$	10
28	$\{10,0,0\}$	10

Gray cells : optimal portfolios (cf. Table 7.V).

Tableau 7.VI – Index of the portfolios for the powers-of-three structure.

y_c \ y_p	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
2	(1)	(1)	(1)	(1)	(1)	(1)	(1)	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	x	x	x	x	x
3	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	o	o	o	o	o	o	o	o	x	x	x	x	x
4	o	o	(1)	o	(1)	o	(1)	(1)	(1)	o	(1)	(1)	o	o	o	x	x	x	x	x	x	x	x	x	x	x	x	x
5	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	x	x	x	x	x
6	(1)	(1)	(1)*	(1)	(1)*	(1)	(1)*	(1)*	(1)*	(1)	(1)*	(1)*	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	x	x	x	x	x
7	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	x	x	x	x	x
8	(1)	(1)	(1)*	(1)	(1)*	(1)	(1)*	(1)*	(1)*	(1)	(1)*	(1)*	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	x	x	x	x	x
9	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	x	x	x	x	x
10	(3)	(2)	(1)	o	(1)	(2)	(1)	(1)	(1)	o	(1)	(1)	o	o	o	x	x	x	x	x	x	x	x	x	x	x	x	x
11	(1)	(1)	(1)*	(1)	(1)*	(1)	(1)*	(1)*	(1)*	(1)	(1)*	(1)*	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	x	x	x	x	x
12	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	x	x	x	x	x
13	(1)	(1)	(1)*	(1)	(1)*	(1)	(1)*	(1)*	(1)*	(1)	(1)*	(1)*	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	x	x	x	x	x
14	(1)	(1)	(1)*	(1)	(1)*	(1)	(1)*	(1)*	(1)*	(1)	(1)*	(1)*	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	x	x	x	x	x
15	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	x	x	x	x	x
16	(1)	(1)	(1)*	(1)	(1)*	(1)	(1)*	(1)*	(1)*	(1)	(1)*	(1)*	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	x	x	x	x	x
17	(1)	(1)	(1)*	(1)	(1)*	(1)	(1)*	(1)*	(1)*	(1)	(1)*	(1)*	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	x	x	x	x	x
18	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	x	x	x	x	x
19	o	o	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
20	(3)	(2)	(1)	(3)	(1)	(2)	(1)	(1)	(1)	(3)	(1)	(1)	(3)	(3)	(3)	x	x	x	x	x	x	x	x	x	x	x	x	x
21	(2)	(1)	(1)*	(1)	(1)*	(1)	(1)*	(1)*	(1)*	(1)	(1)*	(1)*	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	x	x	x	x	x
22	(2)	(1)	(1)*	(1)	(1)*	(1)	(1)*	(1)*	(1)*	(1)	(1)*	(1)*	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	x	x	x	x	x
23	(2)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	x	x	x	x	x
24	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	x	x	x	x	x
25	(1)	(2)	(2)	(1)	(1)*	(1)	(1)*	(1)*	(1)*	(1)	(1)*	(1)*	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	x	x	x	x	x
26	(2)	(2)*	(2)*	(1)	(1)*	(1)	(1)*	(1)*	(1)*	(1)	(1)*	(1)*	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	x	x	x	x	x
27	(2)	(2)*	(2)*	(1)	(1)*	(1)	(1)*	(1)*	(1)*	(1)	(1)*	(1)*	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	x	x	x	x	x
28	(2)	(2)	(2)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	(1)	x	x	x	x	x

X : no wealth transfer is feasible.
O : a wealth transfer is feasible but no transfer occurs.
* : two or more possibilities of wealth transfer for the given transfer value.

Tableau 7.VII – Values of the optimal wealth transfers in a single coincidence meeting under the powers-of-three structure.

CHAPITRE 8

CONCLUSION

Les études empiriques relatives aux coûts sociaux des paiements s'accordent toutes à reconnaître que les pièces et billets sont les instruments de paiement les plus coûteux pour la société. Depuis quelques années, nous assistons donc à la mise en place de diverses stratégies par les banques commerciales et certaines banques centrales pour réduire la détention et l'usage des espèces. Parmi celles-ci, des stratégies non-tarifaires ont été mises œuvre à l'image des modifications du système de divisions monétaires ou des changements de technologie. Toutefois, leur réelle efficacité fait débat. Notre thèse s'est donc donnée pour ambition de mieux appréhender les incidences des comportements de paiement, des technologies et de la structure du SDM sur la distribution des transactions en espèces dans l'économie et leur coût social.

Articulée autour de six papiers, cette thèse contribue à améliorer la compréhension de ces relations sur plusieurs dimensions.

Premièrement, nous analysons les conséquences de l'assouplissement de certaines hypothèses restrictives des études théoriques qui comparent l'efficacité des SDM. En particulier, nous invalidons le principe d'invariance de Caianiello *et al.* (1982) en prouvant, d'une part, qu'il n'y a *pas d'égalité* entre la valeur moyenne des transactions en espèces et la valeur moyenne des pièces et billets en circulation et, d'autre part, qu'il n'y a *pas d'invariance* de cette moyenne suite au raffinement de la structure du SDM. De même, nous montrons qu'une distribution uniforme tend à surestimer la demande pour les valeurs faciales élevées et à sous-estimer celle des divisions de faible valeur, en l'occurrence la division 1 euro, en raison d'une occurrence plus élevée des montants ronds et d'une sous-représentation des montants faibles dans la distribution empirique.

Deuxièmement, nous contestons l'idée répandue selon laquelle un SDM efficace réduit le coût social des espèces. En effet, si un SDM efficace diminue certes les coûts variables de production, il comporte généralement un plus grand nombre de divisions qui augmentent les coûts fixes de production de sorte qu'il existe une condition limite pour qu'un SDM soit à la fois le plus efficace et le moins coûteux. En outre, nous mettons en évidence que des solutions de paiement efficaces du point de vue du principe du moindre effort peuvent être plus coûteuses pour la banque centrale en raison du coût d'usage des divisions mobilisées. En définitive, des solutions moins efficaces peuvent être, *ceteris paribus*, préférables du point de vue du coût pour la société. Ces résultats sont démontrés plus généralement dans le cadre d'un modèle d'équilibre général, dit *modèle de search*, dans lequel nous prenons en considération différentes composantes des coûts de production et de maintenance de la masse monétaire en circulation.

Troisièmement, nous proposons un cadre formel original pour mesurer la profitabilité de l'adoption d'une nouvelle technologie de paiement pour une banque centrale qui intègre plusieurs dimensions comme les coûts de production des technologies et les durées de vie des matériaux ainsi que leurs effets sur les revenus du seigneuriage et les coûts de la contrefaçon.

Au final, nos recherches permettent de renouveler un certain nombre de résultats établis dans la littérature économique même si nos travaux ne sont pas également exempts de toutes critiques.

Premièrement, notre première contribution invalide certes la preuve formelle de Caianiello *et al.* (1982) qui tente de justifier la loi empirique formulée par Hentsch (1973) sur la masse monétaire en circulation de chaque division. Toutefois, notre travail ne fournit aucune preuve alternative pour trouver une justification légitime à la loi empirique. Nous disposons seulement de quelques intuitions. Plus précisément, nous pensons en effet que la décroissance rapide des volumes en circulation en fonction de la valeur faciale des divisions est le résultat de deux effets combinés. Le premier est lié à la distribution observée des transactions en espèce dans l'économie qui suit une distribution log-normale et qui implique une décroissance des occurences des transactions en fonction de leurs montants. Le deuxième effet a trait au comportement de paiement efficace et neutre des agents économiques qui induit un usage 'équilibré' des divisions ; *i.e.* que les agents n'ont pas de préférences particulières pour les divisions.

Deuxièmement, nous utilisons la plupart du temps dans nos modèles sur la comparaison entre les SDM, le principe du moindre effort en supposant que les agents n'ont pas de restrictions dans leur portefeuille. Cette simplification n'altère pas la portée de nos résultats généraux. Mais dans un souci de réalisme économique, il serait sans doute souhaitable à l'avenir d'introduire un principe de rareté dans le portefeuille des agents à l'image de Bounie et Soriano (2006). Dans leur modèle, les auteurs supposent que les agents ne disposent pas de pièces et billets pour réaliser des transactions et qu'un retrait préalable au distributeur automatique de billets est nécessaire. Lors du règlement du premier prix, le consommateur fournit l'un des billets retirés au marchand qui lui rend en contrepartie certaines divisions et ainsi de suite. Cette hypothèse sur le comportement de paiement est sans doute plus réaliste et se rapproche des résultats empiriques décrits par Franses et Kippers (2007).

Troisièmement, dans notre tentative de fournir un cadre général sur les changements de technologie, nous raisonnons sur des moyennes. Il serait techniquement plus juste de construire un modèle intertemporel qui actualise les coûts et les bénéfices dans le temps. Cette prise en compte de la dimension temporelle serait plus appropriée également pour considérer des économies d'échelle qui évoluent dans le temps et en fonction des technologies.

Quatrièmement, pour avoir une mesure plus précise du coût que représente pour la banque centrale le paiement efficace selon le principe du moindre effort, il serait souhaitable, dans Bouhdaoui et Bounie (2011), de tenir compte des économies d'échelle dans les coûts de production des coupures monétaires en distinguant les coûts fixes par division et les coûts variables. Un autre point d'amélioration serait d'assouplir notre hypothèse restrictive sur l'indépendance entre le comportement de paiement des agents et les vitesses de circulation des divisions.

Cinquièmement, les résultats établis dans Bouhdaoui (2011) reposent sur un seul paramétrage du modèle. Une analyse d'élasticité donnerait sans doute plus de portée à ces résultats et permettrait d'affiner la compréhension des relations entre les variables exogènes du modèle et les indicateurs économiques de bien-être social et de coût social des espèces mesurés à l'état stable.

Ces réserves mises à part, nous pensons que nos contributions présentées dans ce manuscrit ouvrent néanmoins la voie à de nouvelles pistes de recherche nécessaires pour comprendre les relations entre la structure du SDM, la distribution des transactions dans l'économie, le comportement de paiement en espèces du public et le coût social généré par cet instrument de paiement. Nous indiquons pour conclure quelques unes de ces voies sur lesquelles nous souhaiterions à l'avenir nous concentrer. Elles se fondent en partie sur la littérature des *modèles de search*, dans la continuité de la dernière contribution Bouhdaoui (2011).

Premièrement, nous souhaiterions élargir notre cadre d'étude pour évaluer l'efficacité des stratégies tarifaires dans la réduction du coût social des espèces. Dans ce cadre, nous pourrions évaluer l'effet d'une taxation de l'usage des espèces en endogénéisant les coûts de maintenance, générés par la dépréciation, dans la fonction d'utilité des agents. Nous comparerions ensuite le coût social des espèces obtenu par ce mode de tarification au scénario initial où le coût de maintenance est financé par une taxe forfaitaire identique pour tous les agents. Cette étude pourrait également intégrer d'autres effets annexes tels que l'impact de la tarification à l'usage sur la distribution des transactions, par exemple l'occurrence des prix ronds, et sur le comportement de paiement des agents.

Deuxièmement, nous avons le projet d'élargir davantage notre problématique en traitant la question de la demande d'espèces dans l'économie. Nous souhaiterions en particulier modifier le cadre de Bouhdaoui (2011) en autorisant les dépôts bancaires rémunérés. Ainsi, le problème du consommateur consisterait au préalable à déterminer la part des ressources à faire fructifier sur le dépôt bancaire et la part à transporter dans le portefeuille pour réaliser des transactions génératrices d'utilité. L'étude fournirait notamment des indices utiles sur l'influence des taux d'intérêt sur la demande d'espèces dans l'économie tout en comparant les résultats obtenus avec ceux des modèles classiques d'équilibre général de la demande de monnaie.

Enfin, pour terminer et élargir notre champ d'étude, on notera que les *modèles de search* sont également utilisés pour évaluer les politiques monétaires (Lee, 2009 ; Berentsen *et al.*, 2009). Dans ce cadre, une approche originale consisterait à introduire un nouvel acteur dans notre modèle de base qui incarne l'autorité monétaire. Ce dernier serait doté d'une fonction objectif et interviendrait à chaque période pour ajuster les variables exogènes du modèle en se basant sur les indicateurs économiques. Les perspectives de cette approche pourraient être prometteuses pour comparer les politiques monétaires. Cependant, avant de pouvoir l'implémenter, il est d'abord nécessaire de déterminer les conditions pour que l'itération des périodes converge vers un état stable que l'on pourra étudier. Cette évolution nécessite en d'autres termes de prouver théoriquement l'existence d'états stables.

CHAPITRE 9

RÉFÉRENCES INTRODUCTION ET CONCLUSION

Attanasio O., Guiso L. et Japelli T., 2002, “The Demand for Money, Financial Innovation, and the Welfare Cost of Inflation : An Analysis with Household Data”, *Journal of Political Economy*, 110(2) : 317-351.

Banque Nationale de Belgique, 2006, “Costs, Advantages and Drawbacks of the Various Means of Payment”, *Economic Review*, 41-47.

Baumol W. J., 1952, “The Transaction Demand for Cash - An Inventory Theoretic Approach”, *Quarterly Journal of Economics*, 66(4) : 545-56.

Berentsen A., Menzio G. and Wright R., 2009, “Inflation and Unemployment in the Long Run”, IEW - Working Papers, University of Zurich.

Boeschoten W.C. et Fase M.M.G., 1989, “The Way We Pay With Money”, *Journal of Business and Economic Statistics*, 7(3) : 319-326.

Bolt, W., Jonker N. and van Renselaar C., 2010, “Incentives at the Counter : An Empirical Analysis of Surcharging Card Payments and Payment Behaviour in the Netherlands”, *Journal of Banking and Finance*, 34(8) : 1738-1744.

Bounie D., 2002, “Quelques implications bancaires et monétaires du développement des systèmes de paiement électronique”, *Revue Économique*, 52 : 313-331.

Bounie D. et Soriano S., 2006b, “La Substitution de la Monnaie Electronique à la Monnaie Fiduciaire : Modèle et Simulations”, *Revue Française d’Economie*, 20(3) : 153-182.

Bounie D. et François A., 2006, “Les Déterminants de la Détention et de l’Usage des Instruments de Paiement : Eléments Théoriques et Empiriques”, *Revue d’Economie Financière*, 83(3) : 159-173.

Bounie D. et François A., 2008, “Is Baumol’s ’Square Root Law’ Still Relevant ? Evidence from Micro-Level Data”, *Applied Financial Economics*, 18(13) : 1091-1098.

Borzekowski R. et Shaista A., 2006, “Consumers’ Use of Debit Cards : Patterns, Preferences, and Price Response”, *Journal of Money, Credit and Banking*, 40 : 149-172.

Caianiello E.R., Scarpetta G. et Simoncelli G., 1982, “A Systemic Study of Monetary Systems”, *International Journal of General Systems*, 8 : 81-92.

Coventry L., 2001, “Polymer Notes and the Meaning of Life”, Currency Conference, Barcelona, Spain, April.

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

De Grauwe P. *et al.*, 2006, “Towards a More Efficient Use of Payment Instruments”, Discussion Papers, University of Leuven.

De Nederlandsche Bank, 2004, “The cost of payments”, *Quarterly Bulletin*, 57-64.

Drehman M., Goodhart C. et Krueger M., 2002, “The challenge Facing Currency Usage : Will the Traditional Transaction Medium be able to Resist Competition from the New Technologies”, *Economic Policy*, 34 : 193-227.

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

Garcia Swartz D.D., Hahn R.W. et Layne-Farrar A., 2006, “The Move Toward a Cashless Society : A Closer Look at Payment Instrument Economics”, *Review of Network Economics*, 5(2) : 175-198.

Gowrisankaran G., et Krainer J., 2005, “The Welfare Consequences of ATM Surcharges : Evidence from a Structural Entry Model, ” Federal Reserve Bank of San Francisco, Working Paper 2005-01.

Guibourg G. et Segendorff B., 2007, “A Note on the Price and Cost Structure of Retail Payment Services in the Swedish Banking Sector 2002”, *Journal of Banking and Finance*, 31 : 2817-2827.

Hayashi F. et Klee E., 2003, “Technology Adoption and Consumer Payments : Evidence from Survey Data”, *Review of Network Economics*, 2(2) : Article 8.

Hentsch J.C., 1973, “La Circulation des Coupures qui Constituent une Monnaie”, *Journal de la Société de Statistique de Paris*, 114(4) : 279-286.

Humphrey D.B., Kim M. et Vale B., 2001, “Realizing the Gains from Electronic Payments : Costs, Pricing, and Payment Choice”, *Journal of Money, Credit and Banking*, 33(2) : 216-234.

Humphrey D. B., Lawrence B. P., Vesala I. M., 1996, “Cash, Paper, and Electronic Payments : a Cross-Country Analysis”, *Journal of Money, Credit and Banking*, 28(4) : 914-939.

Kiser E., R. Borzekowski et S. Ahmed, 2008, “Consumers’ Use of Debit Cards : Patterns, Preferences, and Price Response”, *Journal of Money, Credit and Banking*, 40(1) : 149-172.

Knotek E.S., 2008, “Convenient prices, currency, and nominal rigidity : Theory with Evidence from Newspaper Prices”, *Journal of Monetary Economics*, 55(7) : 1303-1316.

Lee M., 2009, “Carrying Cost of Money and Real Effects of Denomination Structure”, *Journal of Macroeconomics*, 32 : 326-337.

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

Lee M. et Wallace N., 2006, “Optimal Divisibility of Money When Money is Costly to Produce”, *Review of Economic Dynamics*, 9(3) : 541-556.

Lotz S. et Rocheteau G., 2004, “The Fate of One-dollar Coins in the U.S," Economic Commentary, Federal Reserve Bank of Cleveland, issue Oct 15.

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

Mot, E.S. and J. Cramer, 1992, “Mode of payment in household expenditures”, *De Economist*, 140(4) : 488-500.

Rinaldi L., 2001, Payment Cards and Money Demand in Belgium, International Economics Working Papers Series, University of Leuven.

Rochet, J-C. et Tirole J., 2002, “Cooperation Among Competitors : Some Economics of Payment Card Associations”, *RAND Journal of Economics*, 33(4) : 549-570.

Rochet, J-C. et Tirole J., 2003, “Platform Competition in Two-Sided Markets”, *Journal of the European Economic Association*, 1(4) : 990-1029.

Rysman M., 2007, “An Empirical Analysis of Payment Card Usage”, *Journal of Industrial Economics*, 55(1) : 1-36.

Snellman J.S., Vesala J.M. et Humphrey D.B., 2000, “Substitution of Noncash Payment Instruments for Cash in Europe”, Bank of Finland discussion papers, 1/2000.

Stavins J., 2001, “Effect of Consumer Characteristics on the Use of Payment Instruments”, *New England Economic Review*, (3) : 19-31.

Stix H., 2004, “How Do Debit Cards Affect Cash Demand ? Survey Data Evidence”, *Empirica*, 31 : 91-115.

Sumner S., 1993, “Privatizing the Mint”, *Journal of Money, Credit and Banking*, 25 : 13-29.

Telser L.G., 1995, “Optimal Denominations for Coins and Currency”, *Economics Letters*, 49 : 425-427.

Tschoegl A.E., 1997, “The Optimal Denomination of Currency : A Conjecture”, *Journal of Money, Credit and Banking*, 29 : 546-554.

Van Hove L. et Heyndels B., 1996, “On the Optimal Spacing of Currency Denominations”, *European Journal of Operational Research*, 90 : 547-552.

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., 2004a, “Cost-Based Pricing of Payment Instruments : the State of the Debate”, De Economist, 152(1) : 79-100.

Van Hove L., 2004b, “Electronic Purses in Euroland : Why Do Penetration and Usage Rates Differ ? ”, SUERF Studies, No. 2004/4.

Williamson S. et Wright R., 2010, “New Monetarist Economics : Models”, in Friedman B. and Woodford M., *Handbook of Monetary Economics*.

Zhu T., 2003, “Existence of a Monetary Steady State in a Matching Model : Indivisible Money”, *Journal of Economic Theory*, 112(2) : 307-324.