ABSTRACT

Title of dissertation: ESSAYS ON INFORMATION CONTAGION AND MEDIA OF EXCHANGE

David McArthur, Doctor of Philosophy, 2013

Dissertation directed by: Professor John Rust
Department of Economics

CHAPTER 1: INFORMATION CONTAGION

In social media, information spreads like a contagion from person to person. When many pieces of information are spreading and competing in the space of social media messages, their propagation rates become very unequal. This is because contagion creates a positive feedback which amplifies small, random differences in prevalence. A framework for modeling social media is developed that suggests how to infer the communication choices of social media participants from the observed heavy-tailed count distribution of messages containing different pieces of information. In a Monte Carlo simulation where agents with rational expectations make individually optimal communication choices, the feedback effect is only partially mitigated. Even with fully rational behavior, information that no-one has an especially high propensity to pass along can "go viral".

CHAPTER 2: SEARCH AND BARGAINING WITH PLASTIC: MONEY AND CHARGE CARDS AS COMPETING MEDIA OF EXCHANGE

Charge cards are introduced into the Lagos-Wright money search model as an alternative medium of exchange competing with money. I explore why cards and money may coexist, and examine the implications of intermediated exchange for monetary policy. Charge cards lower the social cost of inflation because they overcome the hold up problem with money that otherwise results in too little exchange. Some inflation can even be beneficial if a higher cost of holding money pushes agents to become cardholders. Moreover, higher nominal interest rates help card companies set higher spending limits, which can also increase the level of exchange and improve welfare.

ESSAYS ON INFORMATION CONTAGION AND MEDIA OF EXCHANGE

by

David McArthur

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2013

Advisory Committee:
Professor John Rust, Chair/Adviser
Professor Ginger Jin
Professor Enrique Mendoza
Professor John Shea
Professor Ben Shneiderman

CONTENTS

1. *Information Contagion* 1
   1.1 Introduction 2
      1.1.1 Inference 4
      1.1.2 Empirical challenges 7
      1.1.3 Theoretical refinements 9
      1.1.4 Structural models 11
      1.1.5 Related literature 14
      1.1.6 Organization 16
   1.2 Empirical evidence 17
      1.2.1 Distribution of counts 19
      1.2.2 Arrival analysis 23
      1.2.3 Binomial sub-sampling 28
   1.3 Model 30
      1.3.1 Agent based simulation 32
      1.3.2 Pólya urn process 33
      1.3.3 Birth of facts – Yule-Simon process 38
      1.3.4 Birth and death – generalized Yule-Simon 41
   1.4 Microeconomic model 44
      1.4.1 Solution for binary model 48
      1.4.2 Solution for intensity model 53
   1.5 Conclusion 61
   1.A Top quoted phrases 69

2. *Search and Bargaining with Plastic:*
   Money and Charge Cards as Competing Media of Exchange 71
   2.1 Introduction 72
   2.2 Facts about charge cards 78
   2.3 Model 81
      2.3.1 Lagos and Wright environment 81
      2.3.2 Charge cards 83
      2.3.3 Bargaining 90
      2.3.4 Distribution of money holdings 105
      2.3.5 Card-accepting and cardholding decisions 109
      2.3.6 Spending limits 113

2.4 Equilibrium 116
2.4.1 Monetary equilibrium 117
2.4.2 No-trade equilibrium 118
2.4.3 Charge card equilibrium 119
2.4.4 Mixed equilibrium 119
2.5 Simulation 120
2.5.1 Endogenous spending limit 122
2.5.2 Welfare benefit of charge cards 124
2.5.3 Cost of inflation 126
2.6 Extensions 127
2.6.1 Endogenize card company fees 127
2.6.2 Heterogeneous sellers 128
2.6.3 "No surcharge rules" and bargaining over price 129
2.7 Conclusions for monetary policy 130
2.A Proofs 132

1. INFORMATION CONTAGION

"If the news is that important, it will find me."

– unidentified college student in a focus group

quoted in Stelter (2008)

1.1 Introduction

In social media, what is the relationship between individual participants' decisions to pass along a piece of information and the total number of copies of that information that are produced? A common metaphor for the communication of information from person to person is that of a virus spreading through a population. Indeed there is a longstanding literature in mathematics that uses stochastic models of epidemics to describe the spread of "rumors". The classic model is that of Daley and Kendall (1965)[1]. There are two problems with using these models to analyze social media. First, there is no place for *media* in epidemic rumor models. Rumors pass directly from person to person whenever an "infective" meets a "susceptible". In contrast, most social media consist of messages – tweets, posts, blog entries. The message is both the mechanism by which information is transmitted and the record which researchers are able to observe. Second, there is no *decision* in epidemic rumor

[1] The feature that distinguishes epidemic rumor models from the classic SIR (Susceptible, Infective, Removed) models of epidemiology is that infectives become removed upon meeting another individual who is either infective or removed (i.e. when attempting to pass the rumor to someone who has already heard it). In rumor models, removed individuals are sometimes called "stiflers" because they actively affect the progress of the epidemic. In a sense it is not only the "disease" that is contagious, but also the "cure".

models. Rather, transmission of information is involuntary. So, in order to connect communication choices with observable outcomes in terms of messages produced, I put forth a new framework for modeling information contagion: agents who know a piece of information choose to post messages in a medium that is consumed by other agents.

Explicitly including the medium in the framework allows models to naturally capture a striking empirical feature of social media: the number of copies of a given piece of information, unconditional on the information's age, has a heavy-tailed probability distribution. Section 1.2 presents data collected by Spinn3r, a search engine focusing on social media web sites. The Spinn3r data was previously analyzed by Leskovec et al. (2009) who identified phrases of words that recur on multiple pages, and made the resulting dataset available on memetracker.org. The number of pages containing a given phrase is roughly Pareto distributed (for a phrase chosen at random *regardless of age*, the probability of observing k or more occurrences is proportional to $k^{-\rho}$ above some minimum value of k). My proposition is that these phrases serve as markers for distinct pieces of information which are spreading through social media. However, such a count distribution does not arise in models that consider a single contagion in isolation. Instead the heavy-tailed distribution suggests an interaction between many contagions competing for the chance to spread (see section 1.3). This competition is captured naturally by models in which information spreads through messages and agents read messages at random. The heavy-tailed distribution of messages arises because of a feedback effect whereby the more messages there are containing a piece of information, the more likely that information is to be learned by a new agent who will go on to re-post it in additional messages.

1.1.1 Inference

The implication of the feedback effect is that some pieces of information spread much faster than others, purely by chance. Suppose it was observed that phrase A appears in ten times as many posts as phrase B. One could not infer that social media participants are ten times as likely to re-post A as they are to re-post B. Moreover, given the value of $\rho = 1.24$ estimated from the overall memetracker dataset, one could not even reject the hypothesis that A and B are *equally likely* to be re-posted[2]. The outcome of information being reproduced many times may be due to agents having a higher propensity to re-post it, but could just as well be because the information happened to spread quickly and this early advantage has snowballed.

That the feedback effect can dominate and obscure individual choices is illustrated by a simple agent-based simulation (see section 1.3.1): agents

[2] In a two-tailed statistical test of the null hypothesis that the number of posts containing A and the number containing B come from the same Pareto distribution with parameter $\rho = 1.24$, observing a test statistic $\max(x_A, x_B)/\min(x_A, x_B)$ of 10 has a p-value of 0.0515. One could not reject the null at a 5% confidence level.

communicate via an idealized message board where each message contains a single piece of information and agents read messages at random. Different pieces of information have different values, and each one is initially known by a single agent. Every period each agent posts the most valuable information she knows, up to the point where the marginal cost of posting an additional message exceeds the value of the marginal information. Figure 1.8 shows a typical run of this simulation. The information which spread fastest was only of medium value, but it reached half the population before a majority of information spread to more than a handful of agents. The correlation between value and prevalence is quite poor.

Little can be inferred from the prevalence of an individual piece of information. However, given a collection of different pieces of information, their occurrence counts can be fit to a Pareto distribution and some inference might be made from the fitted value of ρ. A basic analytical model is laid out in section 1.3.3 which is equivalent to a Yule process (also known as preferential attachment in the literature on network formation). If agents find messages and re-post the information they contain at a fixed rate λ and new pieces of information arrive at rate μ, the steady state distribution of the number of occurrences will be Pareto with parameter $\rho = (\mu + \lambda)/\lambda$. If the birth rate μ is observed, the rate of re-posting by individual agents can be deduced from ρ.

This brings us a step closer to being able to observe agents' choices from outcomes. By fitting separate Pareto distributions to different *sets* of information, it may be possible to infer a common factor in the propensity to re-post across members of one set relative to members of another set. For example, figure 1.2 in section 1.2.1 shows two sets of phrases: those containing the word "Lehman" which mainly refer to the financial crisis, and phrases containing the sequence of words "broke up with" which seem to be predominantly about celebrities. Each set of phrases has it's own count distribution of the number of messages containing a random phrase from the set. For both "Lehman" phrases and "broke up with" phrases, the most frequent outcome is the search engine finding only a single message containing the phrase. The second most common outcome is two occurrences, and so on. The typical phrase from either set has hardly spread at all. Yet for both types, there are some phrases that have been reproduced many times. About one in 70 "Lehman" phrases are found in ten or more messages. For "broke up with" phrases the fraction is about one in 90. For both types, a tiny fraction of phrases have made it onto hundreds of web pages.

More succinctly, both count distributions seem to be moderately good fits to Pareto distributions. We would like to be able to use their respective Pareto coefficient estimates $\hat{\rho}_{Lehman} \approx 1.35$ and $\hat{\rho}_{brokeupwith} \approx 1.95$ and their respective arrival rates of new phrases $\mu_{Lehman} = 1427/month$ and $\mu_{brokeupwith} = 147/month$ to determine the relative propensity of social media participants to pass along each type of information. Here we find $\lambda_{Lehman} = 26.6 \times \lambda_{brokeupwith}$. One might conclude that people will re-post news about the crisis at a rate 26.6 times higher than the rate they pass along celebrity gossip! Unfortunately there are both practical and theoretical issues with this result.

1.1.2 Empirical challenges

First there is the issue of data sampling. Social media participants have many channels for communication and it is unlikely that a single dataset will capture all of the messages containing a given piece of information. Certainly the memetracker data does not contain every post on every platform. This matters because the measurements of interest are the counts of occurrences, and the *relative counts in a sample are not the same as the relative counts in the population.* Section 1.2.3 on binomial sub-sampling demonstrates how observing a message with with probability p distorts the count distribution. The sample distribution will have relatively fewer large counts than the population distribution. Fortunately, if the population is Pareto distributed the sample distribution will be *asymptotically* Pareto (see Stumpf et al., 2005, for a treatment of Pareto count distributions and sampling). This means for a sufficiently large value of k_{min}, the sample distribution will follow $Prob(x \geq k \mid x \geq k_{min}) = \left(\frac{k}{k_{min}}\right)^{-\rho}$ and it may still be possible to estimate ρ. However, this requires sufficiently large samples that the Pareto can be fit using only the small portion of information that attains large counts. Moreover, sampling complicates estimating the birth rate μ of new information. Many pieces of information in the population will have zero counts in the sample. Inferring the number of missing pieces of information requires knowing the full count distribution that was thinned by the binomial sub-sampling, not just the upper tail.

Inference from a sample also depends on the sample being representative of the population, i.e. the probability p of inclusion in the sample should be independent of the message's content or other characteristics. This may not be the case for real world datasets. For example, there is evidence that the memetracker data is *not* representative of the whole population of social media messages. Figure 1.3 shows a quantile-quantile plot comparing two empirical distributions: the count distribution of phrases containing "inflation" and the count distribution of phrases containing "playoff". If these counts came from different Pareto distributions with different values for ρ, then the QQ plot would show an exponential relationship between the two. The quantile pairs would fall on a straight line only if the QQ plot had a logarithmic scale (the slope would be $1 + (\rho_{inflation} - \rho_{playoff})$). Instead the relationship between the quantiles is roughly linear, but with a slope around 1.5 rather than one. This means the two sample distributions have the same shape, but differ by a scale factor: the probability of a "playoff" phrase occurring in k messages is the same as the probability of an "inflation" phrase occurring in $1.5k$ messages (perhaps only for $k > 6$). One explanation of this is that the population distributions are actually the same – both Pareto with ρ values too similar to distinguish – but the probability of a "playoff" message being *sampled* is higher, with $p_{playoff}/p_{inflation} \approx 1.5$.

1.1.3 Theoretical refinements

The theoretical issues revolve around the questions of whether we really expect to find a fixed re-posting rate λ and why we should not expect to find ideal Pareto distributions. First, the Yule process itself produces a distribution that is only asymptotically Pareto. Section 1.3.3 gives the exact count distribution which is named Yule-Simon and has somewhat fewer low count occurrences than Pareto (see figure 1.9 for a comparison). In itself this is not much of as issue since the two distributional forms converge quickly and are practically indistinguishable for counts above 10. If complete datasets (containing the entire population of messages through which information spreads) were available, one might use the goodness of fit of the Yule-Simon on low counts as a test of the basic mechanism. But when working with a sample, the binomial sub-sampling effect distorts the low counts and the Pareto coefficient can only estimated from the higher counts anyway.

More problematic is that the basic social media model built on the Yule process assumes that the total number of messages can grow without bounds. Even though each piece of information continues moving up in message count, the distribution converges to a steady state when the *proportion* of information at each size is such that the growth balances the constant addition of new information which each start out with a single message. The far upper tail of the Yule-Simon/Pareto distribution contains the pieces of information that have been spreading for a very long time yet continue to find new agents who will start re-posting them.

But in a finite population of agents there must be an upper-bound to the distribution, even if all agents re-post old information forever. On the other hand, with an upper bound and continuous birth of new information there cannot be a steady state distribution unless there is also *death* of information. At some point agents must stop re-posting information, otherwise an ever growing proportion of information would be at that upper bound. Section 1.3.4 introduces a social media model consistent with a finite population, in which agents stop re-posting information when it exceeds an age limit. The resulting count distribution is a generalized version of the Yule-Simon that is Pareto-like for counts above the single digits but has a truncated upper tail.

1.1.4 *Structural models*

The generalized Yule-Simon may be a useful reduced form. But understanding how the structure of social media and changes in communication technology will affect the availability of information to consumers requires a structural model of agents' communication decisions. Solving such a model necessitates specifying the costs and benefits faced by social media participants. Section 1.4 presents a micro-founded dynamic programming model of social media in which posting messages is costly, each piece of information yields a value to each agent who learns it, and agents who know information can choose to post it in a message because they can capture some of the value from the agent who reads their messages. Agents in this model do not know how many other agents have already learned each piece of information. But agents do know the age of information and have rational expectations about the probability distribution of how many agents know information as a function of its age.

Two examples of this structural model are presented that differ in the communication technology assumed. In the first, the social media platform permits agents to choose only whether or not to post messages for each piece of information they know. If they post, they do so at a fixed rate. Agents will choose to keep posting messages until the information reaches a threshold age. After this threshold the value of communicating has dropped to the point that the share accruing to the message poster, times the probability a given message will be read, is below the cost of posting. The benefit of communicating decreases with information's age because of the increasing probability that an agent reading a message has already learned the information and gets no additional value. In addition to the direct value, an agent who learns information can begin re-posting it and capture some value from additional agents. This "resale" value of information depends on the entire future path of all the agents who will be infected indirectly. Each agent can extract a share of all the value from the agents downstream. But each additional agent who knows a piece of information also increases the competition. This example rationalizes the reduced form Yule process with birth and death, and therefore yields the same generalized Yule-Simon distribution. But here the age limit, and thus the distribution's truncation, are derived from model primitives (value of information, size of population and cost of posting). This dynamic programming example is solved numerically in section 1.4.1 using a Monte Carlo nested fixed point algorithm.

A fixed message posting rate, per piece of information, may or may not not be a good description of the communication behavior of social media participants today. But it seems unrealistic to impose it as a restriction of the message posting technology. A second example of the micro-founded structural model allows agents to choose the rate at which they post, facing a cost function that is increasing and convex in the number of messages per period. Agents with rational expectations will post more messages containing new information and taper off posting as the information ages. This allows us to address the important question of whether the contagion effect is only transitory or is fundamental to social media. Might not fully optimizing agents who anticipate the feedback effect arrive at a more equal distribution of information?

Section 1.4.2 presents a numerical solution of this example. The equilibrium of this system is an aggressive policy function that posts many messages for young information and reduces the rate quickly as information ages (see figure 1.13). This does produce a more equal distribution than the generalized Yule-Simon. But the distribution still has a (truncated) heavy tail. There are 10,000 agents in this example model. The equilibrium policy transmits half of all pieces of information to over 100 agents (one percent of the population) where a much smaller fraction of information reaches that many agents under the model with fixed re-posting rates. On the other hand, the equilibrium distribution is a good match to the generalized Yule-Simon *conditional* on having reached about 98 people (see figure 1.16). This is the expected number of agents who know the information when the equilibrium policy transitions from being very age sensitive to less sensitive. The total number of messages posted for a piece of information (# posts/agent $\times$ # agents infected) has a sharp peak at this age.

The conclusion is that optimizing social media participants with rational expectations, who take into account the feedback mechanism, will only partially mitigate that feedback. The resulting spread of information will still result in some “lucky” information spreading a lot. Researchers observing the messages from such an equilibrium system will still find a count distribution that has a Pareto-like range over orders of magnitude of counts.

1.1.5 Related literature

As mentioned above, the classic model epidemiological model of rumors is Daley and Kendall (1965). For a modern extension that accounts for the structure of the social network through which rumors can spread see Nekovee et al. (2007).

It has long been known that the number of citations to academic journal articles and the number of hyperlinks to documents on the World Wide Web both have highly skewed distributions. Since at least the late 1990’s it’s been widely accepted that these follow Pareto distributions (see Redner 1998 on citations and Barabasi and Albert 1999 on hyperlinks), though some argue that these distributions are in fact log-normal (see Mitzenmacher 2004 for a good overview of this debate.) The burgeoning literature on topic detection and tracking has occasionally noted a similar highly skewed distribution in the number of documents containing a given phrase or set of words (Gruhl et al., 2004; Leskovec et al., 2009).

A recurring explanation is that the generative processes creating new hyperlinks, references to past research, and articulations of phrases all exhibit *preferential attachment* (a.k.a. "cumulative advantage"). New instances occur in proportion to the number of existing instances, presumably via some form of copying what has come before. The first treatment of such a process dates back to Yule (1925). The Yule process was most famously revived by Simon (1955), but similar mechanisms for generating asymptotically Pareto distributions have been reinvented many times. See Simkin and Roychowdhury (2006) for a comparison of the numerous derivations of this result.

Information contagion bears a superficial resemblance to information cascades. Contagion describes how information passes from person to person, whereas information cascades describes the situation where agents infer others' information by observing their actions. The resulting herd behavior, where all agents copy the action of the first few actors in disregard of their own private information, is wholly unrelated to the situation where some pieces of information quickly spread to a large portion of the population because of the feedback effect of contagion.

1.1.6 Organization

The remainder of this paper is organized as follows: Section 1.2 presents the Memetracker data and shows (a) that the count distribution is heavy tailed and the same for subsets of phrases on completely different topics, and (b) that conditional on already having n mentions, the arrival of the $n+1$th mention looks the same regardless of n, except that its overall magnitude increases in n. Section 1.3.1 shows the simple agent-based model where each agent only passes along the most valuable information she knows, but the outcome shows little correlation between value and prevalence. Section 1.3 develops a branching process model of information contagion. A simple one-shot version of the model where agents follow basic rules is introduced and solved analytically using a combinatorial urn model is described in section 1.3.2. A birth process for information is introduced in section 1.3.3 which makes the urn model equivalent to a Yule-Simon process. This would have an ergodic distribution as its solution, but perpetual growth is not compatible with contagion with a finite number of susceptible agents. Section 1.3.4 adds a maximum age for information to be communicated, which means that contagion can continue without saturating the population. Section 1.4 replaces the basic rules-based behavior of agents with a full microeconomics-founded model of communication choice. Agents are fully rationally utility maximizers with rational expectation. Sections 1.4.1 and 1.4.2 describe the computational solution to this problem under two different assumptions for the communication technology. In both solutions, the resulting count distribution is heavy tailed despite agents making individually optimal choices. Section 1.5 concludes.

1.2 Empirical evidence

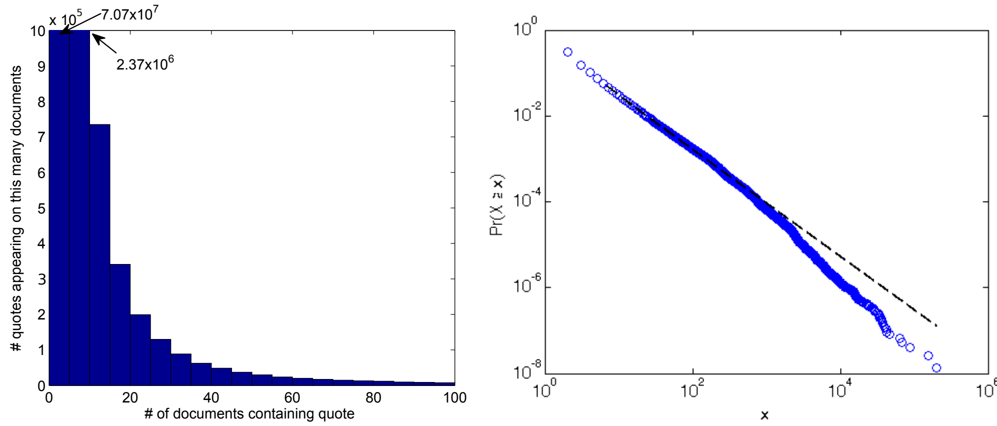
Spinn3r is a company that provides a back-end web crawler service which indexes social media, blogs, and mainstream news web sites. Their customers are companies that analyze trends, track the impact of news stories, or provide front-end search engines for the "blogosphere". Leskovec et al. (2009) analyzed nine months of Spinn3r data, and have made their sample publicly available at memetracker.org. The Memetracker data contains records of 97 million documents that were indexed between August 1, 2008 and April 30, 2009. The data contains the URL of each document and the time stamp from when the search engine visited it. In addition, from each document, every phrase of three or more words appearing inside quotation marks is identified. Most of these are actual quotations where the document's author is quoting the words of another person. Titles of books, films and television shows also appear within quotation marks. These are rare overall, but make up a sizable number of the most common quoted phrases. See appendix 1.A for a list of the top phrases and their counts.

I interpret each phrase as a marker for a narrow topic, and assume all documents that include a given phrase contain one or more pieces of information – fact or opinion – about that topic. Sometimes, the phrase itself conveys a specific piece of information, such as "Palm Beach County residents, claim your economic stimulus payment" which is from a press release put out by Florida senator Mel F. Martinez inviting the public to a workshop to help them fill out paperwork. The phrase went on to appear in 28,374 documents indexed by Spinn3r. If you lived in Palm Beach County, learning this fact before October 15, 2008 may have been quite valuable. Sometimes the phrase expresses an opinion, such as "Phil Gramm is the single most important reason for the current financial crisis" which was said by a former S.E.C. lawyer and first published in an article in the New York Times on Nov 16, 2008. This quotation appeared in 35,617 documents, many of them blog postings or message boards. The set of documents tagged by this phrase probably contain multiple, perhaps divergent, opinions and facts about the extent to which deregulation was to blame for the crisis.

The most repeated quoted phrase was the epithet "gang of ten" which referred to a group of US Senators who negotiated a compromise on the *New Energy Reform Act of 2008.* This phrase appeared in numerous news stories throughout August and September 2008, and went on to appear in 154,715 documents in the Spinn3r dataset. On the one hand the phrase is associated with many distinct pieces of information about the political process and the numerous components of this bill. On the other hand, all of these distinct facts or opinions are closely related under the umbrella of 'bipartisan agreement passes energy bill' and it can be argued that "gang of ten" is a good marker for this nexus of topics. Through the data mining techniques of topic detection and sentiment analysis it should be possible to obtain a more fine grained list of topics, and ultimately attempt to identify the specific information content of each document. This is beyond the scope of this paper, and is not possible using the public Memetracker data, which contains only the quoted phrases from each document and not the complete text.

1.2.1 Distribution of counts

Leskovec et al. (2009) identified 75 million distinct quoted phrases in the spinn3r data. These phrases appeared a total of 201 million times, disregarding multiple instances of the same quotation in the same document. The average number of documents mentioning a quote is 2.69 and the modal number is 1. The phrase with the most mentions appeared in 154,715 documents. Figure 1.1 shows the distribution of the number of documents containing a given phrase. The graph in panel 1.1b is the counter-cumulative distribution function $Pr(X \geq x) = 1 - \mathrm{CDF}(X)$. This is shown rather than the more familiar CDF so that exponentially small probabilities can be represented on a log scale. On the bottom axis is the count x, also in log scale. Since these



(a) Histogram of phrases by number documents in which they appear. The heavy tail is hard to capture on a linear scale.

(b) Counter-cumulative distribution function on a log-log scale. The line is the best fit power law (a.k.a. Pareto) distribution.

Fig. 1.1: Number of copies of quoted phrases appearing on 97M web pages from blogs, social media and news websites indexed by spinn3r.com. 75M distinct phrases appeared a total 201M times. The distribution is heavy-tailed with the top 0.01% of phrases appearing on thousands of documents while the average phrase appears on 2.69. 70.7M phrases appeared five times or fewer while one phrase appeared 154,715 times. The distribution is approximately power law over a wide range, falling a little below is the upper tail.

data appear to fall very nearly on a straight line on a log-log plot, we may suspect their distribution is approximately Pareto (also known as a power-law or scale-free distribution) $Pr(x \geq k) = Ck^{-\rho}$ where $C = x_{min}^{\rho}$. It seems unlikely that the distribution of the "values" of the information associated with these phrases would have such a shape. On the other hand, branching processes can readily give rise to Pareto distributions.

A Pareto distribution was fit to the data using the technique of Clauset et al. (2009) which finds the best fit exponent ρ for each subset of data points excluding more or fewer points from the left, to find the pair of ρ and cut-off x_{min} that maximizes likelihood. The best fitting parameters are $\rho = 1.24$ and $x_{min} = 6$ and this distribution is shown as a dashed line on figure 1.1b. A chi-squared test for goodness of fit cannot reject that the distribution of the data for counts 6 and above is Pareto($\rho = 1.24$,$x_{min} = 6$) with p-value=1. It is common for heavy-tailed distributions to be well described by a Pareto distribution only in their upper tails. Here the Pareto is a very good fit from $x = 6$ up to around $x = 1000$. A candidate distribution that may fit the smallest counts better while matching the shape of the Pareto for larger counts is the Yule-Simon distribution $Pr(x = k) = \rho B(k, \rho + 1)$ (described in section 1.3.3 below). The number of phrases with counts above 1000 falls off somewhat faster than the Pareto distribution. These represent only a small portion of the data: the top 0.01% of all phrases. Nevertheless, an ideal model would reproduce this distribution at all scales. A generalized version of the Yule-Simon distribution may also capture the fall-off from Pareto at the very highest counts.

Figure 1.2 shows evidence that subsets of phrases pertaining to very different topics have similar count distributions. On the left is the distribution of the number of documents mentioning a phrase that contains the word 'Lehman' (which are almost all about the bankruptcy of Lehman Brothers). On the right is the distribution of the number of documents mentioning a

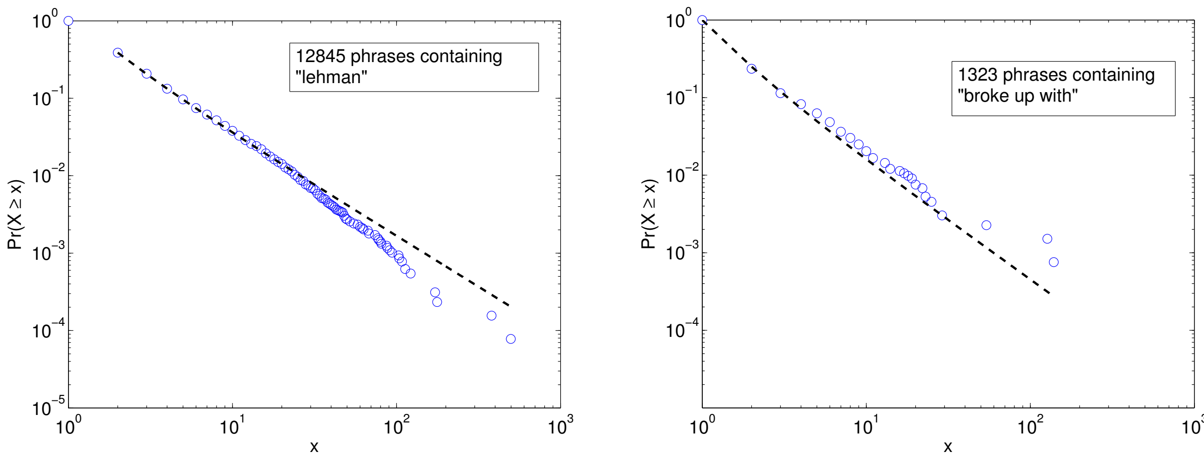


Fig. 1.2: The number documents containing a phrase has similar distribution for different subsets of phrases, an indication that the distribution is unrelated to information content.

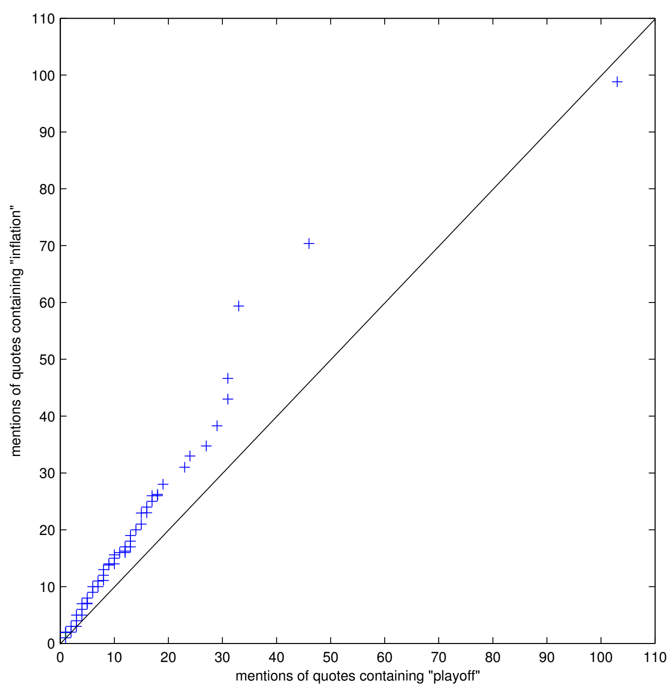


Fig. 1.3: Two-sample QQ plot comparing distributions for subsets of quoted phrases about different topics “playoffs” and “inflation”. The quantiles of the two distributions are proportional (they fall on a straight line) and close to the 45 degree line meaning they are close to identical

phrase that contains 'broke up with', which are mainly celebrity gossip. Another two subsets, phrases containing 'inflation' and phrases containing 'play-off' are compared on a quantile-quantile plot in figure 1.3. The quantiles from the two distributions fall on a straight line, implying that the two distributions differ only in scale (if they fell along the 45 degree line the two distributions would be identical). Why would the distribution of importance of the information associated with sports topics be the same as the distribution of importance for information associated with macroeconomics topics? In fact, all subsets of phrases have the same basic power law shape. One either has to accept that all information has the same distribution of importance regardless of topic, or conclude that the counts of these phrases is not determined by importance.

1.2.2 Arrival analysis

The daily volume of phrases identified by Leskovec et al. (2009) is shown in figure 1.4. There appears to be two regimes, with a break point sometime around the beginning of February 2009. Before this point the volume exhibits a weekly cycle but is otherwise very sable, hovering around 500K mentions per day. After February 2009 the volume more than doubles, but also becomes much more volatile. This latter period is excluded from the analysis of arrivals that follows but is included in the analysis of overall count distribution above. Aside from the change between regimes, there does not appear to be much growth in the overall rate of mentions. The arrival rate of new phrases is also fairly stable at around 280K per day.

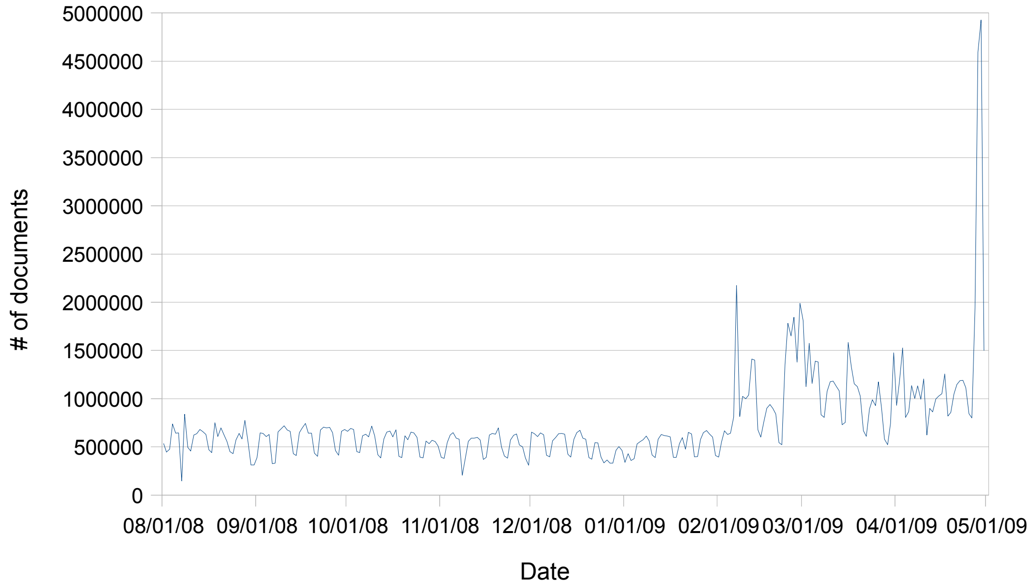


Fig. 1.4: Memetracker data: daily volume of phrases identified by Leskovec et al. (2009) in documents indexed by spinn3r search engine. The first six months of the dataset, up until early February 2009, the volume exhibits a weekly cycle but hovers around 500K per day. In the last three months the volume more than doubles, but also becomes much more volatile. Data after February 1, 2009 are excluded in the following arrival analysis.

The full dataset is too large for convenient computation, so a sample from the dataset was produced by selecting the first 335K phrases whose first mention occurred after October 1, 2008. A date well into the dataset was selected so as to exclude "evergreen" phrases that occur at a certain background level all of the time. The phrases in the sample were not seen in the first two months, and then seen for the first time during the first 29 hours of October. The sample consists of all mentions for this set of phrases up until the end of

January 31, 2009, for a total of 487K mentions. The majority of these phrases, 248K or 74%, were not observed again in the sample. Figure 1.7 shows the number of phrases that receive *at least* n mentions for n up to 10.

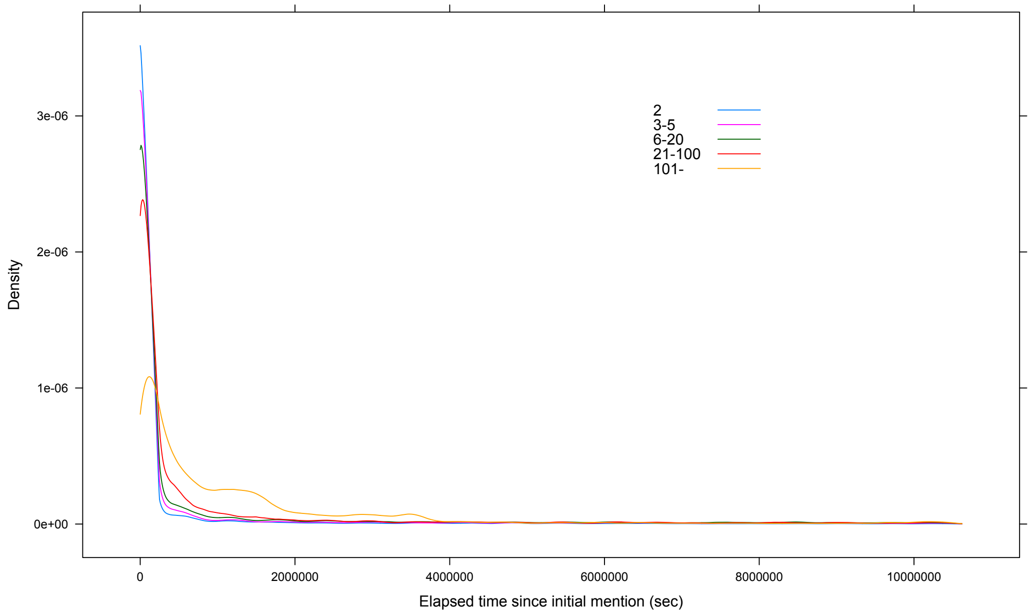


Fig. 1.5: Memetracker data: normalized arrival intensity of additional documents mentioning a phrase after the 1st mention. Phrases are put into five different groups depending on the total number of mentions the phrase receives. Data is from a sub-sample of 335K phrases whose first mention was on October 1, 2008. The window of time is four months. Most mentions occur within a few days of the initial mention. The intensity function is similar for phrases of very different mention counts, indicating it is not longevity that produces high counts.

Figure 1.5 shows the arrival of additional mentions after the initial one. The intensity falls off quickly, with most mentions arriving within a few days of the first. This is true for the phrases that only ever receive a few mentions, but also for phrases that receive hundreds. In the figure, phrases are grouped by the total number of mentions they receive. For example, the 2-mention group contains only phrases that had exactly 2 mentions in the sample. The 21-100 mention group had between 20 and 100 mentions. The groups were chosen to have roughly the same number of *mentions* in each group. There are about as many phrases (43K) in the 2-mention group as in all the other groups combined. The time profiles of all these groups is very similar. The magnitudes are roughly similar as well, although that cannot be seen from this graph since these intensity functions have been normalized (the area under each curve is 1, so they are in fact density functions rather than intensity functions). This figure dismisses the idea that the high count phrases are the ones that continued to spread for a long time.

The distribution of inter-arrival times is shown in figure 1.6. Only the times between the $(n-1)$th and nth mention is shown, for select values of n (e.g. the "1000" curve shows the wait time between the 999th mention and 1000th mention, for all those phrases in the sample that had at least 1000 mentions). The variability of wait times is about the same, regardless of how many mentions have already occurred, although there is some tightening of the distribution for higher order mentions. This graph dismisses the idea that the high count phrases are ones for which new mentions intrinsically arrive more quickly. The first 99 mentions may *happen* to have come quickly, but conditional on already having 99, the arrival of the 100th doesn't look much different from the arrival of the 3rd conditional on the 2nd – except

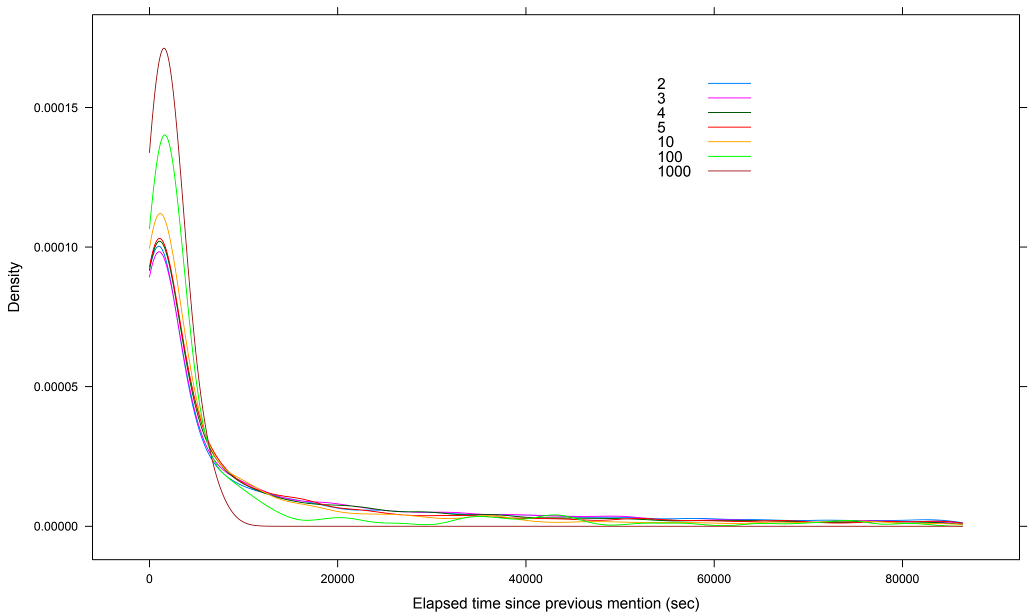


Fig. 1.6: Memetracker data: normalized arrival intensity of n^{th} mention, conditional on phrase already having $n-1$ mentions (sub-sample of 335K phrases). The window of time is one day. The intensity function for the 100th mention is not substantially different from that of the 2nd. Most mentions arrive within 2000 seconds of the previous mention, regardless of how many mentions have come before.

that the overall *magnitudes* are different. These curves are again normalized densities rather than intensities. There are many more 2nd mentions than 100th mentions. However, conditional on having $n-1$ mentions, the arrival intensity actually increases in n.

The contagion hypothesis is that the very fact of having more mentions in social media increases a phrase's likelihood of getting yet another. Additional mentions occur when an individual finds an existing mention and then creates a new document containing the same phrase. Perhaps the rate of new mentions depends *only* on the number of existing mentions. If contagion explained all of the difference in final mention counts, the fraction of phrases appearing two times that go on to appear a third would be twice as great as the fraction of those appearing once that go on to appear a second time. This is in fact the case (see table 1.1). However, the rate to go from three mentions to four is 2.37 times the rate from one to two, less than the predicted 3. The arrival rate of higher order mentions falls increasingly short of the rate predicted by contagion. This seems like evidence of contagion, but contagion may not be the whole story.

Tab. 1.1: $K(n)$ is the number of phrases that have at least n mentions. The contagion model predicts that the rate is proportional to the number of previous mentions.

n	$K(n)$	$\lambda(n) = K(n+1)/K(n)$	$\lambda(n)/\lambda(1)$	predicted $\lambda(n)/\lambda(1)$
1	334592	0.260	1	-
2	87065	0.508	1.95	2
3	44204	0.617	2.37	3
4	27292	0.718	2.76	4
5	19602	0.772	2.97	5
6	15140	0.808	3.11	6
7	12237	0.823	3.16	7
8	10073	0.858	3.30	8
9	8641	0.860	3.30	9
10	7431	...	...	...

1.2.3 Binomial sub-sampling

These relative rates of observed mentions probably do not reflect the true relative rates of mentions in all media. This is because the Memetracker dataset, large though it is, is surely only a small sample of all media. And with count data, the counts in the sample are not proportional to the counts in the population. The spinn3r search engine likely visits only a subset of the web pages where one individual can place information to be found by others. Moreover, a lot of news is passed from person to person online via email, which spinn3r does not collect. And, of course, information can pass from person to person via offline media as well. Suppose our dataset only contains a fraction p of all media mentions. If k is the true number of documents containing a phrase then x, the number we observe, will be a random variable with a binomial distribution. $P(x|k) = \binom{k}{x} p^x (1-p)^{k-x}$. This is called binomial sub-sampling, or "p-thinning". It completely changes the absolute counts, and heavily skews relative counts. If the population of media contains 10 mentions of A and 100 mentions of B, we are much more likely to observe 5 of the 10 A mentions than we are to observe 50 of the 100 B mentions. Our observed ratio of B/A will very likely be less than 10 to 1. Furthermore, some phrases may not be observed at all.

In figure 1.7 I show a simple numerical example of contagion where the number of mentions is thinned by binomial sub-sampling. Parameter values have been chosen so the resulting numbers are close to the data in table 1.1. A reasonably good fit comes from a latent population of 2.5 million phrases of which we only observe mentions for 335 thousand, because the sample only contains 4.5% of all mentions. In the population all 2.5 million are mentioned

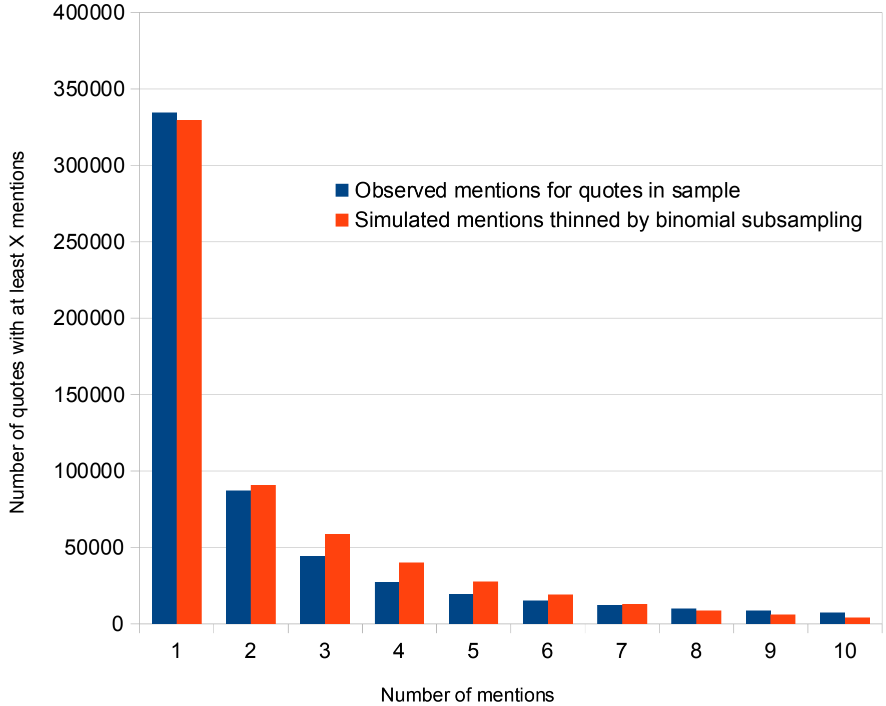


Fig. 1.7: Data vs. simulation: number of phrases observed to have at least n mentions. Data is from a 335K phrase sub-sample of Memetracker dataset. Simulation is a simple branching model thinned by binomial sub-sampling: starting with $k(1)$ distinct phrases initially mentioned once, each mention induces additional mentions at a rate λ, but mentions are only observed with probability p. Fitted values are I=2.5M, λ=0.196, p=0.045.

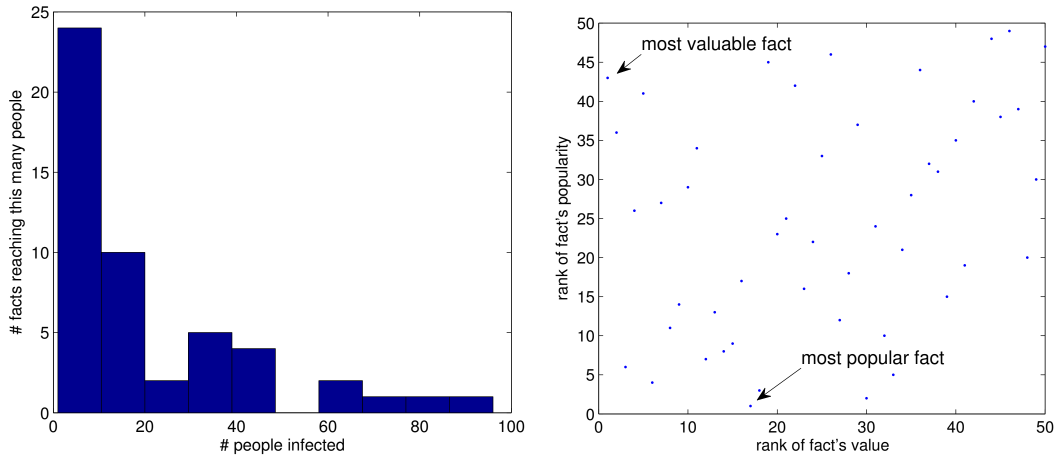
at least once, and of the phrases mentioned at least n times, λn of them are mentioned at least $n+1$ times. The fitted value of $\lambda = 0.196$. Contagion and binomial sub-sampling alone can do a fair job of explaining the observed data.

1.3 Model

Let's go beyond the numerical example in section 1.2.3, and build some models of the spread of information. Imagine the Internet is one giant idealized bulletin board. Agents browse the bulletin board and find messages to read in a random process. Each message contains one "fact", meaning one unit of information.[3] When an agent reads a message she immediately learns the fact it contains. She then decides whether or not to re-post the fact in a new message. All agents are identical except for their information sets. In particular, their rates of finding messages are identical. This means each agent is equally likely to be the next to read a message. Also assume that all messages are identical except for what fact they contain. Moreover, agents cannot tell what fact a message contains without reading the message. This means each message is equally likely to be the next one read.

The number of messages that will come to contain each fact depends on (a) whether and how new facts are added, and (b) when the system is observed. We'll start with a stripped down version of the model in which the number of facts is fixed and the population is finite and small. We allow each fact to have a different value which is exogenously given. The agents in this model are not forward looking, and they do not have rational expectations. But they do preferentially post the higher value facts over lower ones, which lets us see how the feedback effect from contagion can outweigh individual choices. In section 1.4 I will add rational expectations, but at the cost of needing to solve self-consistent dynamic paths for the evolution of the entire probability distribution

[3] I do not mean to imply that information needs to be objective or true, simply that it comes in indivisible units.



(a) Histogram of facts by popularity. Half of the facts infected 1-9 people, while one fact infected 96 people.

(b) The popularity of a fact is not highly correlated with its value

Fig. 1.8: Agent-based simulation of 50 facts spreading through a population of 200 agents.

of counts. That solution would be computationally infeasible with separate distributions for different valued facts. By replacing rational expectations with a moderately sophisticated heuristic we can explore here the relationship between value and prevalence using agent-based modeling.

1.3.1 Agent based simulation

Agents communicate facts to each other by posting them on a bulletin board at some cost. Each period, every agent decides which facts to post from the set of facts she knows. Then agents read posts at random, with every agent equally likely to read each post. The probability a given agent will learn a given fact is proportional to the fraction of posts which contain that fact.

In the following period, old posts are replaced by new ones. Each fact has a different intrinsic value, and agents choose what to post using a rule of thumb: each agent posts the most valuable fact she knows, then second most valuable, and so on, up to the point where the value of the marginal fact is less than the cost of making one more posting. The bulletin board was simulated with two hundred agents and fifty facts, each initially known by one agent.[4] A very unequal distribution of awareness emerges (figure 1.8): after three hundred periods a single fact has been learned by ninety-six agents, while half of the facts has been learned by less that ten agents. There is only a weak correlation between value of a fact and number of agents who came to learn it: the most well-known fact is only the sixteenth most valuable and the most valuable is forty-third out of fifty, learned by only three agents.[5]

1.3.2 Pólya urn process

Keeping the number of facts fixed, now let the population be unbounded. If we eliminate the value and costs, then the sophisticated communication choice becomes a basic rule: for every fact the agent learns, she posts one mes-

[4] Each fact has a value drawn from normal distribution with mean one and standard deviation 0.25 (but constrained to be positive). Every agent faces a cost of posting facts, $c = \beta m^2$ where m is the number of facts posted in a period. The value of $\beta = 0.16$ was chosen so that each agent would rarely post more than four facts at once.

[5] The Pearson correlation coefficient is $\rho = 0.2928$, but since we would not necessarily expect a linear relationship the plot shows the rank of the value versus the rank of popularity. The Spearman rank correlation coefficient is $\rho = 0.3512$. This is significantly different from zero (p-value = 0.0124) according to a permutation test on the simulated data, but the correlation is quite weak.

sage containing that fact as soon as she learns it. Then the number of messages containing each fact can be described as a multivariate Pólya-Eggenberger urn process. The sequence of transmissions and re-postings of facts is equivalent to the sequence of colored balls drawn from an urn when after each draw the ball is replaced along with an additional ball of the same color. Let there be I different colors. Initially the urn contains a_i balls of color i for $i = \{1, \dots I\}$. A ball is drawn at random from the urn and it's color is observed. The ball is then replaced and a new ball of the same color is added to the urn.

For the first draw there are $a = \sum a_i$ balls to choose from. For the second draw there are $a + 1$ balls, and so on until the last draw where there are $a+t-1$ balls to choose from. So the total number of ways to make t draws is $a(a+1)\cdots(a+t-1) = (a+t-1)!/(a-1)!$. Now consider a given sequence of colors $\iota_{\tau=\{1,\dots,t\}}$ where ι_τ is the color of the τth ball drawn. Let $d_i(t)$ be the total number of balls of color i drawn: $d_i(t) = \sum_{\tau=1}^{t} \mathbf{1}(\iota_\tau = i)$ where $\mathbf{1}(\cdot)$ is the indicator function. How many ways can a particular sequence of colors be produced? For each color i, the first time that color is observed there are a_i balls of that color, and so a_i ways that color could have been chosen. For the second time color i is observed there are $a_i + 1$ ways, and so on, up to the $a_i + d_i(t) - 1$ ways to make the $d_i(t)$th observation of color i. So the total number of ways to draw $d_i(t)$ balls of color i is $a_i(a_i + 1)\cdots(a_i + d_i(t) - 1) = (a_i + d_i(t) - 1)!/(a_i - 1)!$. The total number of ways to draw the particular sequence of colors $\iota_{\tau=\{1,...,t\}}$ is: $\prod(a_i + d_i(t) - 1)!/\prod(a_i - 1)!$. The probability of drawing a particular sequence of colors is

$$Pr(\iota_{\tau=\{1,...,t\}} = j_{\tau=\{1,...,t\}}) = \frac{(a-1)!}{(a+t-1)!} \frac{\prod(a_i + d_i - 1)!}{\prod(a_i - 1)!} \tag{1.1}$$

where as above $d_i = \sum_{\tau=1}^{t} \mathbf{1}(j_\tau = i)$.

We would like to know the probability distribution of $\mathbf{x}(t) = (x_1(t), ..., x_I(t))$ where $x_i(t)$ is the number of balls of color i in the urn after t draws. Note that $x_i(t) = a_i + d_i(t)$. There are $t!/\prod d_i!$ different particular sequences of colors $\iota_{\tau=\{1,...,t\}}$ that produce a given list of counts $d_i(t)$, and the probability of drawing a particular sequence is given in (1.1). Putting these together, and using the gamma function $\Gamma(n) = (n-1)!$ for integer values n, yields

$$Pr(\mathbf{x}(t) = \mathbf{k}|\mathbf{a}) = \frac{t!}{\prod(k_i - a_i)} \frac{\Gamma(\sum a_i)}{\Gamma(\sum k_i)} \prod_i \frac{\Gamma(k_i)}{\Gamma(a_i)}. \tag{1.2}$$

This is a Dirichlet compound multinomial distribution, also known as a **multivariate Pólya distribution**. A draw from (1.2) is equivalent to first drawing a probability vector $\mathbf{p}$ from a Dirichlet distribution with parameter $\mathbf{a}$ and then making t draws from a categorical distribution with parameter $\mathbf{p}$ and counting the results for each type. An asymptotic result described in Johnson and Kotz (1977, p378) gives probability distribution of ratios of different colored balls in the urn after many draws:

$$\lim_{t\to\infty} f(\mathbf{x}(t)/t) = Dirichlet(\mathbf{a}). \tag{1.3}$$

This limit result also applies to ratios of pairs of colors:

$$\lim_{t\to\infty} f\left(\frac{x_i(t)}{x_i(t)+x_j(t)}\right) = Beta(a_i, a_j). \tag{1.4}$$

The numbers of balls of other colors do not affect the horse race between color i and color j, since we can simply ignore draws that affect neither of them. Moreover, we can always lump together sets of colors by relabeling (the distribution of white balls in an urn with red and green balls remains the same if the experimenter is color blind and does not distinguish green and red.) The marginal distribution for the number of balls of a single color j, which is beta-binomial:

$$Pr(x_j = k) = \binom{t}{k-a_j}\frac{\Gamma(a_j+b)}{\Gamma(a_j)\Gamma(b)}\frac{\Gamma(k)\Gamma(b+t-k)}{\Gamma(a_j+b+t)} = \binom{t}{k-a_j}\frac{B(k,t-k+b)}{B(a_j,b)} \tag{1.5}$$

where $b = \sum_{i\neq j} a_i$ is the initial number of balls of all other colors. A draw from (1.5) is equivalent to first drawing a probability p from a beta distribution with parameters (a_i, b) and then making t draws from a Bernoulli distribution with parameter p and counting the successes.

If each fact starts out being known by only a single agent, the initial number of messages is always one $a_i = 1\ \forall\, i \in \{1, \ldots, I\}$. Given the interpretation of (1.2) as a multinomial distribution with a probability vector drawn from a Dirichlet, we can see that for $\mathbf{a} = (1, 1, 1, \ldots, 1)$ all probability vectors are equally likely. We are just as likely to choose a point near one of the vertices of the $I - 1$ dimensional simplex of probability space as we are to choose a point near the center. This means it is not uncommon to make draws from (1.2) where one color appears much more often than others. A generalization of the Pólya process allows for more than one new ball to be added each draw. If after drawing and replacing a ball, s new balls of the same color are added, then the distribution is the same as (1.2) if the parameter $\mathbf{a}$ is replaced with $\boldsymbol{\alpha} = s^{-1}\mathbf{a}$. The effect of fractional parameters in the multivariate Pólya distribution is to make unequal outcomes more likely. With $\boldsymbol{\alpha} = (1/s, 1/s, \ldots, 1/s)$ the points near the center of the simplex become less likely as s increases. Thus, modifying the bulletin board model to have agents post multiple messages when they learn a fact is one way to generate highly unequal distributions of counts like those observed in the web spider data. However, as shown below a different modification – introducing a birth process for facts – matches the distribution even better.

The multivariate Pólya distribution is a classic result for "infectious" processes. Yet it holds only approximately for populations with a finite number of susceptible individuals. In situations where the approximation is good, the Pólya result shows us that the distribution after many draws is sensitive to initial conditions. But the urn mechanism is also memory-less in the sense that the probability of each draw depends only on the current number of balls of each type. This suggests it may be fruitful when fitting a Pólya process to time series data to fit a piecewise Pólya process allowing occasional corrections to prevent the error in the approximation from growing large. As a model for contagion, the Pólya distribution overstates the probability of large infections since it ignores the possibility that transmissions could fail because the recipient is already infected. Asymptotic results such as (1.3) should especially be used with caution. In the exact model, it can be seen (in the definition of $\mathbf{\Omega}(\boldsymbol{k})$) that with a finite number of facts I, as $t \to \infty$, $x_i(t) \to N \ \forall i \in \{1, \ldots, I\}$. Eventually all agents will learn all facts and the process stops because no more transmissions can occur.

1.3.3 Birth of facts – Yule-Simon process

The urn mechanism above describes a fixed set of facts. This can be useful as a one-shot model to compare the number of copies of facts that propagate together for a set length of time (measured in transmissions). But we would like a model of ongoing communication. We need to introduce a birth process for facts into the bulletin board model. Suppose that each agent discovers new facts in a Poisson process with rate μ independently of learning facts by reading messages. Upon discovering a fact, an agent again follows the basic rule and posts one message containing the new fact. Meanwhile agents read messages and re-post facts at rate λ. The overall arrival of new facts is a Poisson process with rate $N\mu$, and the total rate for new messages is the sum of the discovery rate and the re-posting rate: $N(\mu + \lambda)$. A new message will contain a new fact with probability $\delta = \mu/(\mu + \lambda)$. The urn scheme can be modified to accommodate new facts: with probability $1 - \delta$ a ball is drawn and replaced with addition just as before, but with probability δ a ball of a new color is added instead. Chung et al. (2003) point out that a Pólya urn model generalized in this way is equivalent to another common model: the Yule-Simon process, also called preferential attachment in the literature on networks.

The Yule-Simon process gives an ergodic probability distribution for the number of messages containing a fact chosen at random. New facts always appear in only one message. Maintaining for the moment an assumption of an infinite population of susceptibles, the number of messages containing an older fact grows without bound. The portion of all facts that appear in exactly x messages decreases when one of them is transmitted, so that it now appears in $x + 1$ messages, and increases when a fact that previously appeared in exactly $x - 1$ messages is transmitted. Suppose that the portions of facts in different numbers of messages are in a stationary state. It can be shown in numerous ways that these portions follow a **Yule-Simon distribution**:

$$Pr(x = k) = \rho \frac{\Gamma(k)\Gamma(\rho + 1)}{\Gamma(k + \rho + 1)} = \rho B(k, \rho + 1) \tag{1.6}$$

where $\rho = 1/(1 - \delta)$ and $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b)$ is the beta function. See Simkin and Roychowdhury (2006) for a comparison of the numerous derivations of this result that have been given over the past hundred years as this mechanism for producing this distribution has been repeatedly rediscovered. Many authors are interested only in the upper tail of the Yule-Simon distribution which is approximately Pareto (power-law) distributed:

$$Pr(x = k) = \rho \frac{\Gamma(k)\Gamma(\rho + 1)}{\Gamma(k + \rho + 1)} \propto \frac{\Gamma(k)}{\Gamma(k + \rho + 1)} \approx k^{-(\rho+1)} \tag{1.7}$$

for large k, where the final term is due to Stirling's approximation.

The Yule-Simon process has been used to explain many different heavy-tailed distributions from the distribution of incomes observed by Pareto to the distribution of the number of species belonging to different biological genera. Barabasi and Albert (1999) have popularized the equivalent "preferential attachment" model as the explanation for the apparent Pareto distribution of hyperlinks between documents on the World Wide Web. The Yule-Simon distribution is a good fit for the distribution of counts observed in web spider data. But without some modification, the model is fundamentally at odds with a contagion process in a population with a finite number of susceptibles. The setup of the Yule-Simon process supposes that facts continue to propagate forever. In the stationary solution, a fact chosen at random can be arbitrarily old and be known by arbitrarily many agents. To be compatible with a finite population a fact must eventually stop spreading.

1.3.4 Birth and death – generalized Yule-Simon

Perhaps the most natural modification of the Yule-Simon process to accommodate a bounded population would be to allow facts to spread as in the exact bulletin board model above, keeping track of the remaining susceptible portion of the population and having the rate of growth slow until finally stopping when the last agent learns the fact. This variation is unsatisfactory because it predicts that the stationary count distribution for the number of messages containing a fact will have a large mass at N. This is contrary to the observed distribution which is smooth over its entire domain, has most of its mass at the low end around one or two counts, and is well approximated by a Pareto distribution for most of it's domain, perhaps falling a little short of Pareto in the very upper tail. A more realistic modification is to suppose that facts stop spreading when they reach a maximum age. Spierdijk and Voorneveld (2009) in an appendix sketch a variation on the Yule-Simon process modified with a cut-off age T_0. For the stationary distribution of observed counts they derive the **generalized Yule-Simon** distribution:

$$Pr(x = k) = \frac{\rho}{1 - \alpha^\rho} \int_0^{1-\alpha} z^{k-1}(1 - z)^\rho dz = \frac{\rho}{1 - \alpha^\rho} B_{1-\alpha}(k, \rho + 1) \qquad (1.8)$$

where α is the inverse of the expected number balls of a color that has been in the urn for T_0 draws, and $B_x(\cdot, \cdot)$ is the (non-regularized) incomplete beta function $B_x(a, b) = \int_0^x z^{a-1}(1 - z)^{b-1} dz$.

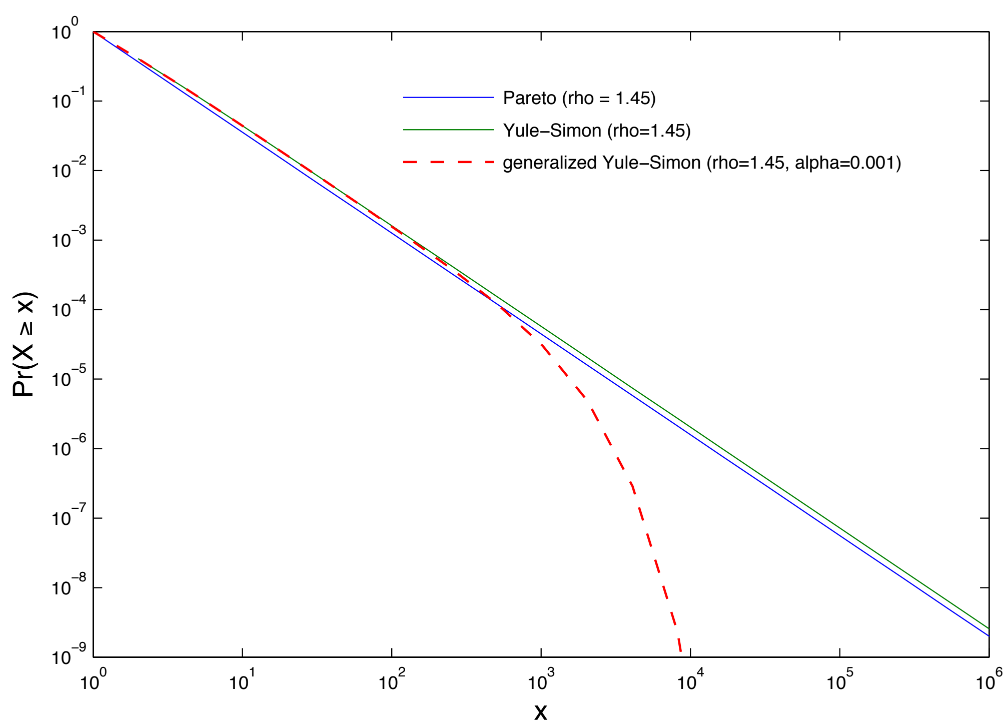


Fig. 1.9: Comparison of generalized Yule-Simon to regular Yule-Simon and Pareto distributions

Figure 1.9 compares the counter-cumulative distribution functions $Pr(X \geq x)$ for the generalized Yule-Simon and unmodified Yule-Simon distribution along with a Pareto distribution all with the same value for $\rho = 1.45$. Since the graph is on a log-log scale the Pareto follows a straight line over it's entire domain $[x_{min}, \infty)$ (in this case $x_{min} = 1$). The Yule-Simon has a somewhat lower probability of single digit counts, though as with the Pareto these account for 90% of the events. For counts above the mid-teens (perhaps 5% of all events) the Yule-Simon exhibits a Pareto tail. The generalized Yule-Simon is practically identical to the unmodified Yule-Simon for the first few decades, but has much fewer counts in the upper tail. With a cut-off parameter $\alpha = 0.001$ the generalized Yule-Simon matches the unmodified for 99.9% of all realizations, but starts to deviate substantially for counts above 500. The generalized Yule-Simon has a *truncated* "heavy-tail": like the Yule-Simon (and Pareto), most realizations are small while realizations thousands of times larger than the mean still occur regularly. But unlike the Yule-Simon, exceedingly large events do not occur. A Yule-Simon process with a cut-off time can capture the salient features of the data and be consistent with contagion in a finite population.

For the bulletin board model to have a cut-off, we not only need agents to stop posting messages containing facts older than T, we also need agents to stop reading them. Otherwise messages from old facts will build up, and in the long run the chance of finding a message with a new fact would go to zero. One way to achieve this is for messages to be removed from the bulletin board over time. Suppose that messages depreciate rapidly – they are covered up or torn down – and agents must continually re-post them if they want them to be found and read. Extending the basic rule for agents: when an agent learns a fact, she posts a message containing that fact, and then maintains it until the fact is age T, after which the message is removed. Some messages only remain on the board for a short time if the fact they contain is already old when the message is posted. No message remains on the board longer than time T. On average the rate of new messages $N(\lambda + \mu)$ must equal the rate of messages being removed $N\mu\,\mathrm{E}[x(T)]$. The expected number of agents to learn a fact before it stops spreading is $\mathrm{E}[x(T)] = (\lambda + \mu)/\mu = 1/\delta$.

1.4 Microeconomic model

Now suppose that instead of following a rule, agents are fully rational. Posting and maintaining a message has some cost. A fact has value for the agent who learns it. To induce an agent who knows a fact to communicate it, there must be some mechanism for transferring value from the agent who learns it to the agent who communicated it. On the Internet this is often achieved by selling advertising. On the other hand, many authors of blogs and users of social media simply value getting the attention of their readers. Somehow the sender and recipient of information split the benefit of communicating. For exposition we will suppose that an agent who posts a message can charge a price p to the agent who finds it. An agent who knows a fact can sell it repeatedly. An agent who buys a fact derives a direct value v from knowing the fact, and she is also able to earn a profit from re-posting it. This "resale value" V of a fact is high when few other people know it. But each new agent who learns a the fact becomes an additional seller. As the number of agents x who know the fact increases, the price that can be charged drops, so V and therefore p are decreasing functions of x. Eventually the expected profit of posting a message falls to zero and the message stops spreading.

If agents had full information about how many other agents knew each fact, then they could keep their messages on the bulletin board until the cost of reaching an additional person outweighed the benefit. With population sized N, this would happen when x/N is such that p times the probability that a given message is the next to be read falls below the cost of keeping a message posted for the expected time until the next transmission. All facts would come to be known by this same portion x/N of the population. But suppose agents do *not* know what other agents know. Agents know only the age of a fact t. Agents have rational expectations: they know the probability distribution over the number of agents who know a fact of age t. The price of a fact can depend only on t, not the actual realization of x which agents cannot observe. Agents will continue to keep messages on the bulletin board until the fact reaches age T such that $p(T)$ is less than the cost.

I consider two different production technologies for posting messages:

- With the "binary" technology each agent can have zero or one messages on the bulletin board for each fact she knows. The cost of keeping each message on the board is c per unit of time (one unit is the expected time until the next message is read).

- With the "intensity" technology each agent can post any non-negative integer number of messages. Maintaining m messages costs $c(m)$ per unit time.

Under either technology, the probability that a given message is the next message read is $1/M$ where M is the total number of messages on the bulletin board. Let t_i be the age of fact i and $x_i(t_i)$ be the number of agents who know fact i when it is age t_i. Let $m(t)$ be the number of messages posted by each agent who knows a fact aged t (zero or one in the binary version, non-negative integers in the intensity version). Then the total number of messages is $M = \sum_i x_i(t_i)m(t_i)$. When a message containing fact i is read, the probability that the agent reading does not already know the fact is $(N - x_i(t_i))/N$.

If agents were to keep communicating a fact long enough, eventually there would be an age $t = \bar{t}$ such that $Pr(x_i(\bar{t}) < N) < \epsilon$ for $\epsilon = c(1)/(Nv)$. That is the age at which the expected profit from communicating is negative even if there are no other messages and the price is the highest conceivable price for a fact (the entire benefit of the fact to society). After this point there is no benefit to keeping a message on the bulletin board. Therefore we can be sure there exists a maximum age $T \leq \bar{t}$ after which posting will stop.

Let v be the direct value of knowing a fact, which an agent enjoys once, immediately upon learning it. Let $V(t)$ be the indirect value of knowing a fact aged t net of v due to the possibility of reselling it one or more times between t and T. Let $W(t)$ be the reservation value of the recipient. This is the value from the opportunity of learning the fact later. The price $p(t)$ splits the expected surplus from reading a message aged t. By the time an agent finds a message, the cost of posting the message has already been sunk, so this cost is not included in the price. The price cannot be contingent on whether the reader already knows the fact, since the poster has no way to verify this after communication has taken place[6]. There is also no way to ascertain whether the reader knows the fact before communication occurs. This is because there is no way to label a fact: identifying a fact fully divulges it. However, the age of a fact is verifiable so the price can be conditional on t. Suppose that the price splits the expected value from reading equally:

$$p(t) = \frac{1}{2}\left(1 - \frac{\mathrm{E}[x(t)]}{N}\right)[v + V(t+1) - W(t+1)]\,. \tag{1.9}$$

The value of knowing a fact due to the possibility of selling it depends on the

[6] We might imagine that the reader could write down every fact she knows and seal this list before reading. She could then refuse to pay if the fact she reads is already on her list. We rule this out by assuming the number of facts an agent knows is too many to list.

number of messages the agents will choose to make (zero or one in the binary version):

$$V(t) = \frac{m(t)}{M}p(t) - c(m(t)) + V(t+1). \tag{1.10}$$

If an agent did not realize the direct value v today, she might realize it later. With probability $1/N$ the agent will be the next to read a message, and with probability $m(t)x(t)/M$ the message she reads will again contain this same fact.

$$W(t) = \frac{1}{N}\frac{m(t)x(t)}{M}\big(v + V(t+1)\big) + \left(1 - \frac{1}{N}\frac{m(t)x(t)}{M}\right)W(t+1) \tag{1.11}$$

Both $v(t)$ and $W(t)$ are decreasing functions since $x(t)$ is monotonically increasing.

1.4.1 Solution for binary model

In the model version with a binary technology for posting messages, each agent chooses for each fact she knows whether or not to keep a message on the bulletin board. The decision is contingent only on the age of the fact. Agents take the total number of messages as given. They make the binary choice for each fact independently since they do not internalize the effect posting one message will have on their chances of transmitting another. For each fact aged

t each agent solves

$$\underset{m\in\{0,1\}}{\operatorname{argmax}}\left\{\frac{m}{M}p(t)-mc\right\} \tag{1.12}$$

and the choice is equivalent to choosing a threshold age T such that $m(t \leq T) = 1$ and $m(t > T) = 0$. T must solve $p(t \leq T) > c(1)M$ and $p(t > T) < c(1)M$. Note that $p(t)$ depends on the entire future path of the expected number of agents who know a fact $\mathrm{E}[x(s)]\ \forall s \in \{t, \ldots, T\}$. The evolution of $\mathrm{E}[x(t)]$ depends on M because the probability a given fact will spread depends on the number of messages being posted about other facts. Furthermore, $M = \sum_i \mathbf{1}(0 \leq t_i \leq T)x_i(t_i)$ where $\mathbf{1}(\cdot)$ is the indicator function, and therefore M depends on the choice of T. This means $\{T, M, p(t), x(t)\}$ need to be solved together.

The key object in the solution to (1.12) is $f(x|t)$ $x \in \{1, \ldots N\}$ $t \in \{0, \bar{t}\}$ the probability mass function for the distribution of $x(t)$. Agents have beliefs about $f(x|t)$ which must be consistent with their choice of T. A nested fixed-point algorithm is used to find $f(x|t)$. Given an initial guess for $f(x|t)$:

1. Solve nested fixed-point algorithm for T.

   Given a guess for T:

   (a) $M = \sum_{t=1}^{T} \delta\, \mathrm{E}[x(t)]$

   (b) Solve $W(t)$ by backwards induction from $W(T + 1) = 0$

   (c) Solve $V(t)$ and $p(t)$ together from $V(T + 1) = 0$

(d) Solve for optimal T'

(e) If $T' = T$ proceed; otherwise repeat with new guess T'

2. Find $f'(x|t)$ determined by M

3. If $\max_t |\mathrm{E}'[x(t)] - \mathrm{E}[x(t)]| < \epsilon$ stop; otherwise repeat with new guess $f'(x|t)$.

The stopping condition 3 for the outer loop depends only on the expected value $\mathrm{E}[x(t)]$ since it is only the expected value that enters into the agent's decision[7].

In the binary choice model, $f(x|t)$ is completely determined by M and parameters N and δ. Let $P_{tk} = Pr(x(t) = k)$. The model assumes initial conditions for each fact $P_{11} = 1$ and $P_{1(k>1)} = 0$. The contagion process then determines the evolution of the distribution:

$$P_{(t+1)k} = \Psi(P_{t(k-1)})P_{t(k-1)} - \Psi(P_{tk})P_{tk} \tag{1.13}$$

where $\Psi(P)$ is the probability that a given fact that is known by a portion P of the population and age $t \leq T$ will be the next fact successfully transmitted.

$$\Psi(P) = \frac{PN}{M}(1-P)\frac{1}{1-\Phi} \tag{1.14}$$

[7] If the updated guess $f'(x|t)$ has the same time path for $\mathrm{E}[x(t)]$ as the previous guess $f(x|t)$ but differs on other moments, iteration should stop anyway. Otherwise the algorithm enters an endless loop since agents will make the same choices as under the previous iteration. An extra check that higher moments of f and f' are close enough can be performed after the stopping condition has been met.

where Φ is the probability the next transmission fails because the reader already knows it. However, Φ itself depends on the entire distribution $f(x|t)$ for every t, and so solution requires another fixed-point algorithm. It can be more straightforward to generate $f(x|t)$ by Monte Carlo simulation.

Figure 1.10 shows the solution of the binary technology version of the model where the distribution $f(x|t)$ was found by simulation. The distribution was generated for each age t, but is only displayed at times $\{50, 150, \ldots, 1950\}$. The solid line shows the (smoothed) mean of the distribution at every t. The cost of posting was tuned such that agents choose $T = 2000$ at which point $\mathrm{E}[x] = 56$. The distribution of the number of agents who know a given fact becomes increasingly skewed as the fact becomes older. At age 1950 when the mean is $\mathrm{E}[x] = 49$, the median is 26 and the mode is 1 while the maximum observed over 50,000 draws was 717.

Figure 1.11 shows the overall count distribution from a snapshot of the bulletin board including facts of all ages $0 \leq t \leq 2000$. The counter-cumulative distribution function $Pr(X \geq x)$ over 50,000 observations is shown on a log-log scale. Visual inspection suggests that a generalized Yule-Simon might describe this simulated data. Using method of moments estimators for parameters ρ and α, figure 1.12 shows a QQ plot comparing the quantiles of simulated distribution against the quantiles of the fitted theoretical distribution. The fit is very good. Only the largest seven observations out of 50,000 lie very far

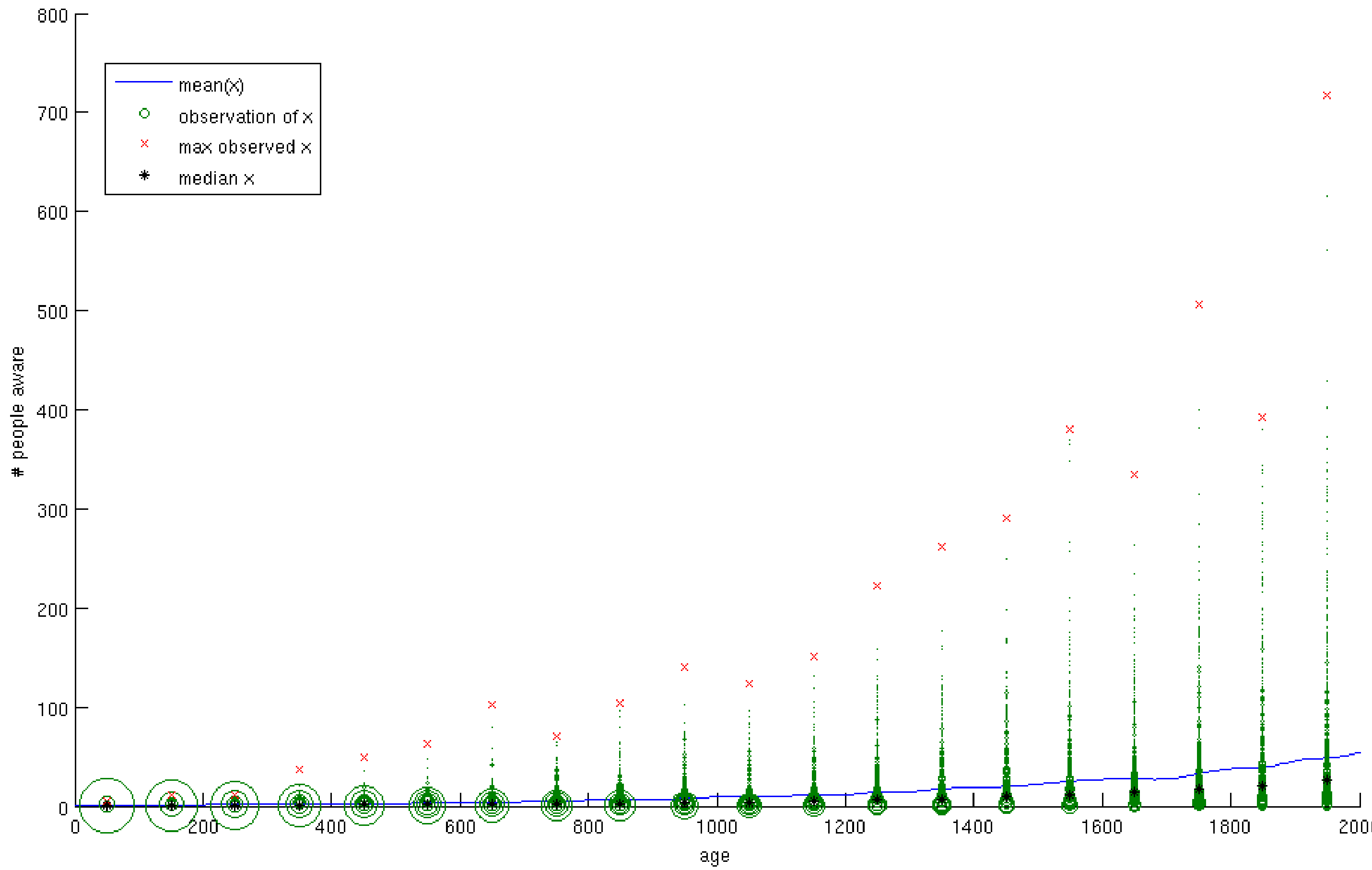


Fig. 1.10: Solution of binary choice model by nested fixed point algorithm. A fact aged t will be known by x agents where the distribution $f(x|t)$ is determined by the contagion process and agents' choice of a maximum age T (here 2000) after which facts are no longer re-posted. Samples from the count distribution at ages $\{50, 150, \ldots, 1950\}$ show it becomes increasingly skewed as facts get older. (Larger circles indicate more mass on a given count.)

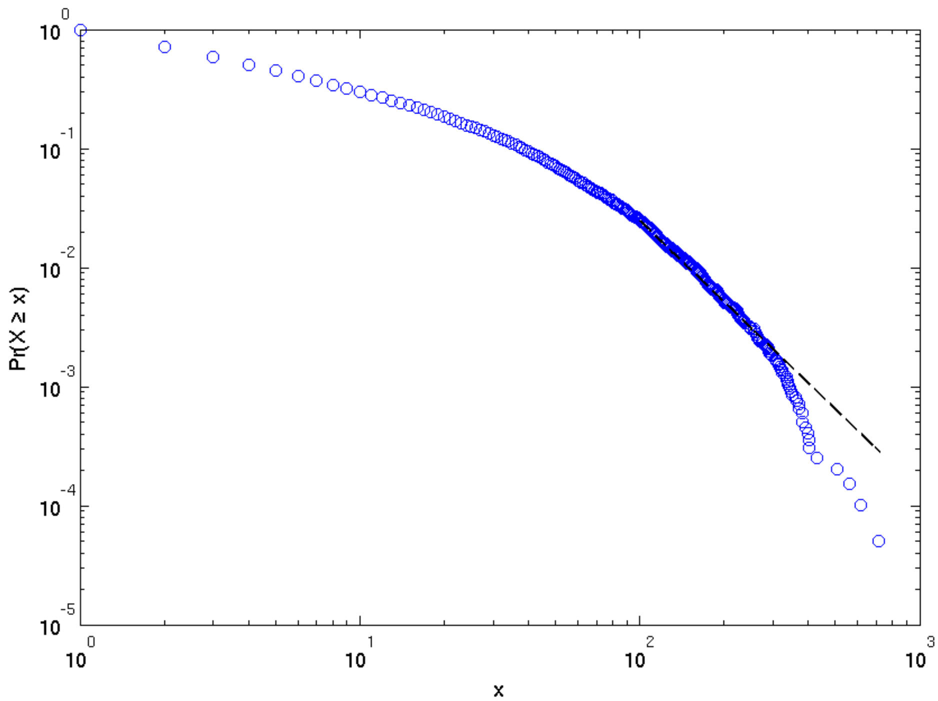


Fig. 1.11: Simulated distribution of number of messages containing a fact under binary choice model

from the 45 degree line. The median of this distribution is two, so the smallest 25,000 observations lie on top of each other on the bottom two points. The two distributions were compared using a Kolmogorov-Smirnov goodness of fit test with bootstrapped quantiles for the test statistic. With a p-value of 0.30, the test did not reject the null that the distributions are the same.

1.4.2 Solution for intensity model

In the model version with an intensity technology for posting messages, each agent chooses for each fact she knows *how many* messages to maintain on the bulletin board. As above, agents take the total number of messages

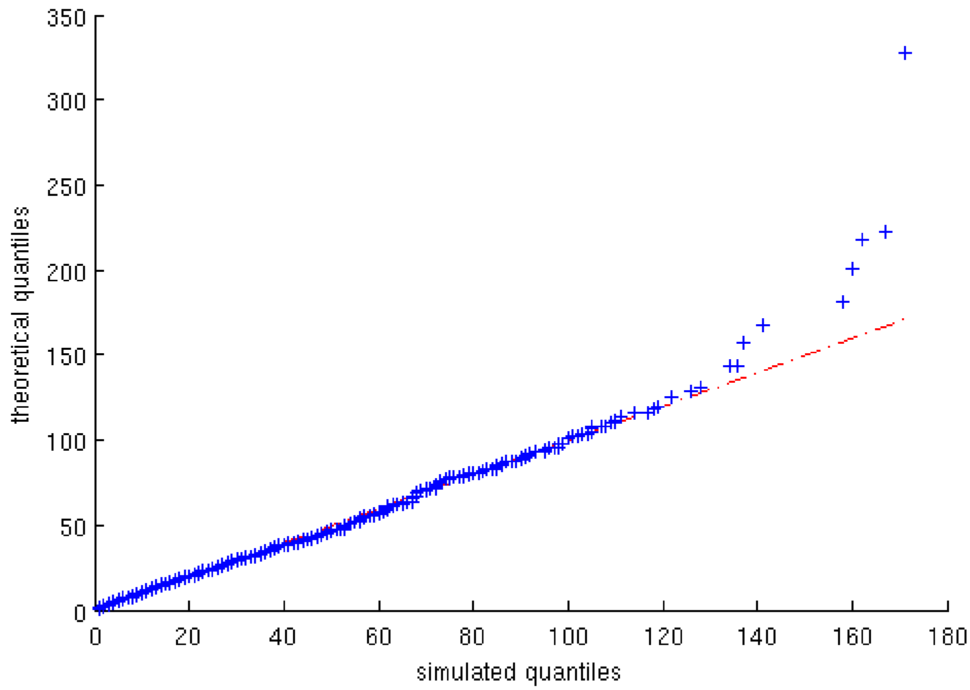


Fig. 1.12: QQ plot comparing binary choice simulated data with generalized Yule-Simon distribution

as given. They face a cost function $c(\mathbf{m})$ for $\mathbf{m} = (m_1, \ldots, m_I)$ where m_i is the number of messages the agent posts containing fact i. For simplicity, suppose that the cost is additively separable in m_i so that $c(\mathbf{m}) = \sum_i c(m_i)$ and agents can solve the problem of how many messages to post for each fact independently. Suppose that $c(\cdot)$ is non-negative, increasing and convex. For each fact aged t each agent solves

$$\operatorname*{argmax}_{m \in \mathbb{N}} \left\{ \frac{m}{M} p(t) - c(m) \right\}. \tag{1.15}$$

As in the binary model, the choice depends on $p(t)$ which is a function of the entire future path of $\mathrm{E}[x(s)]\ \forall s \in \{t, \ldots, T\}$. The evolution of $\mathrm{E}[x(t)]$ depends on both $m(t)$ and on M, where M itself is a function of $m(t)\ \forall t \in \{0, 1, 2, \ldots\}$. This means $\{m(t), M, p(t), x(t)\}$ need to be solved together.

The solution follows a fixed-point procedure similar to that for the binary model. Given an initial guess for $f(x|t)$:

1. Solve nested fixed-point algorithm for $m(t)$.

   Given a guess for $m(t)$:

   (a) $M = \sum_{t=1}^{\infty} \delta\, \mathrm{E}[x(t)] m(t)$

   (b) Define $T = maxt : m(t) > 0$

   (c) Solve $W(t)$ by backwards induction from $W(T+1) = 0$

   (d) Solve $V(t)$ and $p(t)$ together from $V(T+1) = 0$

   (e) Solve for optimal $m'(t)$

   (f) If $m'(t) = m(t)\ \forall t$ proceed; otherwise repeat with new guess $m'(t)$

2. Find $f'(x|t)$ determined by $m(t)$ and M

3. If $\max_t |\mathrm{E}'[x(t)] - \mathrm{E}[x(t)]| < \epsilon$ stop; otherwise repeat with new guess $f'(x|t)$.

The inner loop finds the policy that is consistent with the resulting paths for prices, given beliefs about the proportion of the population of agents that will

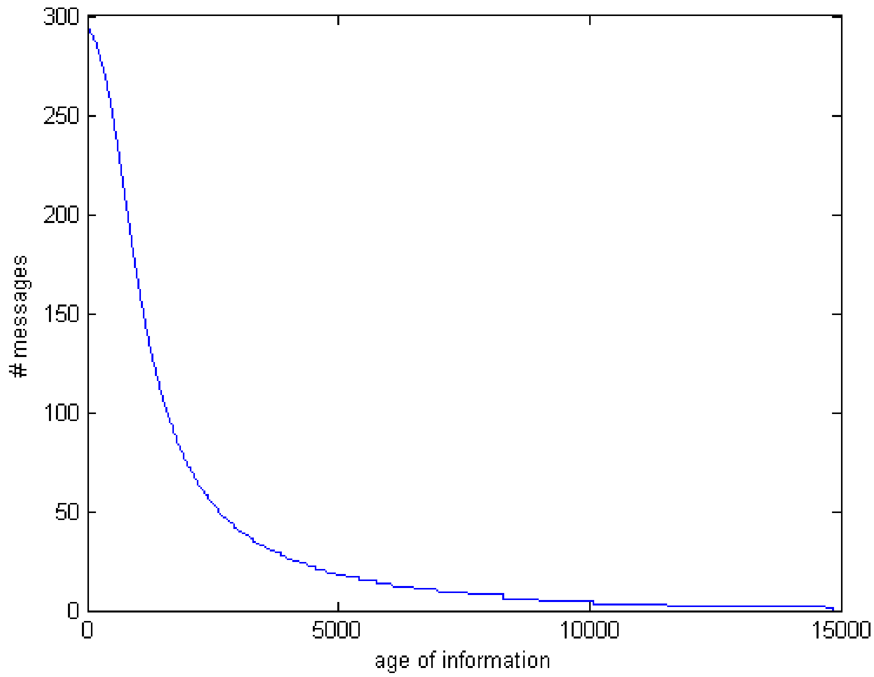


Fig. 1.13: Solution to intensity choice problem: decision rule used by each agent for the number of messages to post for fact aged t

know information of a given age. The outer loop finds the beliefs that are consistent with the resulting policy and price paths.

Assume a logarithmic cost function

$$c(m) = \begin{cases} -ln(1 - \beta m) + c_0, & \text{if } m > 0 \\ 0, & \text{if } m = 0 \end{cases} \tag{1.16}$$

such that an agent will post at most $1/\beta$ messages and c_0 is a fixed cost that can be avoided if the agent posts no messages. A motivation for this cost function might be that posting messages takes time and agents have logarithmic utility from leisure. The parameters were tuned so that agents would choose $T \approx 15000$, that is agents choose $m(t) = 0 \ \forall t > 15000$. Figure 1.13 shows the policy function that solves this example. With these parameters, agents will post about 300 messages containing a fact that was just discovered. As the fact becomes older the *expected* number of other agents who already know it steadily increases. Not only does the chance of a successful transmission decrease (because a reader is more likely to already know the fact), but the value of a successful transmission decreases too, because the resale value of the fact decreases.

The fixed-point solution to this system involves an aggressive policy function that dictates posting many messages for young information and reduces the rate quickly as information ages. For the system to be in equilibrium, the policy function must fall off faster than the growth in the expected number of agents who learn the fact. Without this, the probability that a given agent will learn a given piece of information stays too high over time. This makes the reservation value high because the readers know they can walk away and still learn the same information later from another poster. The price, negotiated between message poster and message reader, collapses and agents end up posting no messages. To support an equilibrium in which messages do get posted, readers must know that they are not very likely to learn the information later. Figure 1.14 shows the resale value, reservation value, price, policy

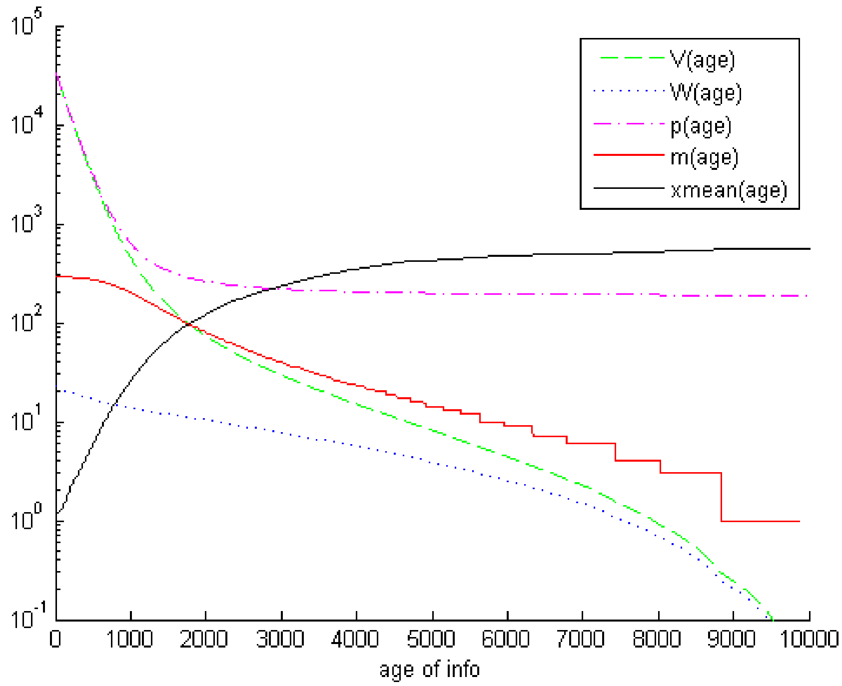


Fig. 1.14: Solution to intensity choice problem: the resale value V, reservation value W, price p, policy function m and resulting mean number of agents who lean the information $xmean$ all as a function of age. Note that the resale value must remain above the reservation value for the system to be in equilibrium.

and resulting mean number of agents who lean the information all as a function of age. Note the essential ingredient for convergence: the resale value is above the reservation value for all ages of fact.

The resulting count distribution of messages per fact is shown in figure 1.15. With optimizing agents incentivized to preferentially transmit the least well known facts but who only observe the age of a fact, the simulation does produce a more equal distribution then the generalized Yule-Simon. But the simulated distribution still has a (truncated) heavy tail. There are 10,000

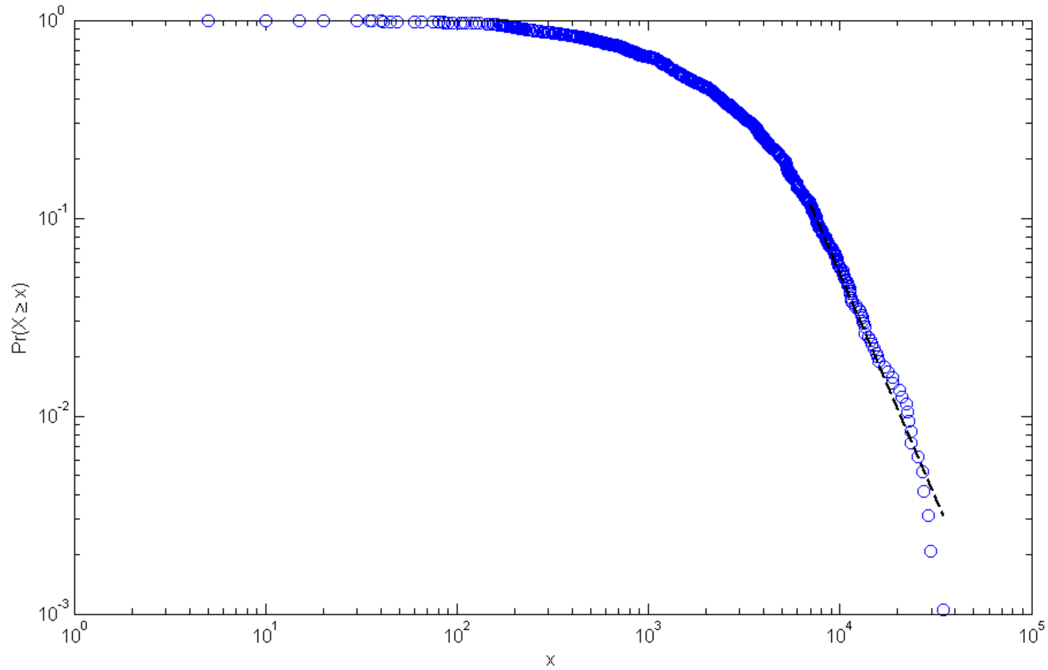


Fig. 1.15: Simulated distribution of number of messages containing a fact under intensity model

agents in this example model. The equilibrium intensity policy transmits half of all pieces of information to over 100 agents – one percent of the population – whereas a much smaller fraction of information reaches that many agents under the model with fixed re-posting rates (the model which generates the generalized Yule-Simon). On the other hand, the equilibrium distribution is a good match to the generalized Yule-Simon *conditional* on having reached a threshold of about 98 people. The upper tails of the two distributions above this threshold are compared in a quantile-quantile plot in figure 1.16. The threshold value of 98 is the expected number of agents who know the information when the equilibrium policy transitions from being very sensitive to age to

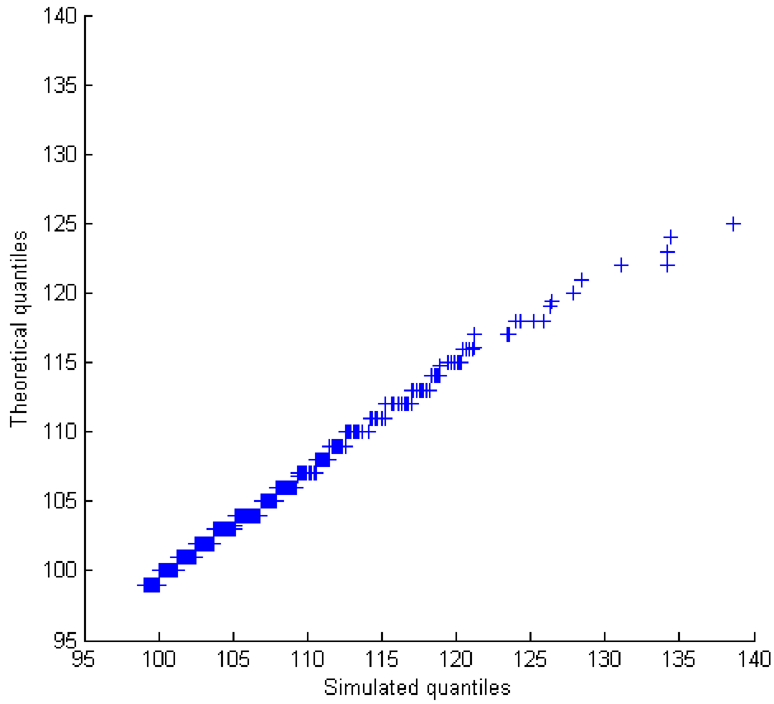


Fig. 1.16: QQ plot comparing the upper tails of simulated vs. generalized Yule-Simon distributions of the number of agents learning a fact. The generalized Yule-Simon is a good fit, but only here in the upper tail.

less sensitive. The total number of messages posted for a piece of information (# posts/agent $\times$ # agents infected) has a sharp peak at this age, as shown in figure 1.17.

The interpretation is that when the equilibrium optimal policy is strongly sensitive to age, agents strongly promote facts that are less likely to be well known. The intermediate outcome for young facts is less unequal. During this phase a significant portion of facts are spread to a significant (though still relatively small) portion of the population. While a few facts still spread very rapidly, it is no longer the case that the large majority get left behind.

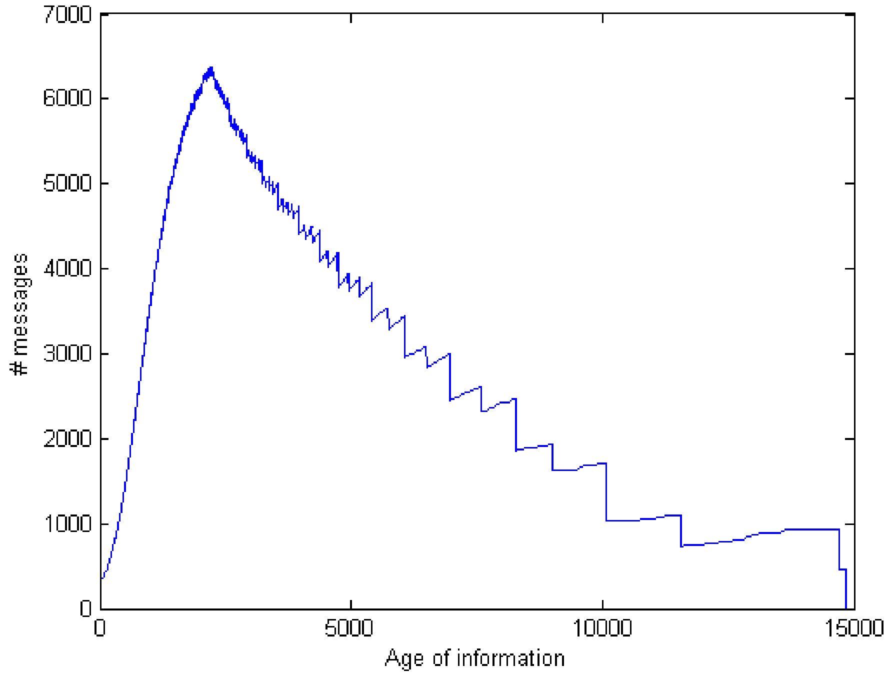


Fig. 1.17: Total number of messages per period containing a fact posted by all agents as a function of the fact's age in the solution to the intensity model. Each agent follows the decision rule shown in figure 1.13. As a fact gets older, more agents learn it but each agent posts fewer messages about it. The peak corresponds to an inflection point in the policy function. After $t = 15000$ the fact stops spreading.

Then, as the policy function becomes less steep and more closely resembles a constant rate, the feedback effect from contagion dominates once again. The further spread of facts is highly unequal.

1.5 Conclusion

Information spreading from person to person rather than from a centralized source, is increasingly common today. According to a recent survey (Pew Research Center, 2011) 41% of Americans now get most of their news on the Internet. Purcell et al. (2010) report that, among Americans who get their news online, 75% get news forwarded through email or posts on social networking sites and 52% share news with others via those means. Consumers are in the midst of a rapid shift to a *distributed* information environment. A majority say it is easier to keep up with the news as a result. On the other hand, 70% of respondents feel "the amount of news and information available from different sources is overwhelming". Information is now abundant but attention is a limited resource. Baresch et al. (2011) argue that, until recently, the allocation of attention was tightly coupled to the allocation of scarce distribution capacity, which was the responsibility of professional editors. As social networks replace broadcast media, this allocation decision has become decentralized and spread across millions of news consumers who are also "news participators" contributing to the creation, commentary and dissemination of news. Half of the respondents in Purcell et al. (2010) say they rely on the people around them to tell them when there is news they need to know.

Like the focus group member quoted at the outset, social media consumers might presume that more important news is more likely to be repeated by their peers and therefore conclude that more important news is more likely to reach them. I argue this is not the case, presenting evidence from social media data as well analytical results from a reduced-form epidemiological model of social media and simulated results from a structural model. The reason that information's importance and prevalence may not be well correlated is not that social media users are biased or that they show a predilection for novelties (though both of these may well be true). Instead I assume that people's communication choices reveal their preferences, so that the "value" of a piece of information can be *defined* as its propensity for being communicated. The problem is this: a greater propensity for re-transmission by each individual does *not* guarantee a higher prevalence in media. This is because viral communication, which is intrinsically uncoordinated and stochastic, also causes a snowball effect: the more people "infected" with a piece of information, the more chances that information has to spread again. This positive feedback amplifies small differences in prevalence which can arise simply from the random order of transmissions. News that is less valuable but that happens to spread early can become much more well known than more valuable news.

The contagion effect has implications for all kinds of aggregate decision making, to the extent that decision makers receive information that has been passed from person to person. Some potential applications in economics include: a *voting model* where many positive and negative facts spread through a population and voters make up their minds based on the subset of facts they learn from their peers; an *asset pricing model* where in lieu of noise traders, news about fundamentals is randomly but unevenly distributed across a population of rational traders; a *sticky price model* with information diffusion in the spirit of Mankiw and Reis (2002) where managers hear different reports about demand via their professional networks. In each case the effect of information spreading by contagion will tend to cause over-reaction to a few pieces of information that happen to become very well known.

As the structural model shows, this probablem is fundamental to social media. Even participants with rational expectations, who take into account the feedback mechanism, will only partially mitigate that feedback. The resulting spread of information will still result in some "lucky" information spreading a lot. Researchers observing the messages from such an equilibrium system will still find a count distribution that has a Pareto-like range over orders of magnitude of counts. Even with fully rational behavior, information that no-one has an especially high propensity to pass along can "go viral".

BIBLIOGRAPHY

Barabasi, Albert-Laszlo and Reka Albert, "Emergence of Scaling in Random Networks," *Science*, Oct. 15 1999, *286* (5439), 509–512.

Baresch, Brian, Lewis Knight, Dustin Harp, and Carolyn Yaschur, "Friends Who Choose Your News: An Analysis of Content Links on Facebook," in "International Symposium on Online Journalism, Austin, Texas" April 2011.

Chung, Fan, Shirin Handjani, and Doug Jungreis, "Generalizations of Polya's urn Problem," *Annals of Combinatorics*, 2003, *7*, 141–153. 10.1007/s00026-003-0178-y.

Clauset, Aaron, Cosma Rohilla Shalizi, and M. E. J. Newman, "Power-Law Distributions in Empirical Data," *SIAM Review*, November 2009, *51*, 661–703. arXiv:0706.1062v2.

Daley, D. J. and D. G. Kendall, "Stochastic Rumours," *IMA Journal of Applied Mathematics*, 1965, *1* (1), 42–55.

Gruhl, Daniel, R. Guha, David Liben-Nowell, and Andrew Tomkins, "Information Diffusion Through Blogspace," in "Proceedings of the 13th International World Wide Web Conference (WWW04)" May 2004, p. 491501.

Johnson, Norman L. and Samuel Kotz, *Urn Models and Their Application*, John Wiley and Sons, 1977.

Leskovec, Jure, Lars Backstrom, and Jon Kleinberg, "Meme-tracking and the dynamics of the news cycle," in "Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining" 2009.

Mankiw, N. Gregory and Ricardo Reis, "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve," *Quarterly Journal of Economics*, 2002, *117* (4), 1295 – 1328.

Mitzenmacher, Michael, "A Brief History of Generative Models for Power Law and Lognormal Distributions," *Internet Mathematics*, 2004, *1* (2), 226–251.

Nekovee, M., Y. Moreno, G. Bianconi, and M. Marsili, "Theory of rumour spreading in complex social networks," *Physica A: Statistical Mechanics and its Applications*, 2007, *374* (1), 457–470.

Pew Research Center, "The State of the News Media 2011, an Anual Report on American Journalism," `http://stateofthemedia.org/2011/` 2011.

Purcell, Kristen, Lee Rainie, Amy Mitchell, Tom Rosenstiel, and Kenny Olmstead, "Understanding the Participatory News Consumer. How internet and cell phone users have turned news into a social experience," http://www.pewinternet.org/Reports/2010/Online-News.aspx March 1, 2010.

Redner, Sidney, "How popular is your paper? An empirical study of the citation distribution," *The European Physical Journal B*, 1998, *4*, 131–134.

Simkin, M.V. and V.P. Roychowdhury, "Re-inventing Willis," January 2006. arXiv:physics/0601192v3.

Simon, Herbert A., "On a Class of Skew Distribution Functions," *Biometrika*, 1955, *42* (3), 425–440.

Spierdijk, Laura and Mark Voorneveld, "Superstars without Talent? The Yule Distribution Controversy," *Review of Economics and Statistics*, 2009, *91* (3), 648–652.

Stelter, Brian, "Finding political news online, the young pass it on," *The New York Times*, 2008, *March 27.* retrieved `http://www.nytimes.com/2008/03/27/us/politics/27voters.html`.

Stumpf, Michael P. H., Carsten Wiuf, and Robert M. May, "Subnets of scale-free networks are not scale-free: sampling properties of networks," *Proceedings of the National Academy of Sciences of the United States of America*, 2005, *102* (12), 4221–4224.

Yule, G. Udny, "A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S.," *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 1925, *213* (402-410), 21–87.

Appendix 1.A Top quoted phrases

Tab. 1.2: Top 30 quoted phrases from spinn3r data

Rank	# mentions	Text of quoted phrase
1.	154715	gang of ten
2.	64660	c tait m gara faubourg de carthage
3.	46510	i love you
4.	42606	dancing with the stars
5.	38973	war on terror
6.	38042	saturday night live
7.	37166	needs to be defeated
8.	36810	the dark knight
9.	35637	the high priest of deregulation
10.	35617	phil gramm is the single most important reason for the current financial crisis
11.	34233	that is equal to the task ahead
12.	34106	abandon all hope
13.	32519	federalist wrote your kidding
14.	31836	die rolle der religion in der modernen gesellschaft
15.	29222	von der welt lernen erfolg durch menschlichkeit und freiheit
16.	28374	palm beach county residents claim your economic stimulus payment
17.	28051	yes we can
18.	25880	meet the press
19.	24278	good morning america
20.	24013	i don't know
21.	22929	daytona beach residents claim your economic stimulus payment
22.	22144	il bloggatore cucina
23.	19983	conditioned on an auto industry emerging at the end of the process that actually works
24.	19388	an inconvenient truth
25.	19344	the curious case of benjamin button
26.	18369	alle kids sind vips
27.	17801	joe the plumber
28.	17405	its like youtube of images
29.	17206	bridge to nowhere
30.	17120	sex and the city

Tab. 1.3: Top 20 phrases containing the word "inflation"

Rank	# mentions	Text of quoted phrase
1.	1988	if the american people ever allow private banks to control the issuance of their currency first by inflation and then by deflation the banks and corporations that will grow up around them will deprive the people of all their property until their children will wake up homeless on the continent their fathers conquered the issuing power of money should be taken from banks and restored to congress and the people to whom it belongs i sincerely believe the banking institutions having the issuing power of money are more dangerous to liberty than standing armies
2.	647	iwpr contends that zimbabwe's currency which has been undermined due to astronomical inflation now faces extinction as citizens abandon it in favor of foreign currencies
3.	468	in the 1980s it was clear to people that the inflation rate was going to come way down and it did
4.	466	there was a sense that we were going through a tough time for a while as a price of getting inflation down and that things would come back up today they can't see any gain from what's going on
5.	448	i believe that banking institutions are more dangerous to our liberties than standing armies if the american people ever allow private banks to control the issue of their currency first by inflation ... [*variation of phrase #1*]
6.	353	i believe that banking institutions are more dangerous to our liberties than standing armies if the american people ever allow private banks to control the issue of their currency first by inflation ... [*slight variation of phrase #5*]
7.	323	doesn't account for inflation
8.	295	is bea measuring growth or inflation
9.	279	in the course of six to nine months many investors and many businesspeople have gone from one extreme fearing rip-roaring inflation to the other extreme fearing the extraordinarily negative consequences of deflation
10.	259	its core business of responsibly managing our money supply and inflation
11.	218	some risk of a protracted period of excessively low inflation
12.	185	most investors are unable to profitably time the market and are left with equity fund returns lower than inflation
13.	144	onslaught of inflation
14.	142	inflation is like little rats under your mattress gnawing away at the corners of your cash
15.	135	indeed we expect inflation to be quite low for some time
16.	132	because the number of dollars andrew and jessica pay each year stays fixed while the rise in prices inflation makes those dollars worth less and less each year in terms of the real goods and services they can buy
17.	118	observed that the market is well-supplied with current commercial oil stocks standing above their five-year average the conference further noted with concern that the current price environment does not reflect market fundamentals as crude oil prices are being strongly influenced by the weakness in the us dollar rising inflation and significant flow of funds into the commodities market
18.	103	la prime la casse les soldes et la d sinflation devraient soutenir la consommation
19.	100	whip inflation now
20.	98	although downside risks to growth remain the upside risks to inflation are also of significant concern

2. SEARCH AND BARGAINING WITH PLASTIC:

MONEY AND CHARGE CARDS AS COMPETING MEDIA

OF EXCHANGE

2.1 Introduction

A high and increasing level of intermediated exchange is a fact of life in modern economies. In the U.S., credit and charge cards are the most frequently used payment instrument, representing 30% of all retail transactions by dollar value. Personal checks and debit cards are second and third, whereas cash transactions make up less than 20%.[1] What does this mean for monetary theory that emphasizes distortions arising from the cost of holding money? Some observers speculate that physical currency will disappear entirely. This may not bother theorists who take a reduced-form approach to money and have all along included demand deposits and other highly liquid assets in their definition. However, a micro-founded theory that seeks to explain exactly why money is needed ought to have something to say about the rise of intermediated exchange.

The money search literature has aimed to build a theory of money from the primitives of the economy by explicitly modeling the role of money as medium of exchange. The benefit of this approach is that such a structural model should be more reliable in predicting behavior when policies or the environment changes. The invention of alternative media of exchange is just such a change. This paper introduces a charge card technology into a search-and-matching model of money. I model charge cards, as distinct from credit

[1] Reported in Levitin (2007) p5, but original source is the industry newsletter *Nilson Report*.

cards, to focus on their use as an alternative medium of exchange rather than the role of credit. I address the following basic questions:

1. If charge cards are costly to process, why do merchants accept them as means of payment? One answer is that consumers will purchase more if they aren't limited by the cash they carry with them. This model formalizes that insight and compares the terms of exchange with and without charge cards.

2. How do charge cards affect welfare? In money-search models the cost of holding money results in an inefficiently low level of consumption. Charge cards can improve welfare by overcoming this problem.

3. Can charge cards and money coexist in equilibrium? Spending limits on cards imply that agents may carry some money too. Also, mixed equilibria are possible in which some agents use cards and others don't.

4. How does inflation affect usage of charge cards? A higher cost of holding money can push buyers to use cards. Since this can improve welfare, the Friedman rule may not be optimal.

I take as my starting point the idea that charge cards enable larger purchases than a consumer's typical money holdings would allow. This addresses a distortion in the real economy caused by the cost of holding money: because of specialization there are large potential gains from trade; but the "double coincidence of wants" problem means that money is often essential for trade, so some gains from trade are never realized if consumers economize on how much money they carry. There are numerous costs to holding money, including the possibility of loss or theft, and the foregone return from other assets. Monetary theory is particularly concerned with the effective tax on money holdings due to inflation. Charge cards avoid these holding costs since consumers do not need to allocate resources in advance of engaging in trade.

To properly analyze the macroeconomic impact of alternative media of exchange, it is necessary to model the microeconomic forces that make a medium of exchange essential. The search-theoretic literature on money does exactly this by explicitly modeling the specialization, decentralization, and double-coincidence problem inherent in exchange. The money search literature finds a relatively high welfare cost of inflation. The inefficiently low level of money holdings is exacerbated for two reasons: first, consumers recognize that money only benefits them if they happen to meet someone selling a good they want, so they carry less money than they would if they were certain to make a purchase; second, when buyers and sellers do meet, the terms of trade are determined by bargaining which takes place after the buyer has sunk the investment in money. This produces a holdup problem which further reduces prospective buyer's willingness to hold money. Charge cards avoid both of these problems, and so can dramatically reduce the cost of inflation.

My model is an extension of the Lagos-Wright framework, which is the current state of the art search theoretic model of money. The Lagos-Wright model alternates between decentralized markets (DM) in which participants are anonymous so that some medium of exchange is required, and perfect Walrasian centralized markets (CM) in which agents can re-balance their money holdings. I introduce an intermediary into the decentralized market (DM): the charge card company. The company can circumvent the DM's problem of anonymity because it has ongoing relationships with its customers. Using a charge card, agents can make purchases in the DM without money and pay the card company in the following centralized market (CM). The CM can be interpreted as the portion of economic activity that is mediated by some institution (e.g. a physical market or an employment contract with a firm), where agents are *not* anonymous and reputation effects or "memory" can function. Charge cards extend this sphere by connecting relationships in the CM to interactions in the DM.

This paper is similar in spirit to Camera (2000) in that we both explore the choice between money and a costly technology that can intermediate exchange. In Camera (2000), the technology matches agents to eliminate search frictions. Agents choose whether to participate in the search sector or to pay a cost (in terms of goods) and participate in a mediated sector. In my version the intermediation is weaker. The market remains decentralized and agents still engage in search, but agents can pay to eliminate the problem of anonymity once random matching occurs.

Several papers have extended the basic model of bilateral exchange with random matching by relaxing the assumption of anonymity to allow some form of credit. Townsend (1983) describes a version of the turnpike model with the innovation of a centralized credit-debit system. Shi (1996) invents an agent-specific asset that can serve as collateral to allow for IOUs in a version of the Kiyotaki-Wright model with divisible goods. Berentsen et al. (2007) starts with a Lagos-Wright framework and introduces banks, which can keep records of agent's financial histories. Telyukova and Wright (2008) extend Lagos-Wright by introducing a third sub-period in which agents are not anonymous, so credit can function, but markets are imperfect so credit is useful. Their aim is to show that agents would rationally save cash for the DM rather than pay off debts.

Most recently, Dong (2007) builds a model of credit cards where agents can choose to use a costly record-keeping technology that allows sellers to extend credit to buyers. Agents' choice between media of exchange in the DM is similar in my model and Dong's. However, our setups differ considerably. Dong seeks to replicate an observed inverse-U shaped relationship between the use of credit and inflation when credit also functions as a medium of exchange. Repayment of credit requires money, and the use of credit in the DM effectively shifts the holding cost of money from buyer to seller. In contrast, I introduce an intermediary which can bypass the cost of money and could potentially replace money, but introduces its own costs. I find that inflation always makes intermediation more attractive.

My paper is complimentary to Berentsen et al. (2007), which is also concerned with the effect of monetary policy on the ability of a financial intermediary to function. Their intermediary is a bank which gives agents a chance to *adjust* their money holdings after learning whether they will be buyers or sellers. Money remains the only medium of exchange, but by paying interest to depositors in the DM, the bank reduces the cost of carrying money. Even though our intermediaries play different roles, we reach the same conclusion: expansionary monetary policy relaxes the constraints that the intermediary needs to impose on its customers. For charge cards to exist, money growth above the Friedman rule is also required for a more basic reason. Charge cards are costly, so agents will only carry them if holding money is more costly.

The rest of this paper is organized as follows. Some facts about charge cards are presented in section 2. Section 3 develops the money-search model with charge cards, and derives agents' money holdings and card-holding/card-accepting decision. Section 4 describes the possible equilibria of the model. Section 5 presents a numerical simulation of the model and measures the welfare gain from charge cards and the cost of inflation when charge cards are in use. Section 6 describes some potential extensions to the model. Finally, section 7 reviews the conclusions for optimal monetary policy.

2.2 Facts about charge cards

The idea of charge cards as a general purpose means of payment arose in the U.S. in the 1950's with the Diners Club card, followed shortly by American Express and Carte Blanche. These predated credit cards by about a decade and continue to operate in competition with credit cards today. While charge cards do allow buyers to defer payment for a short period of time, they are first and foremost intended as a medium of exchange. No interest is charged. The entire balance must be paid every billing cycle. I focus on charge cards because I am interested in the role of payment intermediation as distinct from the role of credit. Over time, charge cards have been giving way to credit cards. However, 30% to 40% of credit card customers are "convenience" users who pay off their entire balance every billing cycle, avoiding finance charges, and so treat their credit cards are though they were charge cards.[2]

Credit cards generally have explicit credit limits. In contrast, the spending limits on charge cards are not stated in advance. Many consumers incorrectly believe that there is no spending limit on charge cards. A more accurate description is that the spending limit is determined and revised at the com-

[2] Chakravorti (2003), p52.

plete discretion of the card company, and is not known to cardholders until a transaction is declined. Here is a description of the spending limit from American Express:

> "The American Express® Preferred Rewards Green Card has no pre-set spending limit which gives you purchasing power that adjusts with your use of the Card. No pre-set spending limit does not mean unlimited spending. Your purchases are approved based on a variety of factors, including current spending patterns, your payment history, credit record, and financial resources known to us."[3]

Since these spending limits are never published, information about them is hard to come by.

Of course, there is a cost to charge cards. In fact, cards are among the most expensive forms of payment.[4] Merchants pay over $36Billion per year in fees to accept credit and charge cards in the U.S. alone.[5] Such fees, called "merchant discounts" because they appear as the difference between the sales

[3] American Express website http://www.americanexpress.com/getthecard/learn-about/Preferred-Rewards-Green [Accessed 7 Dec 2008]

[4] The average cost of a charge card transaction is 72¢, twice as expensive as clearing checks or PIN debit cards. Although we typically model accepting cash as costless, this is not really the case: on average the cost of handling cash is 12¢ per transaction. Levitin (2007), p1.

[5] Levitin (2007), p2. If we add in debit cards the fees paid by sellers to use payment cards approach $50Billion. According to Aneace Haddad, the interchange industry is bigger than the biotech industry, the music industry, the microchip industry, the electronic game industry, Hollywood box office sales, or worldwide venture capital investments ("The Interchange Industry Is Bigger than ...", *Aneace's BLOG*, May 12, 2006, at http://aneace.blogspot.com/).

price and the amount the acquiring bank pays the retailer, have both a fixed component, typically 5¢ - 25¢ per transaction, and a proportional component, between 1% and 3% of the sale.[6]

Merchants are prohibited from passing these costs directly on to the customers who pay with cards. Card company merchant agreements include no-surcharge rules which stipulate that cardholders cannot be charged higher than the posted price. Until 1984 this restriction was also enshrined in federal law, the 1968 Truth in Lending Act. In addition, twelve states have laws prohibiting payment card surcharges. The no-surcharge rule, along with other merchant restraints, have been the basis of numerous lawsuits and have partially motivated anti-trust investigations, but remain in effect in some form for all the major card companies.[7]

Historically, card companies have also imposed an annual fee on cardholders. Most credit cards no longer charge this fee, but it is still common for charge cards. Currently, the annual fees on American Express charge cards range from \$95 to \$450.[8] However, the merchant discount is a more important source of revenue for charge card companies than cardholder fees. American Express, which now also issues credit cards but is still a major issuer of charge cards, makes 72% of its revenue from merchant discount fees and less than 12%

[6] Levitin (2007), p9

[7] For a detailed history of merchant restrictions see Levitin (2007) p21-32.

[8] American Express website `http://www.americanexpress.com/getthecard/compare-cards/no-limit-cards` [accessed 7 Dec 2008]

from annual fees.[9] Nevertheless, annual fees remain important in consumers' choice over whether to become a cardholder.

The use of all forms of payment cards has been on the rise since 1993, but the strong gains for credit and charge cards, debit cards, and other electronic payment systems have come mainly at the expense of personal checks, which have seen a long decline from a high of nearly 60% of transactions by dollar value in the mid-1990's to about 20% today.[10] It is interesting to note that cash usage has been fairly stable for a decade or more. Perhaps predictions of a cashless economy are premature. This motivates the search for a microfounded model where money and intermediated exchange can coexist.

2.3 Model

2.3.1 Lagos and Wright environment

As in Lagos-Wright, each period is divided into two sub-periods: the DM and the CM. Agents do not discount between the DM and subsequent CM. In the DM each agent produces a special good, while in the CM every agent produces the same general good. Every agent consumes the general good, but each agent only consumes a subset of the range of special goods, and doesn't

[9] Levitin (2007), p17. By contrast, Discover, a firm that like American Express faces both merchants and consumers (unlike card networks such as Visa and Mastercard) but whose portfolio is purely credit cards, earns 3/4 of its revenue from interest and only 1/4 from merchant discount fees.

[10] Levitin (2007) chart 1.

consume the good she herself can produce. Goods are not storable. In the CM there is a perfect Walrasian market where agents are price takers. In the DM agents meet randomly and anonymously and they bargain over the terms of exchange. With probably σ an agent will encounter another agent who produces a special good she likes and she becomes a buyer. Also with probably σ she will encounter an agent who likes the special good that she can produce and she becomes a seller. Assume for simplicity that there are no double-coincidence meetings.

The production technologies for both special goods and the general good are linear in labor, but agents have different disutilities from labor in the two sub-periods. Let x and X be consumption of special and general goods respectively, and h and H be labor supplied in the DM and CM. Utility is then

$$\mathcal{U}(x, X, h, H) = u(x) - c(h) + U(X) - H \tag{2.1}$$

where $u(0) = c(0) = 0$, $u \geq 0$, $u' \geq 0$, $u'' \leq 0$, $U \geq 0$, $U' \geq 0$, $U'' \leq 0$, $c \geq 0$, $c' \geq 0$ and $c'' \geq 0$. Assume there exists a quantity $q^* > 0$ such that $q^* = \operatorname{argmax} u'(q) - c'(q)$.

Fiat money exists in the CM and any amount can be bought or sold at the going price of ϕ units of general good. The amount of money per-capita is M_t at the start of each period. During the CM, a monetary authority injects a lump-sum transfer of new money such that $M_{t+1} = \gamma M_t$. Take γ to be constant and consider only steady-state equilibria where $\gamma = \phi_t/\phi_{t+1}$. Let T be the per-capita transfer in terms of the general good.

2.3.2 *Charge cards*

In each CM agents can make agreements with the charge card company to use charge cards to intermediate exchange in the following DM. There are two kinds of agreements: agents can become cardholders, and agents can make arrangements to accept charge cards. When two agents meet in the DM, if the buyer is a cardholder *and* the seller accepts cards, then the bargaining between them can include a payment d_c made using the card. Let Π_b be the probability that a randomly chosen agent will be carrying a charge card in the DM, and Π_s be the probability that she accepts cards in exchange. The card company will establish a spending limit l on cardholders, so that d_c cannot be greater than l.

To acquire a charge card an agent pays the card company an "annual" fee (paid each period) of η^b units of general good. Also, if the agent made a purchase with a card in the previous DM, she must pay the full balance on her account (also in terms of general goods). I assume the card company does not charge any usage fee to the cardholder, so the company's claim on the cardholder is just d_c.[11] The card company will not issue a card to an agent

[11] To my knowledge, no card company has ever charged its cardholders a transaction fee on purchases. However, it is very common for cards to *pay* cardholders cash-back rewards or other benefits (e.g. frequent flier miles) in proportion to their purchases. In 2005, 80% of

who has ever failed to pay a claim or broken a merchant agreement (see below) in the past.

In order to be able to accept cards in payment in the following DM, the agent makes an agreement with the card company and pays η^s in fees and other set-up costs.[12] The merchant agreement stipulates rules the agent must follow should she become a seller in the DM. For now this includes only the fairly innocuous rule that after bargaining with the buyer the seller must indeed accept the card in payment. The card company will refuse to allow the agent to accept cards if the agent has ever broken a merchant agreement or failed to pay a cardholder claim in the past. If the agent was a seller and accepted a card in the previous DM, the card company pays the agent $(1-\tau)d_c$ units of general good, whether or not the buyer pays her claim, where d_c is the amount she agreed upon in bargaining with the buyer and τ is the merchant discount fee.[13] This fee causes a wedge between buyer's marginal utility and sellers marginal cost when charge cards are used. The second-best quantity is $\tilde{q}$ which solves $c'(\tilde{q}) = (1-\tau)u'(\tilde{q})$. We shall see that if the spending limit

credit card transactions in the U.S. were made with rewards cards (Levitin, 2007, chart 4). A very interesting exercise would be to allow a negative proportional fee for the cardholder.

[12] Setup fees paid to the card company do not seem to be significant in reality, but sellers may face some other setup costs (e.g time costs). The only role for η^s in the model is to give agents a reason they might not accept cards, so η^s can be viewed as a stand-in for other reasons merchants might decide not to accept cards. This is really only interesting when it makes sellers indifferent between accepting cards and not, leading to a mixed equilibrium.

[13] To be fully realistic, the merchant fee should have both a fixed and proportional component. In this model a fixed, per-transaction fee for sellers would have the same function as τ in that both reduce the surplus from exchange, but their cost is divided between the buyer and seller in bargaining over payment method. In an extension where sellers have heterogeneous costs a fixed fee would have different effect on different sellers.

is high enough, agents will always transact $\tilde{q}$ if a card is used. Let l^* be the minimum spending limit under which $\tilde{q}$ is always attainable (i.e. the limit is high enough that it never binds.)

In exchanges where a card cannot be used, payment can be made using *only* money. The agent's gain from trade would be $B_o(s, \breve{s})$ if she is the buyer or $S_o(\breve{s}, s)$ if she is the seller. Here s represents the set of relevant variables describing the agent and $\breve{s}$ is the set of variables describing her trading partner. Let $W(z, y)$ represent the value of entering the CM with real money balances z and a claim on the card company of y general goods (where y is negative if the card company has a claim on the individual.) Then the gain from trading in the DM is the value from trading minus the value of not trading. If only money is used then

$$B_o(s, \breve{s}) \equiv u\big(q_o(s, \breve{s})\big) + W\big(z - d_o(s, \breve{s}),\ 0\big) - W\big(z,\ 0\big) \tag{2.2}$$

$$S_o(\breve{s}, s) \equiv -c\big(q_o(\breve{s}, s)\big) + W\big(z + d_o(\breve{s}, s),\ 0\big) - W\big(z,\ 0\big) \tag{2.3}$$

where q_o is the quantity of good exchanged and d_o is the real value of money the buyer gives the seller. Note that the cost of holding money is a sunk cost in the DM and does not enter the gain.

When a card can be used, then in principle payment might be made with a combination of *both* money and card. Write the gains from trade in the DM as $B_b(s, \breve{s})$ and $S_b(\breve{s}, s)$.

$$B_b(s, \breve{s}) \equiv u\big(q_b(s, \breve{s})\big) + W\big(z - d_m(s, \breve{s}),\ -d_c(s, \breve{s})\big) - W\big(z,\ 0\big) \tag{2.4}$$

$$S_b(\breve{s}, s) \equiv -c\big(q_b(\breve{s}, s)\big) + W\big(z + d_m(\breve{s}, s),\ (1-\tau)d_c(\breve{s}, s)\big) - W\big(z,\ 0\big) \tag{2.5}$$

where q_b is the quantity exchanged, d_m is the real money paid to the seller and d_c is the promise the buyer makes to the card company. The seller gets a claim on the company of $(1-\tau)d_c$.

During the CM, the agent must decide how much cash to take into the DM. She must also decide whether to be a cardholder in the following period and whether to accept cards in the following period. If she chooses to do *neither* her expected value from entering the DM holding real balances z is

$$V^{\text{neither}}(z) = W(z, 0) + \sigma \int B_o(s, \breve{s}) dF(\breve{z}) + \sigma \int S_o(\breve{s}, s) dF(\breve{z}) \tag{2.6}$$

where $F(\breve{z})$ is the distribution of real balances of other agents in the population.

If she chooses to *hold* a card, but not to accept cards then her expected value is

$$
\begin{aligned}
V^{\text{hold}}(z) = W(z,\ 0) + \sigma \Bigg\{ \Pi_s \int B_b(s, \breve{s})\, dF(\breve{z}|\breve{\pi}_s = 1) \\
+ (1 - \Pi_s) \int B_o(s, \breve{s})\, dF(\breve{z}|\breve{\pi}_s = 0) \Bigg\} \\
+ \sigma \int S_b(\breve{s}, s)\, dF(\breve{z})
\end{aligned}
\tag{2.7}
$$

where $\breve{\pi}_s \in \{1, 0\}$ represents the card-accepting choice of the particular trading partner the agent happens to meet in the DM (1 if the partner accepts cards).

If the agent chooses to *accept* cards but not carry one then

$$
\begin{aligned}
V^{\text{accept}}(z) = & W(z,\ 0) + \sigma \int B_o(s, \breve{s})\, dF(\breve{z}) \\
& + \sigma \Bigg\{ \Pi_b \int S_b(\breve{s}, s)\, dF(\breve{z}|\breve{\pi}_b = 1) \\
& + (1 - \Pi_b) \int S_o(\breve{s}, s)\, dF(\breve{z}|\breve{\pi}_b = 0) \Bigg\}
\end{aligned}
\tag{2.8}
$$

where $\breve{\pi}_b \in \{1, 0\}$ represents the cardholding choice of her trading partner (1 if the partner is a cardholder).

If the agent chooses to *both* hold and accept

$$\begin{aligned}
V^{\text{both}}(z) = W(z,\ 0) &+ \sigma\Bigg\{\Pi_s \int B_b(s, \breve{s})\, dF(\breve{z}|\breve{\pi}_s = 1) \\
&\qquad + (1 - \Pi_s) \int B_o(s, \breve{s})\, dF(\breve{z}|\breve{\pi}_s = 0)\Bigg\} \\
&+ \sigma\Bigg\{\Pi_b \int S_b(\breve{s}, s)\, dF(\breve{z}|\breve{\pi}_b = 1) \\
&\qquad + (1 - \Pi_b) \int S_o(\breve{s}, s)\, dF(\breve{z}|\breve{\pi}_b = 0)\Bigg\}
\end{aligned} \tag{2.9}$$

Each possibility could result in a different choice of X, H, and next period's real balances $\hat{z}$. The agent solves all four problems and chooses the option with the highest expected payoff.

$$\begin{aligned}
W(z, y) = \max\Bigg\{ &\max_{X,\hat{z}}\Big[U(X) - X + T + z + y - \gamma\hat{z} + \beta V^{\text{neither}}(\hat{z})\Big]\,, \\
&\max_{X,\hat{z}}\Big[U(X) - X + T + z + y - \gamma\hat{z} - \eta^b + \beta V^{\text{hold}}(\hat{z})\Big]\,, \\
&\max_{X,\hat{z}}\Big[U(X) - X + T + z + y - \gamma\hat{z} - \eta^s + \beta V^{\text{accept}}(\hat{z})\Big]\,, \\
&\max_{X,\hat{z}}\Big[U(X) - X + T + z + y - \gamma\hat{z} - \eta^b - \eta^s + \beta V^{\text{both}}(\hat{z})\Big]\Bigg\}
\end{aligned} \tag{2.10}$$

In each part of (2.10) I have substituted the appropriate budget constraint for H. Three conclusions follow immediately from this setup of the agent's decision:

Lemma 1. *In the CM, consumption is constant $X = X^*$. The agent's card-*

holding and card-accepting decisions and choice of $\hat{z}$ are independent of the state. $W(z, y)$ *can be written as*

$$W(z, y) = z + y + W_0 + \beta W_1 \tag{2.11}$$

where

$$W_0 = U(X^*) - X^* + T$$

$$W_1 = \max\left\{ \max_{\hat{z}} \left[-(1+i)\hat{z} + V^{neither}(\hat{z}) \right], \right.$$
$$\max_{\hat{z}} \left[-(1+i)\hat{z} - \eta^b/\beta + V^{hold}(\hat{z}) \right],$$
$$\max_{\hat{z}} \left[-(1+i)\hat{z} - \eta^s/\beta + V^{accept}(\hat{z}) \right],$$
$$\left. \max_{\hat{z}} \left[-(1+i)\hat{z} - \eta^b/\beta - \eta^s/\beta + V^{both}(\hat{z}) \right] \right\}$$

Proof. From the first order conditions, $1 = U'(X)$ and $\gamma = \beta V'(\hat{z})$ for whichever V corresponds to the agent's card choices. The state variables z and y affect the agent's payoff in the same way regardless of the cardholding and card-accepting decisions, so they do not influence those decisions. Then simply rearrange (2.10) and note that $\gamma/\beta = \gamma(1 + r) = 1 + i$ to arrive at (2.11). □

Using Lemma 1 the gains from trade can be simplified:

$$B_o(s, \breve{s}) = u\big(q_o(s, \breve{s})\big) - d_o(s, \breve{s}) \tag{2.12}$$

$$S_o(\breve{s}, s) = -c\big(q_o(\breve{s}, s)\big) + d_o(\breve{s}, s) \tag{2.13}$$

$$B_b(s, \breve{s}) = u\big(q_b(s, \breve{s})\big) - d_m(s, \breve{s}) - d_c(s, \breve{s}) \tag{2.14}$$

$$S_b(\breve{s}, s) = -c\big(q_b(\breve{s}, s)\big) + d_m(\breve{s}, s) + (1 - \tau)d_c(\breve{s}, s) \tag{2.15}$$

To go further, we need to describe how the quantities and payments exchanged are determined. In most of the money search literature, the terms of exchange are determined by Nash bargaining. This feature has important ramifications for predictions of money search models generally.

2.3.3 Bargaining

Lagos and Wright find a high welfare cost of inflation when the mechanics of money as a medium of exchange are taken into account. This result is due to the bargaining that buyers and sellers engage in when an exchanging goods for money. If there is a positive nominal interest rate, then carrying money into the DM imposes a sunk cost on buyers. But, by allowing exchange to occur, carrying money benefits both buyers and sellers. This leads to a hold-up problem. Since the cost is sunk, it is not shared between the seller and buyer in bargaining. But the surplus from trade *is* shared, and the size of that surplus depends on z, the money holdings of the buyer. Therefore instead of the socially optimal amount z^*, the buyer only brings the amount that sets her cost of holding money equal to her share of the surplus. This hold-up problem is eliminated only if buyers have all of the bargaining power.

However, Lagos and Wright find that even at the Friedman rule buyers carry too little money. This cannot be the result of a hold-up problem, since with a zero nominal interest rate there is no sunk cost to carrying money. Aruoba et al. (2007) explain that this is due to the use of the Nash bargaining solution. When agents choose z, they are effectively choosing the size of the surplus over which they will bargain should they become buyers in a random match. The Nash solution is *non-monotonic* in the sense that the outcome for a given party does not increase monotonically with the size of the bargaining set, and may actually *decrease*. Buyers can improve their bargaining outcome by bringing less cash to the table, even though this is socially inefficient. By contrast, when a charge card is used, the bargaining set is determined by the merchant discount and spending limit which the buyer takes as given. Money has a strategic value to buyers that charge cards lack.

I develop my model using either the Nash bargaining solution or the proportional bargaining solution where there is no strategic incentive for agents to limit their money balances.

Nash bargaining with money only

In a single-coincidence meeting where charge cards cannot be used, either because the buyer is not a cardholder and/or the seller does not accept cards, bargaining over the quantities of special good and money exchanged takes place exactly as in the Lagos-Wright model: $q_o(s, \breve{s})$ and $d_o(s, \breve{s})$ solve

$$\max B_o^{\theta} S_o^{1-\theta} = \max_{q_o, d_o} \{u(q_o) - d_o\}^{\theta} \{-c(q_o) + d_o\}^{1-\theta} \quad \text{subject to } d_o \leq z \tag{2.16}$$

where z is the real money balance the buyer is carrying, and θ is the relative bargaining power of the buyer. Lagos and Wright show that the solution is

$$\begin{aligned} q_o &= \begin{cases} \hat{q}(z) & \text{if } z < z^* \\ q^* & \text{if } z \geq z^* \end{cases} \\ d_o &= \begin{cases} z & \text{if } z < z^* \\ z^* & \text{if } z \geq z^* \end{cases} \end{aligned} \tag{2.17}$$

where q^* is the optimal quantity that solves $c'(q) = u'(q)$ and z^* is the amount of real money that makes exchange of q^* possible.[14] If the buyer brings $z < z^*$,

[14] To see this, first consider the case where the constraint does not bind. Then necessary and sufficient conditions for the solution are

$$(1-\theta)[u(q_o) - d_o]c'(q_o) = \theta[-c(q_o) + d_o]u'(q_o) \tag{2.18}$$

and

$$(1-\theta)[u(q_o) - d_o] = \theta[-c(q_o) + d_o] \tag{2.19}$$

the outcome is the buyer gives the seller all of her money, and $q_o = \hat{q}(z)$ where $\hat{q}$ is the quantity that solves $g(\hat{q}) = z$ where g is defined as

$$g(q) \equiv \frac{\theta c(q)u'(q) + (1-\theta)u(q)c'(q)}{\theta u'(q) + (1-\theta)c'(q)} \tag{2.20}$$

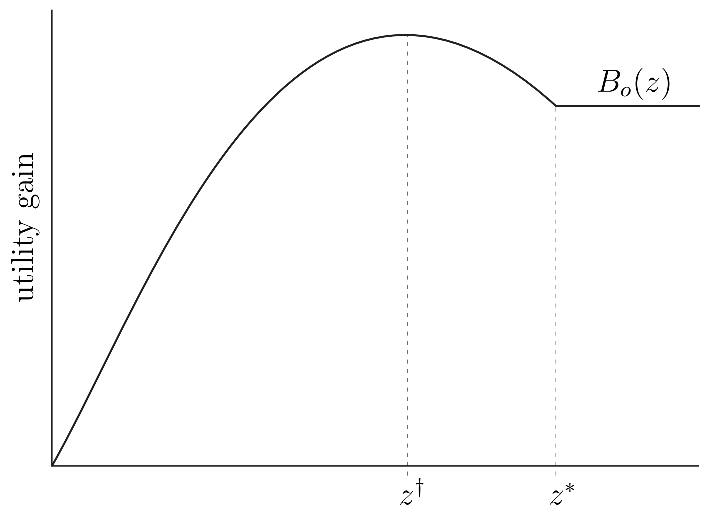


Fig. 2.1: Buyer's gain from trade under Nash bargaining without charge cards.

Notice that the solution depends only on the buyer's real money holdings, so $q_o(s, \breve{s})$ can be written $q_o(z)$ and $d_o(s, \breve{s})$ can be written $d_o(z)$.

Figure 2.1 shows the buyer's gain from trade $B_o(z)$ when she brings money z into the DM. Without assuming more structure on $u(q)$ and $c(q)$ we cannot say much about the shape of B_o. We know $B_o(0) = 0$ and $B_o(z^*) = u(q^*) - z^*$, but at any given point between these two we cannot pin down a

implying $u'(q_o) = c'(q_o)$ and $q_o = q^*$ and $d_o = z^* = \theta c(q^*) + (1-\theta)u(q^*)$. If the constraint does bind, then $d_o = z$. Equation (2.19) no longer holds with equality but equation (2.18) still does. Solving (2.18) for d_o gives the real money needed to pay for q_o, which defines $g(q)$.

first or second derivative. We *can* show that $B_o(z^*)$ is not a maximum. Total surplus is maximized at z^*, but buyers improve their bargaining position by carrying less money. This result is presented in Aruoba et al. (2007), but not explicitly proved for general $u(q)$ and $c(q)$. I state it here and provide the proof in the appendix.

Lemma 2. *Under Nash bargaining with money only, $z \geq z^*$ does not maximize the buyer's surplus. The maximum is at some smaller $z = z^\dagger$ at which the quantity exchanged is less than the social optimum, that is $q^\dagger \equiv \hat{q}(z^\dagger) < q^*$.*

Proof. See appendix. □

Nash bargaining with both charge cards and money

When both cards and money can be used the Nash bargaining solution is the triple (q_b, d_m, d_c) that solves

$$\max B_b^\theta S_b^{1-\theta} = \max_{q_b, d_m, d_c} \{u(q_b) - d_m - d_c\}^\theta \{-c(q_b) + d_m + (1-\tau)d_c\}^{1-\theta}$$
$$\text{subject to } d_m \leq z$$
$$\text{and } d_c \leq l \tag{2.21}$$

Using Lagrange multiplier μ for the first constraint and λ for the second, the Kuhn-Tucker conditions are

$$\theta u'(q)B^{\theta-1}S^{1-\theta} - (1-\theta)c'(q)B^{\theta}S^{-\theta} \leq 0 \qquad = 0 \text{ if } q > 0 \tag{2.22}$$

$$-\theta B^{\theta-1}S^{1-\theta} + (1-\theta)B^{\theta}S^{-\theta} - \mu \leq 0 \qquad = 0 \text{ if } d_m > 0 \tag{2.23}$$

$$-\theta B^{\theta-1}S^{1-\theta} + (1-\theta)(1-\tau)B^{\theta}S^{-\theta} - \lambda \leq 0 \qquad = 0 \text{ if } d_c > 0 \tag{2.24}$$

$$\mu(z - d_m) = 0 \qquad \lambda(l - d_c) = 0 \qquad q \geq 0 \qquad 0 \leq d_m \leq z \qquad 0 \leq d_c \leq l$$

The amount of the money payment can either be an interior solution, $0 < d_m < z$, or a corner solution with either $d_m = 0$ or $d_m = z$. It is also possible for the buyer to carry no money at all, in which case $d_m = z = 0$, which I consider as a fourth distinct alternative. Similarly, the card payment can be interior, $0 < d_c < l$, or a corner solution, $d_c = 0$ or $d_c = l$.[15] I consider each of these 4×3 potential solutions in turn (see appendix) to fully characterize the solution.

First we see what happens if the buyer brings no money into the DM.

Lemma 3. *In Nash bargaining, if the buyer has a card and is carrying no money, then*

[15] I disregard the possibility that the card company sets the spending limit at zero, since this is tantamount to having no cards at all.

1. *The quantity exchanged will be less than the first-best q^*.*

2. *The highest quantity is exchanged when the card's spending limit does not bind. This quantity is $\tilde{q}$ which solves $c'(\tilde{q}) = (1-\tau)u'(\tilde{q})$.*

3. *The card will always be used as long as $\tilde{q} > 0$.*

Thus the solution when the buyer carries no money is

$$
\begin{aligned}
q_b &= \begin{cases} \hat{Q}(0, l) & \text{if } l < l^* \\ \tilde{q} & \text{if } l \geq l^* \end{cases} \\
d_c &= \begin{cases} l & \text{if } l < l^* \\ l^* & \text{if } l \geq l^* \end{cases}
\end{aligned}
\tag{2.25}
$$

where $l^* = g_c(\tilde{q})$. $\hat{Q}(z, l)$ is the quantity of good that will be exchanged as part of the bargaining outcome when the buyer has z money and a charge card with spending limit l. For the special case where $z = 0$ this is the q that solves $l = g_c(q)$, where

$$g_c(q) \equiv \frac{\theta c(q)u'(q) + (1-\theta)u(q)c'(q)}{\theta(1-\tau)u'(q) + (1-\theta)c'(q)} \tag{2.26}$$

(See lemma 6 below for the general solution.)

Proof. See appendix. □

Next I show that if the buyer has some money then she will spend at least part of it.

Lemma 4. *In Nash bargaining, if the buyer is carrying money $z > 0$, then some payment will be made with money; that is $d_m > 0$.*

Proof. See appendix. □

The buyer will spend all the money she is carrying before using the card. If she does not spend all her money, it is because she buys the optimal quantity and pays the same amount of money as she would if she couldn't use a charge card, which is less than her money holdings.

Lemma 5. *In Nash bargaining, if the payment made with money is less than the amount of money the buyer is carrying ($d_m < z$) then*

1. *No payment is made with a charge card; that is $d_c = 0$.*

2. *The optimal quantity is exchanged; $q = q^*$.*

3. *Money payment is the same as without cards; $d_m = z^* = g(q^*)$.*

Proof. See appendix. □

Finally we see how much payment the buyer makes with her charge card after she has spent all her money.

Lemma 6. *In Nash bargaining, if the payment made with money equals the buyer's money holdings ($d_m = z$), then*

1. *If the card is used and the spending limit does not bind ($0 < d_c < l$), the quantity exchanged will be $q = \tilde{q}$.*

2. $d_c = 0$ *if* $q > \tilde{q}$. *The card is not used if the quantity the buyer gets when she spends all her money is already more than the optimal quantity that can be achieved with the card.*

The solution when both money and card are used is

$$q_b = \begin{cases} \hat{Q}(z,l) & \text{if } z < j^{-1}(l) \\ \tilde{q} & \text{if } j^{-1}(l) \leq z < \tilde{z} \\ \hat{q}(z) & \text{if } \tilde{z} \leq z < z^* \\ q^* & \text{if } z > z^* \end{cases}$$

$$d_m = \begin{cases} z & \text{if } z < z^* \\ z^* & \text{if } z \geq z^* \end{cases} \tag{2.27}$$

$$d_c = \begin{cases} l & \text{if } z < j^{-1}(l) \\ j(z) & \text{if } j^{-1}(l) \leq z < \tilde{z} \\ 0 & \text{if } z \geq \tilde{z} \end{cases}$$

Here $j(d_m) \equiv l^* - d_m(1-\tau+\theta\tau)/(1-\tau)$ *is the size of the card payment needed to purchase* $\tilde{q}$ *given that there is also a payment of* d_m *made with cash. This takes into account the fact that the merchant fee applies only to the part of the payment made by card. Hence* $j(0) = l^*$, *because this is the card payment needed to buy* $\tilde{q}$ *when no money is used.* $d_m\theta\tau/(1-\tau)$ *can be seen as the buyer's* *share of the savings from paying d_m of the total amount in money rather than with a card. Now we can see that $j^{-1}(l)$ is the value for z below which the card spending limit will bind. Define $\tilde{z} \equiv g(\tilde{q})$, which is the amount of cash needed to buy $\tilde{q}$ when no payment is made using a charge card. It should be clear that $j^{-1}(0) = \tilde{z}$ (when the card cannot be used at all, the card's second best is only achieved if $z \geq g(\tilde{q})$.)*

As described above, $\hat{Q}(z, l)$ is the quantity exchanged when the buyer brings z money and a card with spending limit l. The general definition (for $z \geq 0$) is this is the q that solves $z = G(q, l)$ where

$$G(q, d_c) \equiv \frac{\theta c(q)u'(q) + (1-\theta)u(q)c'(q) - d_c[\theta(1-\tau)u'(\tilde{q}) + (1-\theta)c'(\tilde{q})]}{\theta u'(q) + (1-\theta)c'(q)} \tag{2.28}$$

Proof. See appendix. □

Figure 2.2 shows the buyer's surplus $B_b(z, l)$ when she is carrying real money z and a charge card with spending limit l and she meets a seller who can accept cards. Note that for money holdings above $\tilde{z} = g(\tilde{q})$ the curve coincides with that of the money-only surplus $B_o(z)$ (the dashed line). This is because when the buyer has this much money, the card is not used: because of the merchant fee τ, it's not worth using the card at all on quantities above $\tilde{q}$. The linear section on the left part of the curve represents the region where $\tilde{q}$ is being traded, but the buyer's money is not enough to pay for it and the

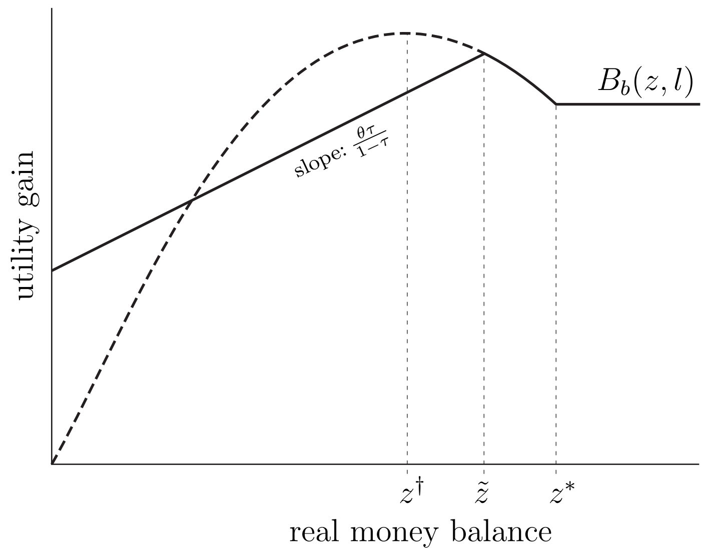


Fig. 2.2: Buyer's gain from trade under Nash bargaining with charge cards.

card is used to make up the difference. B_b is upward sloping in this region because for every additional unit of real money the buyer brings, she avoids paying $1 + \theta\tau/(1 - \theta\tau)$ with the card.

In figure 2.2, τ is such that $\tilde{z} > z^\dagger$ which means that $\tilde{q} > q^\dagger$. The wedge due to the card's merchant fee is not as bad as the inefficiency due to buyers strategically choosing to carry too little money. In this case, charge cards are welfare improving. A larger τ would move $\tilde{z}$ to the left, and also make the slope of the linear section steeper . Also in figure 2.2, the card spending limit never binds. Buyer's can carry $z = 0$ and still exchange $\tilde{q}$. If instead $l < l^*$, then the linear region would not extend all the way down to $z = 0$ (see figure 2.7).

Proportional bargaining with money only

If the money payment and quantity of special good traded is determined by proportional bargaining, the buyer's surplus is always a fixed proportion $\theta/(1-\theta)$ of the total surplus. $q_o(s,\breve{s})$ and $d_o(s,\breve{s})$ solve

$$\begin{aligned}
&\max_{q_o,d_o} u(q_o) - d_o \\
&\qquad \text{subject to } (1-\theta)[u(q_o) - d_o] = \theta[-c(q_o) + d_o] \\
&\qquad \text{and } d_o \leq z
\end{aligned} \tag{2.29}$$

The solution is

$$\begin{aligned}
q_o &= \begin{cases} \hat{q}(z) & \text{if } z < z^* \\ q^* & \text{if } z \geq z^* \end{cases} \\
d_o &= \begin{cases} z & \text{if } z < z^* \\ z^* & \text{if } z \geq z^* \end{cases}
\end{aligned} \tag{2.30}$$

where as before $z^* = g(q^*)$ and $\hat{q}(z)$ solves $z = g(\hat{q}(z))$ but now $g(q) = (1-\theta)u(q) + \theta c(q)$.

Figure 2.3 shows the buyer's gain under proportional bargaining when only money is used. Unlike the surplus from Nash bargaining, here $B_o(z)$ is well behaved.

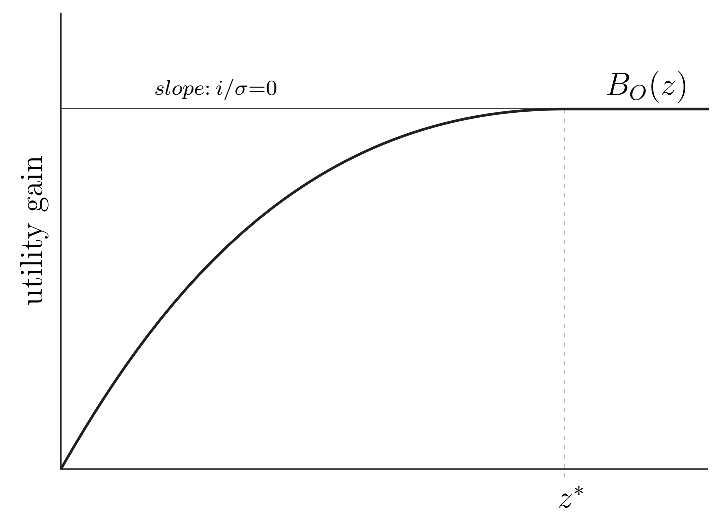


Fig. 2.3: Buyer's gain from trade under proportional bargaining without charge cards.

Lemma 7. *Under proportional bargaining when charge cards are not used,*

1. *The global maximum is attained at $z \geq z^*$.*

2. *There are no local maxima.*

3. *There is no at kink at z^*.*

Proof. $B_o'(z) = \hat{q}'(z)\hat{B}_o'(\hat{q}'(z))$ where $\hat{B}_o(q) = \theta[u(q) - c(q)]$ is the buyer's gain as a function of the quantity of goods she receives.

$$\hat{q}'(z) = 1/g'(q) = \frac{1}{(1-\theta)u'(\hat{q}(z)) + \theta c'(\hat{q}(z))} \tag{2.31}$$

$$B_o'(z) = \frac{\theta[u'(\hat{q}(z)) - c'(\hat{q}(z))]}{(1-\theta)u'(\hat{q}(z)) + \theta c'(\hat{q}(z))} \tag{2.32}$$

both of which are positive for all $z < z^*$. Moreover, evaluated at z^* where

$u'(q) = c'(q)$, $B_o'(z^*) = 0$, which matches the slope to the right of z^* where $B_o(z) = u(q^*) - z^*$. $\square$

Proportional bargaining with both charge cards and money

Proportional bargaining when both charge cards and money can be used solves

$$\begin{aligned}
&\max_{q_b, d_m, d_c} u(q_b) - d_m - d_c \\
&\qquad \text{subject to } (1-\theta)[u(q_b) - d_m - d_c] = \theta[-c(q_b) + d_m + (1-\tau)d_c] \\
&\qquad \text{and } d_o \leq z \\
&\qquad \text{and } d_c \leq l
\end{aligned} \tag{2.33}$$

The solution to this problem is

Lemma 8. *Under proportional bargaining when charge cards are used, the* *outcome is*

$$
q_b = \begin{cases} \hat{Q}(z,l) & \text{if } z < j^{-1}(l) \\ \tilde{q} & \text{if } j^{-1}(l) \leq z < \tilde{z} \\ \hat{q}(z) & \text{if } \tilde{z} \leq z < z^* \\ q^* & \text{if } z > z^* \end{cases}
$$

$$
d_m = \begin{cases} z & \text{if } z < z^* \\ z^* & \text{if } z \geq z^* \end{cases} \tag{2.34}
$$

$$
d_c = \begin{cases} l & \text{if } z < j^{-1}(l) \\ j(z) & \text{if } j^{-1}(l) \leq z < \tilde{z} \\ 0 & \text{if } z \geq \tilde{z} \end{cases}
$$

where all values are as defined in the Nash bargaining case, except that now

$$
G(q,l) = (1-\theta)u(q) + \theta c(q) - (1-\tau\theta)l \tag{2.35}
$$

Just as with Nash bargaining,

1. *If the buyer is carrying money she will spend all of it up to z^*, enough to buy q^*.*

2. *She will not make any payment by card unless her money is not enough*

to buy $\tilde{q}$, in which case she will spend all her money first and pay the remainder with the card.

3. *As long as the card's spending limit doesn't bind, the quantity exchanged will be $\tilde{q}$ if the card is used at all.*

Proof. See appendix. □

Figure 2.4 shows two versions of $B_b(z, l)$, the buyer's gain from proportional bargaining using card and money. The top solid line represents the case where the card's spending limit never binds ($l \geq l^*$) – the linear part of the curve curve extends all the way down to $z = 0$. The lower solid line shows what happens when $l < l^*$. When she brings $z < j^{-1}(l)$, the buyer spends all her money and maxes-out her charge card too, but together these are not enough to pay for $\tilde{q}$ so a smaller quantity is exchanged. Both curves coincide with the money only curve $B_0(z)$ (the dashed line) above $\tilde{z}$.

As with the money-only curve, there is no kink in $B_b(z)$ at z^*. Moreover, there are no kinks at the other breakpoints either. The right-sided derivative of $B_b(\tilde{z})$ and the left-sided derivative at $B_b(j^{-1}(l))$ are both the same as the slope of the linear section in between them: $\theta\tau/(1 - \theta\tau)$.

2.3.4 Distribution of money holdings

Now that we have established the results of bargaining in the DM, both when charge cards are used and when only money is used, let us return to the

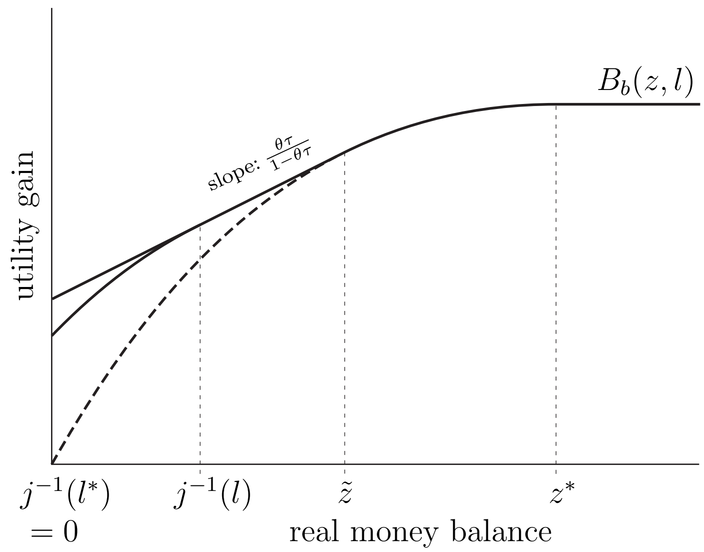


Fig. 2.4: Buyer's gain from trade under proportional bargaining with charge cards.

agent's decision in the CM. As we have seen, the outcome of q_o and d_o or q_b, d_m and d_c depend only on the buyer's money holdings and the charge card spending limit if a card can be used. Thus the gains from trade can be written as

$$B_o(z) = u\big(q_o(z)\big) - d_o(z) \tag{2.36}$$

$$S_o(\breve{z}) = -c\big(q_o(\breve{z})\big) + d_o(\breve{z}) \tag{2.37}$$

$$B_b(z,l) = u\big(q_b(z,l)\big) - d_m(z,l) - d_c(z,l) \tag{2.38}$$

$$S_b(\breve{z},l) = -c\big(q_b(\breve{z},l)\big) + d_m(\breve{z},l) + (1-\tau)d_c(\breve{z},l) \tag{2.39}$$

Only B_o and B_b depend on the agent's own choice of real money holdings. Using this and (2.11) the value of entering the DM for each combination of the cardholding and card-accepting decision can be simplified as

$$V^{\text{neither}}(z) = z + W_0 + \beta W_1 + \sigma B_o(z) + \sigma A_0(\Pi_b) \tag{2.40}$$

$$V^{\text{hold}}(z) = z + W_0 + \beta W_1 + \sigma \Pi_s B_b(z) + \sigma(1 - \Pi_s) B_o(z) + \sigma A_0(\Pi_b) \tag{2.41}$$

$$V^{\text{accept}}(z) = z + W_0 + \beta W_1 + \sigma B_o(z) + \sigma A_1(\Pi_b) \tag{2.42}$$

$$V^{\text{both}}(z) = z + W_0 + \beta W_1 + \sigma \Pi_s B_b(z) + \sigma(1 - \Pi_s) B_o(z) + \sigma A_1(\Pi_b) \tag{2.43}$$

where $A_0(\Pi_b)$ is the agent's expected surplus from being a seller in the DM if she decides not to accept cards, and $A_1(\Pi_b)$ is the her expected surplus if decides to accept cards and the probability of meeting a cardholder is Π_b.

$$A_0(\Pi_b) \equiv \int S_o(\breve{z}) dF(\breve{z}) \tag{2.44}$$

$$A_1(\Pi_b) \equiv \Pi_b \int S_b(\breve{z}, l)\, dF(\breve{z}|\breve{\pi}_b = 1) + (1 - \Pi_b) \int S_o(\breve{z})\, dF(\breve{z}|\breve{\pi}_b = 0) \tag{2.45}$$

Note that neither $A_0(\Pi_b)$ nor $A_1(\Pi_b)$ depend on the agent's own choice of money holdings.

Plugging these expressions into (2.11) and dropping terms that do not depend on $\hat{z}$ we see that an agent's choice of real balances to take into the next period takes one of two forms. If she chooses to carry only money, her choice of $\hat{z}$ solves

$$\max_{\hat{z}} \Big\{ -i\hat{z} + \sigma B_o(\hat{z}) \Big\} \tag{2.46}$$

If she decides to become a cardholder, her choice of $\hat{z}$ solves $\hat{z}$ that solves

$$\max_{\hat{z}} \Big\{ -i\hat{z} + \sigma \Pi_s B_b(\hat{z}, l) + \sigma(1 - \Pi_s) B_o(\hat{z}) \Big\} \tag{2.47}$$

The solutions to these problems can be seen visually from graphs like figure 2.1. The agent chooses the money holding that maximizes the distance between the curve B and a line through the origin with slope i/σ. Alternatively, find the most northwestern point on B that is tangent to a line with slope i/σ.

These maximization problems have no solution if $i < 0$, which would imply that inflation is below the Friedman rule, so I will restrict attention to situations where $i \geq 0$. We know that at nominal interest rate $i = 0$ the buyer will choose $z \geq z^*$ under proportional bargaining, but she will choose $z^\dagger < z^*$ under Nash bargaining.

With either bargaining solution we cannot ensure that $B_o(z)$ or $B_b(z)$ are concave for all z. They would be if the buyer has all of the bargaining power ($\theta = 1$), or if we impose more structure on $c(q)$ and $u(q)$. In an unpublished version of their 2005 paper, Lagos and Wright show that it is sufficient for $\log u(q)$ to be concave. Without concavity of the buyer's surplus, there is a danger that the solution for the money balance will not be unique. At a given nominal interest rate i, a line with slope i might happen to be tangent to $B(z)$ at more than point. Moreover, with charge cards $B_b(q)$ has a linear section. If the interest rate is exactly $i = \theta\tau/(1-\tau)$ (under Nash bargaining or $i = \theta\tau/(1-\theta\tau)$ under proportional bargaining) then any $j^{-1}(l) < z < \tilde{z}$ will be a solution.

Assuming that (2.46) and (2.47) have unique solutions, then for a given interest rate all agents making the same cardholding choice will carry the same real balances. The choice of money balance also depends on the probability Π_s of meeting someone who accepts cards. Note that the two maximization problems are the same if $\Pi_s = 0$. Agents choosing to carry a card will also carry $\hat{z}(i, \Pi_s)$ money. Agents choosing to carry *only* money will carry $\hat{z}(i, 0)$. The distributions of money holdings conditional on cardholding decision are thus degenerate: $F(z|\pi_b = 0) = \hat{z}(i, 0)$ and $F(z|\pi_b = 1) = \hat{z}(i, \Pi_s)$. This simplifies A_0 and A_1.

$$A_0(\Pi_b) = \Pi_b S_o\big(\hat{z}(i, \Pi_s)\big) + (1 - \Pi_b) S_o\big(\hat{z}(i, 0)\big) \tag{2.48}$$

$$A_1(\Pi_b) = \Pi_b S_b\big(\hat{z}(i, \Pi_s), l\big) + (1 - \Pi_b) S_o\big(\hat{z}(i, 0)\big) \tag{2.49}$$

2.3.5 Card-accepting and cardholding decisions

The choice of whether or not to accept cards is made by comparing the expected benefits given the interest rate and the cardholding choices of other agents in the economy. Each agent chooses to pay fixed costs η^s to be able to accept cards next period when

$$\sigma A_1(\Pi_b) - \eta^s \geq \sigma A_0(\Pi_b) \tag{2.50}$$

Similarly, the choice of whether or not to become a cardholder is made by comparing the expected benefits given the interest rate and the card-accepting choices of other agents in the economy. Each agent chooses to pay the annual fee η^b to become a cardholder in the next period when

$$\Pi_s B_b\big(\hat{z}(i,\Pi_s)\big) + (1-\Pi_s)B_o\big(\hat{z}(i,\Pi_s)\big) - \frac{i}{\sigma}\hat{z}(i,\Pi_s) - \frac{\eta^b}{\beta\sigma} \geq B_o\big(\hat{z}(i,0)\big) - \frac{i}{\sigma}\hat{z}(i,0) \tag{2.51}$$

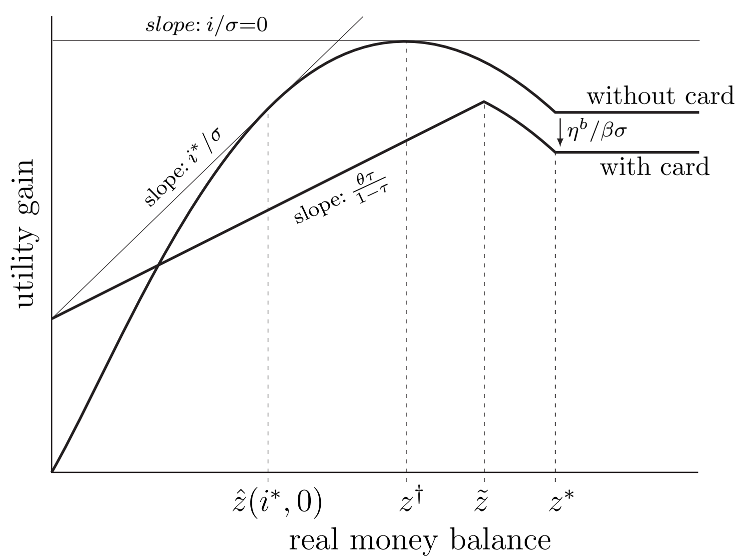


Fig. 2.5: Cardholding decision under Nash bargaining.

Figure 2.5 depicts this choice in a pure charge card equilibrium, meaning all agents accept cards ($\Pi_s = 1$). For a given interest rate i, agents will choose the point on the graph with highest tangent line of slope i/σ. The curve for B_b (the gain from using a card) is shifted down by $\eta^b/\beta\sigma$. In a mixed equilibrium where $\Pi_s < 1$ the curve for card acceptance would be a convex combination of B_b and B_o shifted down by $\eta^b/\beta\sigma$.

At the interest rate drawn in figure 2.5 the agent is just indifferent about becoming a cardholder. Call this rate i^*.

$$i^* \text{ solves } \frac{i^*}{\sigma}[\hat{z}(i^*,0) - \hat{z}(i^*,1)] = B_o\big(\hat{z}(i^*,0)\big) - B_b\big(\hat{z}(i^*,1)\big) + \frac{\eta^b}{\beta\sigma} \qquad (2.52)$$

For $i > i^*$ the buyer would prefer to be a cardholder. In the figure, the card spending limit doesn't bind, so if the agent chooses to hold a card, she will choose to carry zero money and $\tilde{q}$ goods will be exchanged. If she decides not to be a cardholder she'll carry $\hat{z}(i,0)$ money and $\hat{q}(\hat{z}(i,0))$ goods will be exchanged.

At a given interest rate, charge cards are welfare improving if $\hat{Q}(\hat{z}(i,1),l) > \hat{q}(\hat{z}(i,0))$, which translates into $\tilde{q} > \hat{q}(\hat{z}(i,0))$ if the spending limit doesn't bind. By setting $i \geq i^*$ a monetary authority can make charge cards viable. To find whether this is optimal, compare $\hat{q}(\hat{z}(i,1))$ to $\hat{q}(\hat{z}(0,0)) = q^\dagger$. Under proportional bargaining $q^\dagger = q^*$ and it is socially optimal to set $i = 0$ (the Friedman rule). Figure 2.6 depicts this choice for proportional bargaining. In the figure $l < l^*$ so at i^* the spending limit is binding and the agent carries

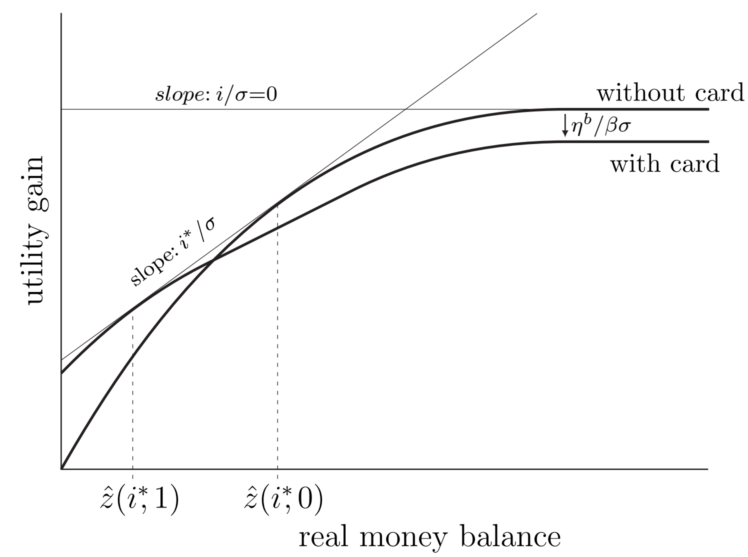


Fig. 2.6: Cardholding decision under proportional bargaining.

some money $\hat{z}(i,1) > 0$ even when she is a cardholder. But the money-only solution at $i = 0$ attains the first best, while the card solution only attains $\hat{Q}(\hat{z}(i,1),l) < \tilde{q} < q^*$.

In contrast, under Nash bargaining $q^\dagger < q^*$. If $q^\dagger$ is also less than $\tilde{q}$ as shown in figure 2.5,[16] then a charge card equilibrium welfare-dominates a monetary equilibrium even at the Friedman rule and $i = i^*$ is the optimal rate – provided the economy can find the charge card equilibrium! The same could be true under any other mechanism for determining the terms of exchange that gives buyers a strategic incentive to limit their money holding.

For a given i, comparing B_b and B_o shows which cardholding choice a buyer prefers, and comparing $\hat{Q}$ and $\hat{q}$ shows us which yields more total

[16] Figure 2.5 shows $z^\dagger < \tilde{z}$. For the sake of argument assume that also $q^\dagger < \tilde{q}$, although this cannot be read from the graph. We have to be careful to compare qs rather than zs since $\hat{q}'(z)$ and $\hat{Q}z(z,l)$ are not guaranteed to be positive.

surplus. However, there is another channel by which interest rates can affect welfare and the cardholding decision. Up to now we have treated the charge card spending limit as exogenous. It is likely, however, that the spending limit will be a function of interest rates. In the next section we endogenize the spending limit.

2.3.6 Spending limits

The size of the spending limit imposed by the card company on buyers will depend on the legal institutions enforcing payment. The higher the cost of defaulting, the higher the card company can set the limit and still expect to be paid. One can imagine a severe punishment (e.g. debtor's prison) which would make the spending limit arbitrarily high. Within the confines of the model, a fairly severe punishment might be exclusion from all future CM sub-periods. This could represent something between a bad credit rating and a criminal record which prevents the agent from engaging in any exchange where reputation matters. The agent would reduced to only anonymous interactions in the DM.

Here I establish a lower bound on the spending limit corresponding to the minimum punishment that should always be available to the card company: revoking the charge card. Cardholders weigh the immediate benefit of not paying their claim to the card company against the present discounted cost of only being able to use money in all future periods. The spending limit will be the threshold amount that leaves the cardholder indifferent between paying and not paying the claim. Under this minimal enforcement mechanism, the limit is $l = \hat{l}(i)$ which solves

$$\hat{l} = \frac{\text{net gain with card} - \text{net gain w/o card}}{1-\beta} \tag{2.53}$$

$$= \frac{\sigma\left[B_b\left(\hat{z}(i,1), \hat{l}\right) - B_o\left(\hat{z}(i,0)\right)\right] + i[\hat{z}(i,0) - \hat{z}(i,1)] - \eta^b/\beta}{1-\beta} \tag{2.54}$$

This implies that i^* does not exist! Recall that i^* is the interest rate that makes agents indifferent between holding a card and not. If agents are just indifferent, then by equation (2.53) $l = 0$. But this is equivalent to having no card at all. The card-and-money curve B_b and the money-only curve B_o are identical. If there is any positive annual fee $\eta^b > 0$ then choosing not to be a cardholder will be strictly preferred which contradicts that agents are indifferent.

This is not to say that the charge card is never preferred, just that agents are never indifferent. For extremely high interest rates $\lim_{i\to\infty} B_o = 0$. Then a positive spending limit that allows $B_c > 0$ can be supported if agents are patient enough, since by (2.53), l can be arbitrarily large if β is close enough to one. This leads to the definition of another threshold interest rate i^{**} which is the minimum i such that (2.51) is satisfied when l is defined by (2.53). At i^{**}

using a card is significantly preferred to using only money. If the interest rate falls below i^{**} the cost of losing the card falls slightly but then the spending limit collapses. Figure 2.7 shows a cardholding choice where there is a binding spending limit that is supported by a positive difference ΔB in buyer's gains at i^{**}. When there are no institutions to promote payment of claims i^{**} is the minimum interest rate at which charge cards are viable. If $\hat{Q}\big(\hat{z}(i, \Pi_s), l\big) > q^{\dagger}$ then the optimal interest rate is $i \geq i^{**}$. As depicted in figure 2.7 the optimal rate is $i = 0$ because $\hat{Q}\big(\hat{z}(i, \Pi_s), l\big) < q^{\dagger}$.

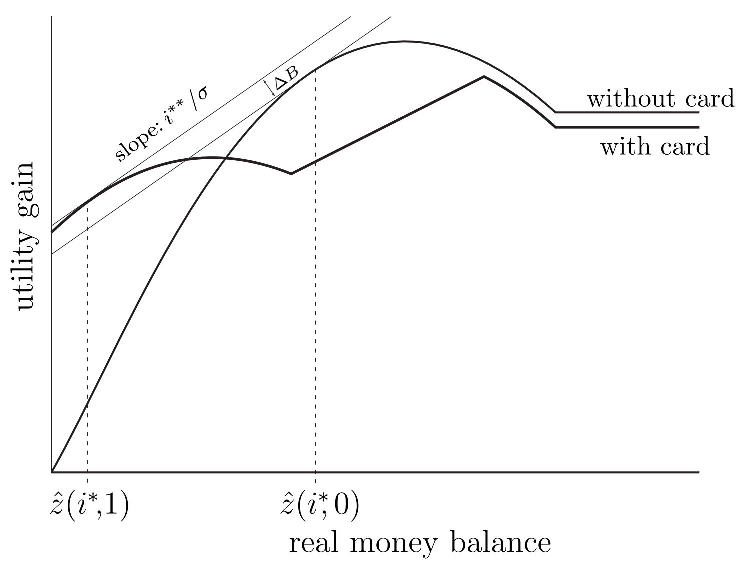


Fig. 2.7: Card spending limit derived from difference in buyer's gains.

Notice that it is possible for higher interest rates to be welfare improving even if charge cards are already being used. If the spending limit is binding, agents' choice of money holdings is to the left of $j^{-1}(l)$ on a curved section section of $B_b(z)$ where $\hat{q}(z, l) < \tilde{q}$. If l were fixed, this would mean increasing

i would decrease z and therefore decrease q. But the effect is ambiguous if the spending limit is set by (2.53). If $B_b(z)$ at this point is less curved than $B_o(\hat{z}(i,0))$ at the point where *it* would be tangent to i/σ – that is if

$$\hat{Q}_{zz}\big(\hat{z}(i,1),l\big) > \hat{q}''\big(\hat{z}(i,0)\big) \tag{2.55}$$

then increasing i also increases $\big[B_b\big(\hat{z}(i,1),l\big) - B_o\big(\hat{z}(i,0)\big)\big]$, which raises the spending limit and pushes $j^{-1}(l)$ to the left. Depending on parameters (especially β) this effect could dominate and actually raise $\hat{Q}\big(\hat{z}(i,1),l\big)$, increasing total surplus.

2.4 Equilibrium

There are four possible equilibria: monetary, charge card, mixed, and no-trade. In the monetary equilibrium cards are not used and all agents carry money. Depending on the interest rate, all agents might be better off if they could collectively switch to using cards. The no-trade equilibrium is the extreme case of this coordination failure, where agents carry neither money nor cards. In the charge-card equilibrium all agents hold charge cards, and all agents accept cards in exchange. If the spending limit does not bind, then payment will be made entirely by charge card and money will not be used. In a mixed equilibrium, charge cards are used for some transactions.

2.4.1 Monetary equilibrium

The monetary equilibrium here is identical to that in Lagos and Wright (2005). Although charge cards are available, no agents choose to accept them as payment and no agents carry a card. This is the only possible equilibrium if $i < i^*$ (or i^{**} in case the spending limit must be supported by the threat of revoking the card). Buyers will not deviate because they prefer using money to cards. Sellers will not deviate because choosing to accept cards costs η^s but provides no benefit since cards will never be used. If $\eta^s = 0$ then a variation of this equilibrium is possible in which agents do accept cards, but there are still no cardholders.

If $i \geq i^*$ then buyers would prefer to pay the annual fee and use a card instead of money, but cards are not accepted. It may be that the merchant setup fee is so high that sellers do not accept cards. If $\eta^s > \bar{\eta}^s(i)$ where

$$\bar{\eta}^s(i) \equiv \sigma[A_1(1) - A_0] = \sigma\big[S_b\big(\hat{z}(i,0), l\big) - S_o\big(\hat{z}(i,0)\big)\big] \tag{2.56}$$

then agents would not pay to accept cards even if all other agents are cardholders. This threshold depends on i because the amount of money that other agents in the economy are carrying depends on i. Presumably, the charge card company would never set a fee so high as to have no customers.

If $i \geq i^*$ and $\eta^s \leq \bar{\eta}^s(i)$ then agents may be stuck in a bad equilibrium.

Collective action would be required to switch to the charge card equilibrium. As i increases and $\hat{z}(i, 0)$ falls, the benefits to both parties from switching to cards becomes greater. Higher nominal interest rates make the monetary equilibrium less stable, in the sense that fewer deviations are required to switch into either a mixed or charge-card equilibrium.

2.4.2 No-trade equilibrium

If $u'(0) = \bar{u}$ is finite, then there will be an interest rate $\bar{i}$ above which agents without charge cards will carry no money. It's possible that agents will still not be able to coordinate on using cards. With no medium of exchange available, trade cannot occur in the DM at all (since we have assumed away double-coincidence meetings, barter is not an option). This is the lowest welfare equilibrium.

A higher interest rate no longer has any effect on the number of deviations required to switch to an equilibrium where cards are used. Since the gains from trade are bounded by $u(\tilde{q})$, this equilibrium is only unstable (in the sense that a small number of simultaneous deviations can switch the economy out of it) if either η^s or η^b are very small. Of course, the charge card company has an incentive to set its fees in just such a way in order to encourage usage.

2.4.3 Charge card equilibrium

In the cards-only equilibrium, all agents carry and accept charge cards. Cards are used in exchange. If the card's spending limit is not binding, then money is not used in exchange. If the spending limit does bind, then agents also carry some money and pay with a combination of card and money. A buyer who deviates and does not carry a card (perhaps because she is being punished for not paying a claim) can still make exchanges using money, but she will purchase a smaller quantity than she would with a card. This and her holding cost of of money reduces her overall utility. A seller who deviates and does not accept cards will either not be able to trade at all if buyers are carrying no money, or only be able to trade a smaller amount if the card's spending limit binds and agents are carrying some money.

Nominal interest rates must be $i \geq i^*$ for a charge card equilibrium to be possible, and $i \geq i^{**}$ if there are no institutions to enforce the payment of claims. The merchant set-up fee must also be $\eta^s < \bar{\eta^s}(i)$. If $\hat{Q}\big(\hat{z}(i,1), l\big) > q^\dagger$, then this is the highest welfare equilibrium.

2.4.4 Mixed equilibrium

All agents are identical in this economy, so for cards to be used only some of the time it must be that agents are indifferent and randomizing on one or both of the ex-ante decisions about charge cards. If agents are indifferent between accepting and not accepting cards, they randomly choose to accept Π^s of the time. Then in any period Π^s is the fraction of sellers who accept cards. Similarly if agents are indifferent between carrying and not carrying a card, then Π^b is the fraction of buyers who are cardholders.

Buyers who do not carry cards do carry money, unless $i > \bar{i}$. The amount of money they hold is the same amount buyers hold in the monetary equilibrium $\hat{z}(i, 0)$. If $\Pi^s = 1$, then buyers who *do* carry cards hold the same amount of money as buyers in the charge card equilibrium. If $\pi^s < 1$ then they carry $\hat{z}(i, \Pi_s)$ money. Sellers' payoffs when money is used will depend on the amount of money agents carry, which in turn depends on the fraction of other sellers that choose to accept cards. However, agents do not take into account the effect of their own card-acceptance decision on other's money holdings.

Any pair of Π^s and Π^b such that both conditions (2.50) and (2.51) hold with equality, is a possible equilibrium in which cards are used. However, such equilibria are unstable. A slight increase or decrease in Π_b makes sellers no longer indifferent, and a slight increase or decrease in Π_s makes buyers no longer indifferent.

2.5 Simulation

Following the numerical analysis in Aruoba et al. (2007), I assume the following functional forms: $U(X) = \Omega \ln X$, $u(q) = q^{1-\zeta}/(1-\zeta)$, and $c(q) = q$.

Tab. 2.1: Parameters for Simulated Model

Parameter	Symbol	Value
Discount factor	β	$(1.03)^{-1}$
CM consumption	Ω	1.65
DM utility parameter	ζ	0.36
Buyer's bargaining power	θ	0.32
Merchant discount	τ	0.015
Meeting probability	σ	0.5
Annual fee	η^b	0.0059
Merchant's set-up fee	η^s	0
Spending limit	σ	0.59

I assume the Nash bargaining solution in the DM. The period is a year. See table 2.1 for the parameter values used. I assume a β that corresponds to a 3% real interest rate. I use the values for Ω, ζ, θ and σ that Aruoba et al. calibrate from U.S. money demand 1900-2000. I have no evidence that merchant set-up fees are important in reality so I assume $\eta^s = 0$.

I assume an annual fee of \$100, and charge card spending limit of \$10,000. In order to translate this into the model I need to convert dollar values into units of general good consumption. With the assumed specification, agents will consume $X^* = \Omega$ general good in the CM. Consumption of special goods in the DM depends on the interest rate and card usage. The first-best would be $q^* = 1$, but this level is never achieved under Nash bargaining. The second-best consumption that can be attained when cards are used is $\tilde{q}$. The price of q^* in terms of general goods depends on how much payment is made with card and how much is made with money. As a quick estimate I use $\tilde{z}$. The

U.S. per capita GDP is a little less than \$50,000.[17] So my rough translation for choosing parameters is $\Omega + \tilde{z} \approx \$50,000$.

Tab. 2.2: Simulation Results

Description	Symbol	Value
Optimal DM consumption	q^*	1
Second best consumption with card	$\tilde{q}$	0.9589
Consumption at Friedman rule w/o card	$q^\dagger$	0.6024
Consumption at i^* with card	q^{card}	0.6597
Real balance needed to buy q^*	z^*	1.3825
Real balance needed to buy $\tilde{q}$ w/o card	$\tilde{z}$	1.3393
Money holdings under Friedman rule w/o card	$z^\dagger$	0.9393
Spending limit that would never bind	l^*	1.3458
Min interest rate for card equilibrium	i^*	0.0078
Min rate for endogenous spending limit	i^{**}	0.0440
Rate the achieves endogenous limit of l^*	i^{***}	0.0921

Simulation results are listed in table 2.2. Given this exogenous spending limit, the charge card equilibrium is viable when the nominal interest rate is above $i^* = 0.78\%$.

2.5.1 *Endogenous spending limit*

If the threat of losing the charge card is the only way the card company can induce cardholders to pay claims, then there must be a strictly positive benefit to having a charge card. At a given interest rate, there is a range of spending limits that can be supported. Figure 2.9 shows the region of feasible interest rate / spending limit combinations in the simulated model.

[17] In 2008 per-capita GDP in the United States was \$47,025 according to the International Monetary Fund's World Economic Outlook Database-October 2008. The World Bank's World Development Indicators database puts the figure for 2007 at \$45,790.

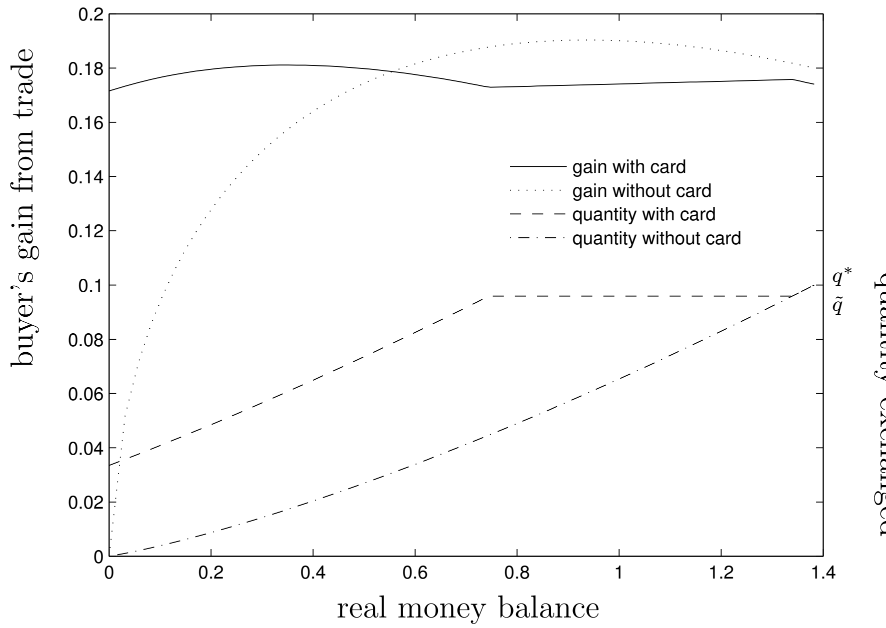


Fig. 2.8: Buyer's gain from trade and quantity exchanged.

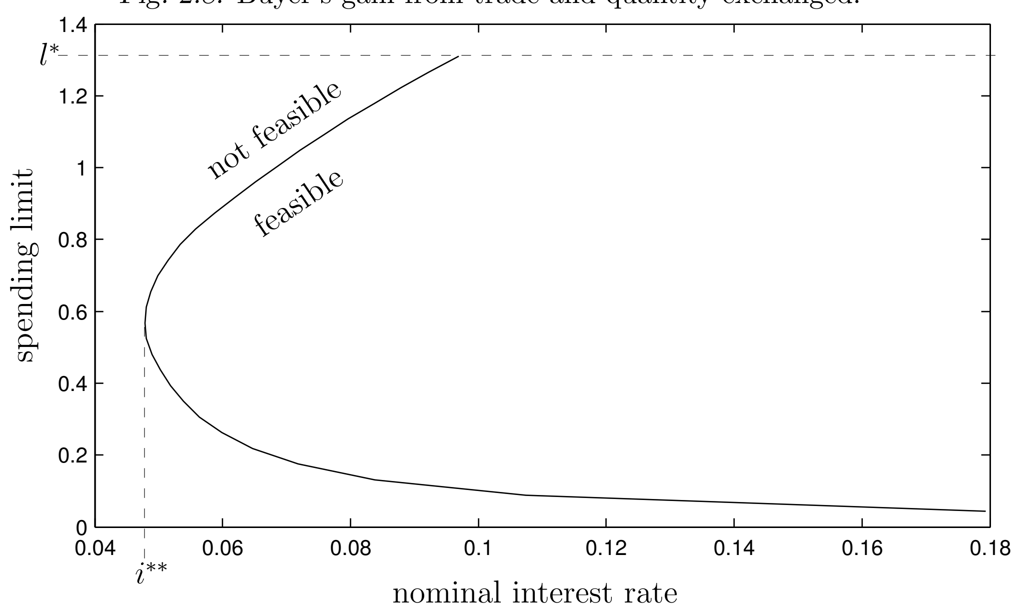


Fig. 2.9: Endogenous spending limits.

Below i^{**} no positive spending limit can be supported, and charge cards are not viable without some other mechanism to induce payment of claims. With the parameters in table 2.1 $i^{**} = 0.0440$. I can also find the interest rate $i^{***} = 0.0921$ that would support l^* the spending limit that allows the second best quantity $\tilde{q}$ to be exchanged even when buyers carry no money.

2.5.2 Welfare benefit of charge cards

Assuming exogenous spending limit $l = 0.59$, at $i^* = 0.0078$ buyers prefer to become cardholders and the quantity traded becomes $q^{card} \equiv \hat{Q}(\hat{z}(i^*, 1), l) = 0.6597$. This quantity is still far from the first best, but it is greater than the quantity that is exchanged under the Friedman rule, since under the Friedman rule buyers prefer not to be cardholders, so only $q^{\dagger} = 0.6024$ is exchanged. This makes i^* the optimal nominal interest rate, or $\gamma^* \equiv \beta(1 + i^*) = 0.9784$ the optimal money growth rate, which is higher than the Friedman rule $\gamma^{FR} = \beta = 0.9709$.

One measure of the welfare benefit of charge cards is the equivalent variation in terms of consumption. This is the fraction $(1 - \Delta)$ of their total consumption that agents would be willing to give up instead of giving up the ability to use charge cards. Δ solves

$$U(\Delta\Omega) - \Omega + \sigma[u(\Delta q^{card}) - q^{card}] = U(\Omega) - \Omega + \sigma[u(q^{\dagger}) - q^{\dagger}] \qquad (2.57)$$

For agents in this simulation comparing optimal monetary policy with and without cards, charge cards are worth just 0.25% of consumption.

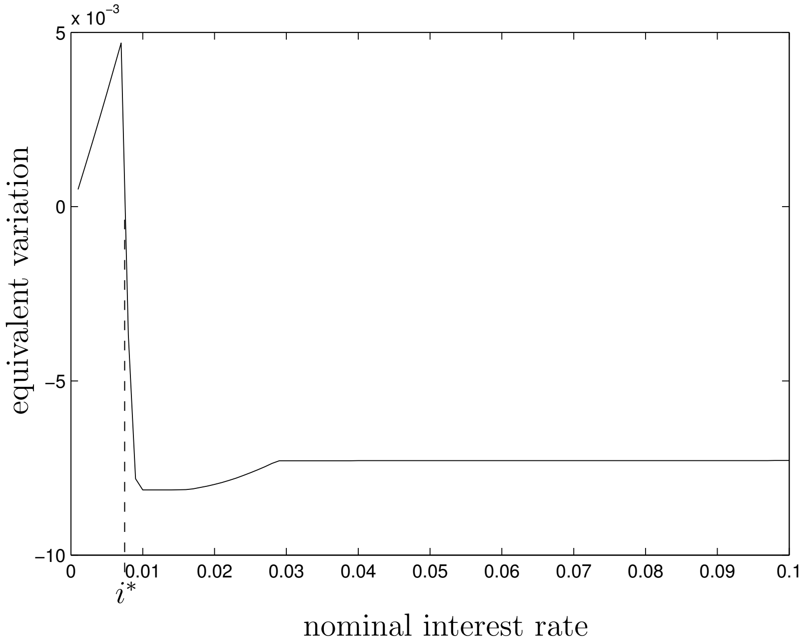


Fig. 2.10: Welfare cost/gain from inflation (exogenous spending limit).

If instead the card spending limit is endogenously determined, then the optimal interest rate is $i^{***} = 0.0921$ which allows $\tilde{q} = 0.9589$ to be exchanged. This corresponds to money growth of $\gamma^{***} = 1.0603$ – too high for most central bankers' tastes, but in the realm of the plausible. In this situation, agents would give up 0.81% of consumption to keep their cards. That this equivalent variation is still so small, even though the quantity of special good is falling by 37% when cards are lost, tells us that the buyer's marginal utility is already close to the seller's marginal cost over this range of quantities.

2.5.3 Cost of inflation

Now consider the welfare cost of inflation in the economy with charge cards. Figure 2.10 shows the equivalent variation for a fixed, exogenous spending limit. This is the fraction of total consumption that agents living under inflation that implies nominal interest rate i would exchange for switching to the Friedman rule ($i = 0$). Figure 2.11 shows the same calculation when the spending limit is a function of inflation. Both graphs start out with welfare costs increasing in interest rates, but then these costs switch to gains as the rate moves above a threshold that permits a charge card equilibrium. At this point agents are using both money and cards. With a fixed spending limit, further increasing i still reduces the money agents carry and reduces welfare until agents stop holding money altogether. After this, only the card is used, and consumption is no longer sensitive to inflation.

When the spending limit is endogenous, it takes a higher interest rate to make card usage viable. Increasing the rate further allows the card company to increase the spending limit and allows a higher quantity to be exchanged until l^* is reached and the quantity becomes the optimal (second-best) $\tilde{q}$.

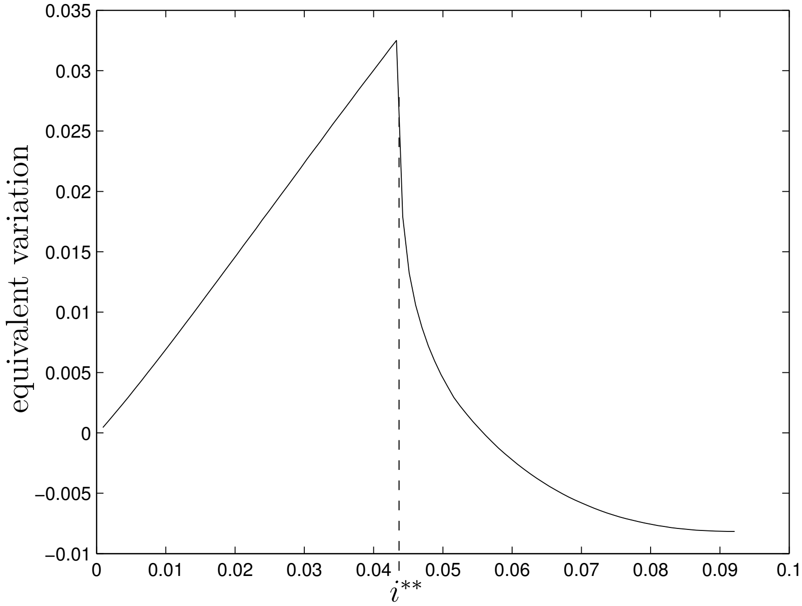


Fig. 2.11: Welfare cost/gain from inflation (endogenous spending limit).

2.6 Extensions

2.6.1 Endogenize card company fees

The most obvious missing piece in this model is a description of how the card company's fees are determined. Fees ought to maximize card company profit. Clearly the company will not set fees that lead to a monetary equilibrium. The structure of the charge card industry matters. Is the card company a monopolist? If there is competition, will companies stake out different positions in the two-sided market for charge cards? Endogenizing card companies' decisions may have serious ramifications for monetary policy.

2.6.2 Heterogeneous sellers

In the real world charge cards and money have coexisted for over 50 years; though great fluctuations in macroeconomic variables and shifting levels of usage for numerous different media of exchange. There remain merchants who do not accept payment by card, but this cannot be viewed as indifferent agents randomizing over a mixed strategy, because with the slightest shift in parameters or even a tiny fluctuation in the fraction of other agents participating, agents are no longer indifferent. In short, the mixed equilibrium is too fragile.

A large dose of realism might be attained while keeping the model tractable by keeping money holdings degenerate but allowing some variation between sellers. In particular, if the cost of production had a fixed component that was heterogeneous across agents, then the merchant discount would matter more in some DM meetings than others. Sellers with high margins would embrace the card and those with low margins would stick to money. Different interest rates and card company fees would move the threshold for seller participation. With a distribution of fixed costs, a mixed equilibrium would become far more tenable.

2.6.3 *"No surcharge rules" and bargaining over price*

A major modeling decision is how buyers and sellers negotiate the method of payment to use. In this paper I stay as close as possible to the original Lagos-Wright model and look for a Nash bargaining solution that specifies quantity and payment amount with each medium of exchange. In reality however, card company "no-surcharge" rules, and sometimes legal restrictions, prohibit sellers from charging more to people who are paying with plastic. It is not clear how to impose this restriction when the bargaining takes place over quantity and total payment. An alternative would be to allow agents to instead bargain over price. The buyer knows the quantity and payment method she will choose for any given price. The seller knows neither of these, and so bargains over expected payout. The Nash solution would be

$$\max_{p} \Big\{ \max\big[b\big(q_m(p),\, p\big)\,,\, b\big(q_c(p),\, p\big)\big] \Big\}^{\theta} \Big\{ \mathrm{E}(s|p) \Big\}^{1-\theta} \tag{2.58}$$

$$b(q,p) = u(q) - \phi q p$$

$$q_m(p) = \begin{cases} q(p) & \text{if } z \geq q(p)\,p \\[2ex] z/p & \text{if } z < q(p)\,p \end{cases}$$

$$q_c(p) = \begin{cases} q(p) & \text{if } l \geq q(p)\,p \\[2ex] l/p & \text{if } l < q(p)\,p \end{cases}$$

$$s = \begin{cases} -c\big(q_m(p)\big) + q_m(p)p & \text{if buyer pays with money} \\[2ex] -c\big(q_c(p)\big) + (1-\tau)q_c(p)p & \text{if buyer pays with card} \end{cases} \tag{2.59}$$

2.7 Conclusions for monetary policy

Optimal monetary policy differs depending on the institution of payment enforcement. If there are strong institutions, then the charge card spending limit might be independent of the interest rate. Then the rate i^* that makes agents indifferent about becoming cardholders is the threshold rate that makes a charge card equilibrium possible. A rate above i^* will decrease quantity and therefore welfare relative to i^* as long as money continues to be used along side cards. If cards completely replace money then welfare becomes independent of inflation.

If the charge card spending limit is derived from the card company's threat of revoking the card, then to make a charge card equilibrium viable the interest rate must be above i^{**}. A monetary authority may be able to further increase l by raising i, but only up to the point $\bar{i}$ where punished agents would cease to carry money altogether. If l^* can be achieved then consumption will be the second-best level $\tilde{q}$. Let $\bar{l}$ be the highest achievable spending limit, and let $\tilde{l} = \min(l^*, \bar{l})$. Let $\tilde{i}$ be the interest rate that induces $\tilde{l}$.

Policy makers should compare welfare under the best possible charge card equilibrium, $u(\hat{Q}(i^*, l)) - c(\hat{Q}(i^*, l))$ or $u(\hat{Q}(\tilde{i}, \tilde{l})) - c(\hat{Q}(\tilde{i}, \tilde{l}))$, with that under the best possible monetary equilibrium $u(q^{\dagger}) - c(q^{\dagger})$.

If the monetary equilibrium is better, then the optimal policy is the Friedman rule. If the economy has been in a equilibrium using cards, it will switch to the monetary equilibrium when the rate of money growth is dropped. If bargaining is proportional (or some other mechanism where buyers have no strategic incentive to limit their money balance) then this should achieve the first best. If the Nash bargaining solution is the outcome, then this will only achieve the first-best if buyers have all the bargaining power.

If instead the charge card equilibrium is better, then the question becomes can the agents in this economy coordinate on the charge card equilibrium? If so, for example if charge cards are already being used, then the optimal monetary policy is to preserve this equilibrium by keeping the nominal rate above i^*. If spending limits are binding and limits are still sensitive to interest rates, then the optimal policy is to increase money growth until the rate is $\tilde{i}$. Such a policy is optimal but will never achieve the first best allocation as long as there is a proportional merchant fee $\tau > 0$.

If the charge card equilibrium is better but the economy is stuck in the monetary equilibrium, a monetary authority might promote a switch between equilibria by raising the interest rate to $\bar{i}$ and reducing the fraction of agents who must simultaneously deviate. This will be very painful, however, if the economy does not move to the charge card equilibrium.

Viewed another way, this model can be used to make positive statements about monetary policy in the real world. Intermediated exchange is a fact of life. Money is a solution to the problem of anonymity, but it is not the only solution. This can provide part of the explanation for why we almost never see the Friedman rule actually used. Positive inflation may not be as costly as we think if consumers have access to a variety of media of exchange. Moreover, some inflation may actually allow intermediaries to flourish and help overcome intrinsic frictions in the economy.

Appendix 2.A Proofs

Lemma 2 *Under Nash bargaining with money only, $z \geq z^*$ does not maximize the buyer's surplus. The maximum is at some smaller $z = z^\dagger$ at which the quantity exchanged is less than the social optimum $q^\dagger \equiv \hat{q}(z^\dagger) < q^*$.*

Proof. The left-sided derivative $B_o'(z^*) < 1$. To see this, first define

$$\hat{B}_o(q) \equiv u(q) - g(q) = [u(q) - c(q)]\frac{\theta u'(q)}{\theta u'(q) + (1-\theta)c'(q)} \tag{2.60}$$

which is the buyer's surplus as a function of quantity. Note that $B_o(z) = \hat{B}_o(\hat{q}(z))$ and note that $\hat{q}'(z) = 1/g'(q)$.

$$\begin{aligned} \hat{B}_o'(q) &= [u'(q) - c'(q)]\frac{\theta u'(q)}{\theta u'(q) + (1-\theta)c'(q)} \\ &+ [u(q) - c(q)]\frac{\theta u''(q)[\theta u'(q) + (1-\theta)c'(q)] - \theta u'(q)[\theta u''(q) + (1-\theta)c''(q)]}{[\theta u'(q) + (1-\theta)c'(q)]^2} \end{aligned} \tag{2.61}$$

which, evaluated at q^* simplifies to

$$\hat{B}_o'(q^*) = [u(q^*) - c(q^*)]\theta(1-\theta)\frac{[u''(q^*) - c''(q^*)]}{u'(q^*)} \tag{2.62}$$

which is negative. It remains to show that $\hat{q}'(z^*) > 0$. Notice that

$$B_o'(z^*) = u'(\hat{q}(z^*))\hat{q}'(z^*) - 1 = \hat{q}'(z^*)\hat{B}_o'(q^*) \tag{2.63}$$

If $\hat{q}'(z^*) < 0$ then the left hand side is negative and the right hand side is positive. Therefore $\hat{q}'(z^*)$ must be positive and $B_o'(z^*)$ is negative. This tells us that as the buyer reduces her money balances going into the DM from z^* to a lower value, the quantity of goods will decrease, but the buyer's surplus will increase. $\square$

Lemma 3 *In Nash bargaining, if the buyer has a card and is carrying no money, then*

1. *The quantity exchanged will be less than the first-best* q^*.

2. *The highest quantity is exchanged when the card's spending limit does not bind. This quantity is* $\tilde{q}$ *which solves* $c'(\tilde{q}) = (1-\tau)u'(\tilde{q})$.

3. *The card will always be used as long as* $\tilde{q} > 0$.

Thus the solution when the buyer carries no money is

$$q_b = \begin{cases} \hat{Q}(0,l) & \text{if } l < l^* \\ \tilde{q} & \text{if } l \geq l^* \end{cases}$$
$$d_c = \begin{cases} l & \text{if } l < l^* \\ l^* & \text{if } l \geq l^* \end{cases} \tag{2.64}$$

where l^* *and* $\hat{Q}(0,l)$ *are defined in the proof below.*

Proof. Consider the potential solutions to the Nash bargaining solution when $z = 0$

Case $d_m = z = 0$, $0 < d_c < l$ $\quad \Rightarrow$ **solution implies** $q = \tilde{q}$

$$q \text{ F.O.C.} \qquad (1-\theta)c'(q)[u(q) - d_c] = \theta u'(q)[-c(q) + (1-\tau)d_c] \tag{2.65}$$

$$d_c \text{ F.O.C.} \qquad (1-\theta)(1-\tau)[u(q) - d_c] = \theta[-c(q) + (1-\tau)d_c] \tag{2.66}$$

This immediately implies $c'(q)/u'(q) = 1 - \tau$. Call this quantity $\tilde{q}$. Since $c(\cdot)$ is convex and $u(\cdot)$ is concave, the ratio $c'(q)/u'(q)$ is non-decreasing in q. Since $c'(q^*)/u'(q^*) = 1 > 1 - \tau = c'(\tilde{q})/u'(\tilde{q})$ I conclude that $\tilde{q} < q^*$.

Solving (2.65) for d_c determines the Nash solution card payment as a function of q. $d_c = g_c(q)$ where

$$g_c(q) \equiv \frac{\theta c(q)u'(q) + (1-\theta)u(q)c'(q)}{\theta(1-\tau)u'(q) + (1-\theta)c'(q)} \tag{2.67}$$

Define $l^* = g_c(\tilde{q})$ to be the card payment that allows this second best $\tilde{q}$ to be exchanged.

Case $d_m = z = 0$, $d_c = l$ $\quad \Rightarrow$ **solution implies** $q = \hat{Q}(0, l)$

If the spending limit constraint binds, then $d_c = l$ and the quantity exchanged will be $q = \hat{Q}(0, l)$ which is defined as the q that solves $l = g_c(q)$.

Case $d_m = z = 0$, $d_c = 0$ $\quad \Rightarrow$ **solution implies** $c'(0)/u'(0) \geq 1 - \tau$

First, note that either $q = 0$ or $q > 0$. With no payment $d_m = d_c = 0$ the seller would have negative gain from trade for any $q > 0$. Even if the buyer has all the bargaining power, I assume that the seller can always walk away from the meeting and attain $S = 0$. Thus $(q, d_m = 0, d_c = 0)$ where $q > 0$ is never a solution.

For (0,0,0) to be the Nash bargaining solution it must be the case that $c'(0)/u'(0) \geq 1 - \tau$. If this were not the case, if $c'(0) < (1-\tau)u'(0)$, then $q = \varepsilon$ and $d_c = u(\varepsilon)$ would be a Pareto improvement over (0,0,0). Some additional structure on $u(q)$ and $c(q)$, such as $\lim_{q \to 0} u'(q) = \infty$ and $\lim_{q \to 0} c'(q) = 0$, would ensure that $d_c = 0$ is never a solution and that if the buyer has no money but does have a card, then the card is used. $\square$

Lemma 4 *In Nash bargaining, if the buyer is carrying money* $z > 0$ *then some payment will be made with money* $d_m > 0$*.*

Proof. Consider the potential solutions to the Nash bargaining solution when $d_m = 0$ but $z > 0$.

Case $d_m = 0$ *(*$z > 0$*),* $0 < d_c < l$ $\Rightarrow$ **not a solution**

d_m F.O.C. $$(1-\theta)B \leq \theta S \tag{2.68}$$

d_c F.O.C. $$(1-\theta)(1-\tau)B = \theta S \tag{2.69}$$

which is a contradiction since $1 - \tau < 1$.

Case $d_m = 0$ *(*$z > 0$*),* $d_c = l$ $\Rightarrow$ **not a solution**

$$d_m \text{ F.O.C.} \qquad (1-\theta)B \leq \theta S \tag{2.70}$$

$$d_c \text{ F.O.C.} \qquad (1-\theta)(1-\tau)B = \theta S + \lambda B^{-\theta} S^{\theta} \tag{2.71}$$

But B, S and λ are all non-negative so

$$(1-\theta)(1-\tau)B \geq \theta S \tag{2.72}$$

$$(1-\theta)B > \theta S \tag{2.73}$$

which is a contradiction.

Case $d_m = z = 0$*,* $d_c = 0$ $\Rightarrow$ **not a solution**

Either $q = 0$ or $q > 0$. Once again note that with no payment the seller will walk away rather than trade $q > 0$. So for $d_m = d_c = 0$ to be a solution we must have $q = 0$. But we have assumed that there exists a $q^* > 0$ such that $c'(q^*) = u'(q^*)$. Since $c(q)$ is convex and $u(q)$ is concave, this means that $c'(0) < u'(0)$. Exchanging $q = \varepsilon$ and $d_m = u(\varepsilon)$ would be a Pareto improvement over $q = 0$, $d_m = 0$, so (0,0,0) is not a solution.

None of the potential exchanges with $d_m = 0$ are in fact solutions to the Nash bargaining problem, so it must be that $d_m > 0$ when $z > 0$. $\square$

Lemma 5 *In Nash bargaining, if the payment made with money is less than the amount of money the buyer is carrying $d_m < z$ then*

1. *No payment is made with a charge card $d_c = 0$.*

2. *The optimal quantity is exchanged $q = q^*$.*

3. *Money payment is the same as without cards $d_m = z^* = g(q^*)$.*

Proof. Consider the potential solutions with $0 < d_m < z$.

Case $0 < d_m < z$, $0 < d_c < l \quad \Rightarrow$ **not a solution**

$$d_m \text{ F.O.C.} \qquad (1-\theta)B = \theta S \tag{2.74}$$

$$d_c \text{ F.O.C.} \qquad (1-\theta)(1-\tau)B = \theta S \tag{2.75}$$

which is a contradiction since $1 - \tau \neq 1$.

Case $0 < d_m < z$, $d_c = l \quad \Rightarrow$ **not a solution**

$$d_m \text{ F.O.C.} \qquad (1-\theta)B = \theta S \tag{2.76}$$

$$d_c \text{ F.O.C.} \qquad (1-\theta)(1-\tau)B = \theta S + \lambda B^{-\theta} S^{\theta} \tag{2.77}$$

Combine to produce $-\tau = \lambda B^{-1-\theta} S^{-\theta}$ which is a contradiction since B, S and λ are all non-negative.

Case $0 < d_m < z$, $d_c = 0$ $\Rightarrow$ **solution implies** $q = q^*$

$$q \text{ F.O.C.} \qquad (1-\theta)c'(q)[u(q) - d_m] = \theta u'(q)[-c(q) + d_m] \tag{2.78}$$

$$d_m \text{ F.O.C.} \qquad (1-\theta)[u(q) - d_m] = \theta[-c(q) + d_m] \tag{2.79}$$

The conditions are identical to those in the Nash bargaining problem with only money when the constraint doesn't bind, and the solution is the same: $c'(q) = u'(q)$ and $d_m = g(q^*)$. □

Lemma 6 *In Nash bargaining, if the payment made with money equals the buyer's money holdings* $d_m = z$*, then*

1. *If the card is used and the spending limit does not bind* $0 < d_c < l$*, the quantity exchanged will be* $q = \tilde{q}$*.*

2. $d_c = 0$ *if* $q > \tilde{q}$*. The card is not used if the quantity the buyer gets when she spends all her money is already more than the optimal quantity that can be achieved with the card.*

The solution when both money and card are used is

$$q_b = \begin{cases} \hat{Q}(z,l) & \text{if } z < j^{-1}(l) \\ \tilde{q} & \text{if } j^{-1}(l) \leq z < \tilde{z} \\ \hat{q}(z) & \text{if } \tilde{z} \leq z < z^* \\ q^* & \text{if } z > z^* \end{cases}$$

$$d_m = \begin{cases} z & \text{if } z < z^* \\ z^* & \text{if } z \geq z^* \end{cases} \tag{2.80}$$

$$d_c = \begin{cases} l & \text{if } z < j^{-1}(l) \\ j(z) & \text{if } j^{-1}(l) \leq z < \tilde{z} \\ 0 & \text{if } z \geq \tilde{z} \end{cases}$$

where $j^{-1}(l)$ *is the* z *that solves* $j(z) = l$, $\tilde{z} = g(\tilde{q})$, *and* $\hat{Q}(z,l)$ *and* $j(z)$ *are defined in the proof below.*

Proof. Consider the potential solutions with $d_m = z$ and $z > 0$.

Case $d_m = z$, $0 < d_c < l$ $\Rightarrow$ **solution implies** $q = \tilde{q}$

q F.O.C. $$(1-\theta)c'(q)[u(q) - z - d_c] = \theta u'(q)[-c(q) + z + (1-\tau)d_c] \tag{2.81}$$

$$d_c \text{ F.O.C.} \quad (1-\theta)(1-\tau)[u(q)-z-d_c] = \theta[-c(q)+z+(1-\tau)d_c] \tag{2.82}$$

It follows immediately that $c'(q)/u'(q) = 1-\tau$ and $q = \tilde{q}$. Solve (2.81) for d_c to find

$$d_c = \frac{\theta[c(\tilde{q})-z]u'(\tilde{q}) + (1-\theta)[u(\tilde{q})-z]c'(\tilde{q})}{\theta(1-\tau)u'(\tilde{q}) + (1-\theta)c'(\tilde{q})} \tag{2.83}$$

$$= \frac{\theta c(\tilde{q})u'(\tilde{q}) + (1-\theta)u(\tilde{q})c'(\tilde{q})}{\theta(1-\tau)u'(\tilde{q}) + (1-\theta)c'(\tilde{q})} - z\frac{\theta\tau u'(\tilde{q}) + \theta(1-\tau)u'(\tilde{q}) + (1-\theta)c'(\tilde{q})}{\theta(1-\tau)u'(\tilde{q}) + (1-\theta)c'(\tilde{q})} \tag{2.84}$$

$$= l^* - z - z\frac{\theta\tau u'(\tilde{q})}{\theta(1-\tau)u'(\tilde{q}) + (1-\theta)c'(\tilde{q})} = l^* - z - z\frac{\theta\tau}{1-\tau} \tag{2.85}$$

Thus when the spending limit does not bind $d_c = j(z)$ where

$$j(z) \equiv l^* - z\frac{1-\tau+\theta\tau}{1-\tau} \tag{2.86}$$

Here $j(d_m)$ is the size of the card payment needed to purchase $\tilde{q}$ given that there is also a payment of d_m made with cash. This takes into account the fact that the merchant fee applies only to the part of the payment made by card. Hence $j(0) = l^*$, because this is the card payment needed to buy $\tilde{q}$ when no money is used. $d_m\theta\tau/(1-\tau)$ can be seen as the buyer's share of the savings from paying d_m of the total amount in money rather than with a card. Now we can see that $j^{-1}(l)$ is the value for z below which the card spending limit will bind. Define $\tilde{z} \equiv g(\tilde{q})$, which is the amount of cash needed to buy $\tilde{q}$ when no payment is made using a charge card. It should be clear that $j^{-1}(0) = \tilde{z}$ (when the card cannot be used at all, the card's second best is only achieved if $z \geq g(\tilde{q})$.)

Alternatively, (2.81) can also be solved for z to write $d_m = z = G(q, d_c)$ where

$$G(q, d_c) \equiv \frac{\theta c(q)u'(q) + (1-\theta)u(q)c'(q) - d_c[\theta(1-\tau)u'(\tilde{q}) + (1-\theta)c'(\tilde{q})]}{\theta u'(q) + (1-\theta)c'(q)} \tag{2.87}$$

Case $d_m = z$, $d_c = l$ $\Rightarrow$ **solution implies** $q = \hat{Q}(z, l)$

When both the spending limit and the money holding constraint bind the quantity exchanged is $q = \hat{Q}(z, l)$ which is defined as the q that solves $z = G(q, l)$. Note that this is consistent with the definition of $\hat{Q}(0, l)$ above.

Case $d_m = z > 0$, $d_c = 0$ $\Rightarrow$ **solution implies** $q > \tilde{q}$

$$q \text{ F.O.C.} \qquad (1-\theta)c'(q)[u(q) - d_m] = \theta u'(q)[-c(q) + d_m] \tag{2.88}$$

$$d_c \text{ F.O.C.} \qquad (1-\theta)(1-\tau)[u(q) - d_m] \leq \theta[-c(q) + d_m] \tag{2.89}$$

from which it immediately follows that $c'(q)/u'(q) \geq 1 - \tau$. Since $c(q)$ is convex and $u(q)$ is concave, this implies $q > \tilde{q}$. If the card is not used, then the quantity is already above the second best using the card. $\square$

Lemma 8 Under proportional bargaining when charge cards are used, the outcome is

$$q_b = \begin{cases} \hat{Q}(z,l) & \text{if } z < j^{-1}(l) \\ \tilde{q} & \text{if } j^{-1}(l) \leq z < \tilde{z} \\ \hat{q}(z) & \text{if } \tilde{z} \leq z < z^* \\ q^* & \text{if } z > z^* \end{cases}$$

$$d_m = \begin{cases} z & \text{if } z < z^* \\ z^* & \text{if } z \geq z^* \end{cases} \tag{2.90}$$

$$d_c = \begin{cases} l & \text{if } z < j^{-1}(l) \\ j(z) & \text{if } j^{-1}(l) \leq z < \tilde{z} \\ 0 & \text{if } z \geq \tilde{z} \end{cases}$$

where all values are as defined in the Nash bargaining case, except that now

$$G(q,l) = (1-\theta)u(q) + \theta c(q) - (1-\tau\theta)l \tag{2.91}$$

Proof. Using α as the Lagrange multiplier for the proportional utility constraint, μ for the money constraint and λ for the spending limit, the Kuhn-

Tucker conditions are

$$u'(q) - \alpha\frac{\theta}{\theta-1}c'(q) - \alpha u'(q) \leq 0 \qquad = 0 \text{ if } q > 0 \tag{2.92}$$

$$-1 - \mu + \alpha\frac{\theta}{\theta-1} + \alpha \leq 0 \qquad = 0 \text{ if } d_m > 0 \tag{2.93}$$

$$-1 - \lambda + \alpha\frac{\theta(1-\tau)}{\theta-1} + \alpha \leq 0 \qquad = 0 \text{ if } d_c > 0 \tag{2.94}$$

$$\mu(z - d_m) = 0 \qquad \lambda(l - d_c) = 0 \qquad q \geq 0 \qquad 0 \leq d_m \leq z \qquad 0 \leq d_c \leq l$$

$$(1-\theta)[u(q_b) - d_m - d_c] = \theta[-c(q_b) + d_m + (1-\tau)d_c] \tag{2.95}$$

Both d_m and d_c could potentially be an interior solution or one of two corner solutions. In addition, consider the case where $d_m = z = 0$. Following the steps in the proofs for lemmas 2 through 6, try each of these 4×3 combinations in turn. The same cases lead to contradictions as under Nash bargaining, and the same conditions on q apply leading to the same breakpoints for z. The only difference between this solution and the Nash solution is that $G(q, l)$ now comes from rearranging the constraint that the buyer's gain must be proportional to the seller's. This difference carries through to $z^* = G(q^*, 0)$, $\hat{Q}(z, l)$ which solves $z = G(q, l)$, and l^* which solves $0 = G(\tilde{q}, l)$ (and thus, by extension $j(z)$ and $j^{-1}(l)$). $\square$

BIBLIOGRAPHY

Aruoba, S. Boragan and Randall Wright, "Search, money, and capital: a neoclassical dichotomy," *Journal of Money, Credit, and Banking*, 2003, *35* (6), 1085–1105.

_ , **Guillaume Rocheteau, and Christopher Waller**, "Bargaining and the value of money," *Journal of Monetary Economics*, 8 2007, *54*, 2636–2655.

Berentsen, Aleksander, Gabriele Camera, and Christopher Waller, "Money, credit and banking," *Journal of Economic Theory*, 2007, *135*, 171–195.

Camera, Gabriele, "Money, search, and costly matchmaking," *Macroeconomic Dynamics*, 2000, *4*, 289–323.

Chakravorti, Sujit, "Theory of credit card networks: a survey of the literature," *Review of Network Economics*, 2003, *2* (2), 50–68.

Dong, Mei, "Money and Costly Credit," *Bank of Canada Working Paper*, 2007, (2011-7).

He, Ping, Lixin Huang, and Randall Wright, "Money, banking, and monetary policy," 2005. University of Pennsylvania.

Kiyotaki, Nobuhiro and Randall Wright, "On money as a medium of exchange," *Journal of Political Economy*, 1989, *97* (4), 927–954.

Kocherlakota, Narayana R., "Money is memory," *Journal of Economic Theory*, 1998, *81*, 232–251.

Lagos, Ricardo and Randall Wright, "A unified framework for monetary theory and policy analysis," *Journal of Political Economy*, 2005, *113*, 463–484.

Levitin, Adam J., "Priceless? The competitive costs of credit card merchant restraints," *Georgetown Law and Economics Research Paper*, 2007, (973974).

Molico, Miguel, "The distribution of money and prices in search equilibrium," *International Economic Review*, 2006, *47* (3), 701–722.

Shi, Shouyong, "Credit and money in a search model with divisible commodities," *Review of Economic Studies*, 1996, *63*, 627–652.

_ , "A divisible search model of fiat money," *Econometrica*, 1997, *65*, 75–102.

Telyukova, Irina A. and Randall Wright, "A model of money and credit, with application to the credit card puzzle," *Review of Economic Studies*, 2008, *75*, 629–647.

Townsend, Robert M., "Financial structure and economic activity," *The American Economic Review*, 1983, *73* (5), 463–484.