Two-sided Market, R&D and Payments System Evolution[*]

Bin Grace Li,[†]James McAndrews[‡]and Zhu Wang[§]

May 2019

Abstract

It takes many years for more efficient electronic payments to be widely used, and the fees that merchants (consumers) pay for using those services are increasing (decreasing) over time. We address these puzzles by studying payments system evolution with a dynamic model in a two-sided market setting. We calibrate the model to U.S. payment card data and conduct welfare and policy analysis. Our analysis shows that the market power of electronic payment networks plays important roles in explaining the slow adoption and asymmetric price changes, and the welfare impact of regulations may vary significantly through the endogenous R&D channel.

Keywords: Two-sided Market, Technology Adoption, R&D, Payments System

JEL Classification: E4, G2, O3

[†]International Monetary Fund. Email: bli2@imf.org.
[‡]TNBUSA Inc. Email: jamesjmcandrews@gmail.com.
[§]Corresponding author. Research Department, Federal Reserve Bank of Richmond, P.O. Box 27622, Richmond, VA 23261, USA. Email: zhu.wang@rich.frb.org.

1   Introduction

The payments system is a vital component of the overall economy and its scale is enormous. It is estimated that U.S. consumers made 158 billion purchases worth about $8.3 trillion in 2011, not counting payments made by businesses or governments. The past several decades have also witnessed significant changes in payment technology, particularly the migration from paper to electronic payments. As shown in Fig. 1, the share of card relative to cash payments increases steadily over time, as does the share of the broadly measured electronic payments relative to paper payments.

In some ways these developments are puzzling. One key question, for example, is why it has taken so long for electronic payments to replace paper payments. For several decades, experts on payments systems have forecast the imminent arrival of a completely electronic, paperless payments system. While the direction of development seems quite clear, the pace has been slow. Most electronic payments that we are familiar with today were introduced decades ago, for example, credit cards in 1950s, ACH in 1970s, and debit cards in 1980s. However, even until the 2000s, their combined market share was just around 50-60 percent (see Fig. 1). The slow transition is also happening to some more recent payment innovations, such as the mobile payments.

Also, there have been controversies regarding the competitive efficiency of electronic payments systems, particularly for card-based payments. Merchants claim that payment card networks (e.g., Visa and MasterCard) and their issuing banks have used market power to drive up the fees that merchants pay for accepting credit and debit card payments (the so-called interchange fees, shown in Fig. 2).[1] This is in sharp contrast to the continuing technological progress and cost reduction in the payment card industry, and has resulted in major regulations and a spectacular number of antitrust litigations in recent years.[2]

---

[1]The fees that merchants pay to accept payment cards are called merchant discount rates. A major component of the merchant discount rates is interchange fees, which are set by card networks and paid by merchants to card issuers through merchant acquirers. In the U.S. and many other countries, the increasing interchange fees have driven up merchants' costs of accepting cards.

[2]For example, Congress passed the Durbin Amendment as a provision of the Dodd-Frank Act in July 2010, which directs the Federal Reserve Board to regulate U.S. debit card interchange fees. Subsequently, the Federal Reserve Board implements Regulation II which caps debit card interchange fees at about half of the pre-regulation level. Recently, Visa, MasterCard and their major issuers reached a $6.2 billion settlement agreement in September 2018 with U.S. retailers in a lawsuit over credit and debit card interchange fees. The settlement, pending court approval, aims to end more than 50 lawsuits that merchants have filed since 2005, and would be the largest private antitrust settlement in U.S. history.
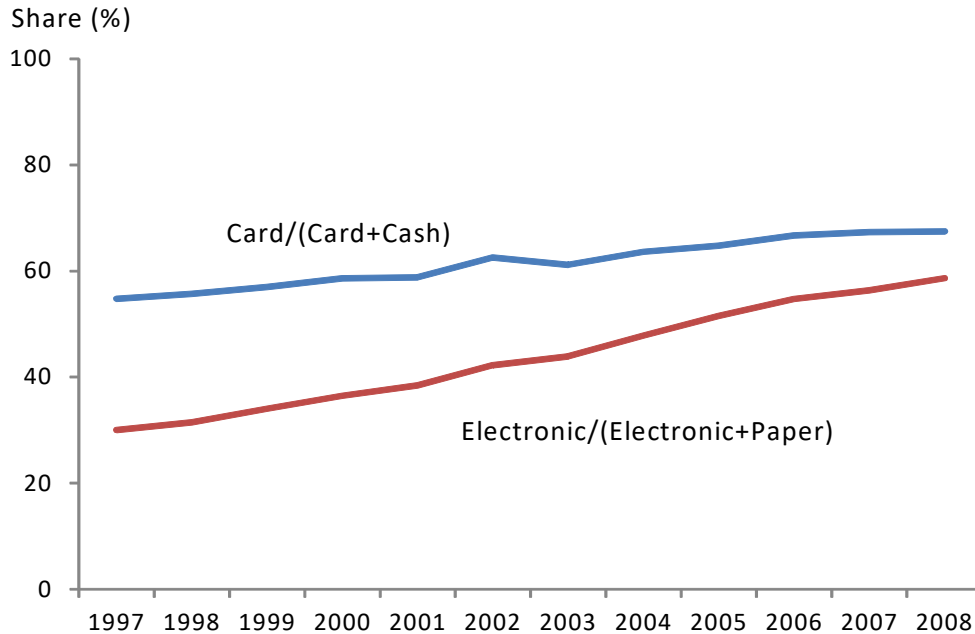
Fig. 1. Evolution of Payment Forms.

Data source: *Nilson Report* (various issues). *Card* includes credit and debit card payments. *Electronic* includes card plus ACH and remote payments. *Paper* includes cash and check payments.

To address these controversies, we develop a unified framework for understanding the payments system evolution. In recent years, a growing literature on payments economics has tried to explore related issues. Some researchers take the money-theoretical approach and use information economics and mechanism design to characterize alternative payments systems (e.g., Kiyotaki and Wright 1989, Lagos and Wright 2005). These theories shed light on how payments arrangements overcome frictions of exchange (e.g., limited information or limited enforcement), but they do not directly explain the slow adoption of electronic payments systems. Moreover, these theories often assume that payment systems are operated by a benevolent planner or by a club of participants, so they do not address the competitive efficiency issues (see Kahn and Roberds 2009, Nosal and Rocheteau 2011 for reviews of the literature).

Some other researchers focus on the industrial organization and pricing issues of payments systems. Particularly, Rochet and Tirole (2002, 2011) argue that the high interchange fees reflect a fundamental failure of the Coase Theorem to hold in the two-sided
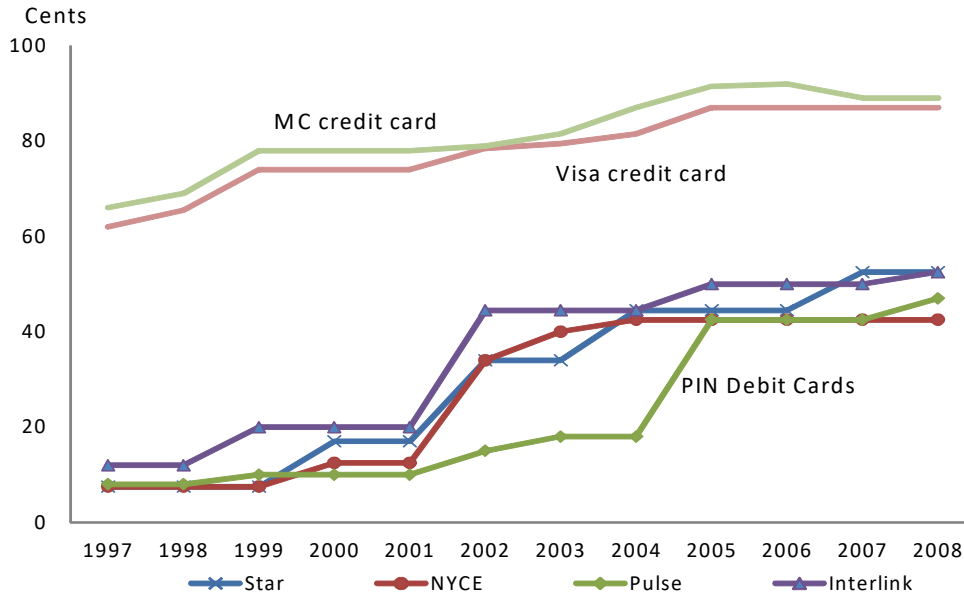
Fig. 2. Payment Card Interchange Fees.

Data source: *American Banker* (various issues). Interchange fees are plotted for a $50 non-supermarket transaction for Visa and MasterCard credit cards plus four largest PIN debit card networks in the U.S.

payment card market (where merchants and consumers are the two sides of card users). While this literature provides important insights into how payment markets function, it has some limitations. For instance, those theories focus on usage externalities of payment means but usually ignore payments adoption decisions and endogenous technological progress. Also, the literature typically imposes an ad hoc distribution of "convenience benefits" from using a particular means of payment and assumes consumers have a fixed demand for goods invariant to their payment choices. Under those assumptions, the adoption and usage patterns of electronic payments are not fully explained, and the welfare analysis is restricted (see Rysman and Wright 2013 for a review of the literature).

In this paper, we provide a new analysis of payments system evolution and the accompanying competitive efficiency issues. Our theory lays out a two-sided market environment where consumers with heterogenous income and merchants of heterogenous size make payment choices under network externalities. This setting is embedded in a fully dynamic model in which a monopoly electronic payment network sets usage fees and conducts R&D

to lower costs. We calibrate the model to match U.S. payment card pricing, adoption and usage data, and conduct welfare and policy analysis.

Our theory has several novelties. First, we model payment adoption decisions under the natural assumption that adopting a payment innovation requires merchants and consumers each to pay a fixed cost, which allows them to potentially benefit from lower marginal cost of making payments. Second, we introduce heterogeneous income (size) among consumers (merchants) to explain their reliance on alternative payment technologies. Finally, we set aside the typical assumption of inelastic demand for final goods by consumers.

Our approach yields clear implications for the adoption and usage pattern of payment innovations. Consider the introduction of payment cards with a high fixed but low marginal cost of use, as compared with cash. More affluent consumers, with higher levels of consumption and purchases, would be more willing to adopt cards than less affluent consumers. Similarly, larger merchants, or those who sell higher-valued goods, would be more likely to accept cards than other merchants. Over time, as the card network conducts R&D to continue lowering card service costs, payment cards penetrate into segments of lower-income consumers and smaller merchants. These predictions are consistent with the empirical evidence (see Figs. A1(a)-(c) in Appendix A1), but the literature that assumes an unspecified "convenience benefit" of using payments cards does not yield such straightforward empirical implications.

It has previously been pointed out (e.g., Wright, 2003) that a merchant serving both cash and card consumers would be competed out of business in a competitive environment. However, in our model setting, we find that large merchants who serve both cash and card customers do survive the threat of entry from specialized merchants. Indeed, our equilibrium is characterized by three categories of merchants based on their size: Large merchants accept both cash and cards, and set a price that is lower than cash-only merchants. They attract both cash users and card users. Medium-size merchants are

specialized. Some of them accept only cash. The others accept both cash and cards, but set a price that is higher than the competing cash-only merchants. They attract only card users, because cards are cost-saving to them overall. Finally, small merchants are all cash-only merchants. These predictions are consistent with what we observe in reality, but are not implied by existing theories.

Our analysis shows that the market power of electronic payment networks plays important roles in explaining the slow adoption of electronic payments. In contrast, we find that a Ramsey social planner who aims to maximize consumer and social welfare would set lower usage fees and conduct more R&D, thus achieving a higher adoption and usage of electronic payments. Our analysis also provides a new explanation for high merchant fees observed in the card payment systems. Imposing a high fee to the merchant side allows a card network to inflate retail prices and extract more rents produced by replacing costly cash payments. According to our model, the card network tends to raise merchant fees but reduce consumer fees as card service costs decline over time, a pattern consistent with the data yet unexplained by prior literature. We find that cash users are disadvantaged by the high merchant fee. However, this is not because they have to subsidize card users as suggested by previous theories (in our model, at merchants where both cash and card are used, cash users are *subsidized* by card users who by their use of a more efficient payment means, contribute to a lower overall costs of payment at the merchant). Instead, the card network sets the merchant fee too high so that fewer stores serve both card and cash users, hence fewer cash users get subsidized by card users.

Our paper also extends the existing literature by offering additional policy insights. Based on the calibrated model, we compare two regulatory approaches: One requires the card network to price at marginal cost, and the other caps the merchant fee. We find that while marginal-cost pricing could maximize consumer welfare in a static setting, it shuts off the card network's R&D and may reduce social welfare in a dynamic setting. In comparison, the merchant fee cap regulation improves consumer welfare without causing

much dynamic inefficiency.

The paper is organized as follows. Section 2 lays out the model setup with two competing payment means, "cash" versus "card." Section 3 characterizes the model equilibrium and dynamic path. Section 4 calibrates the model to U.S. payment card data. Section 5 conducts welfare and policy analysis. Section 6 concludes.

2 The Model

Our model studies pricing, adoption and usage of payment devices. We first lay out a market environment in which only a paper device, "cash," is in use. We then introduce an electronic device, the "card." In our analysis, certain market conditions, such as the variety of goods and the consumer income distribution, are exogenously given.[3]

2.1 A Cash Economy

The market is composed of a continuous distribution of merchants of measure unity. Each merchant sells a distinct consumption good characterized by two parameters: consumer preference α and unit cost μ. The unit cost μ is measured in terms of a numeraire good called *dollar*. For ease of notation, we assume goods with the same value of α share the same value of unit cost denoted as μ_α. It will become clear that μ_α plays no role in our analysis except scaling the level of utility.

Cash is the sole payment device in the economy. Merchants incur a transaction cost τ_m per dollar for accepting cash, which includes handling, safekeeping and fraud expenses. The market is contestable so merchants earn zero profits. As a result, the cash price for a good α is determined as $p_{\alpha,h}$, where

$$p_{\alpha,h} = \frac{\mu_\alpha}{1 - \tau_m}. \tag{1}$$

---

[3]The static environment of the model follows McAndrews and Wang (2012) closely, upon which we construct a fully dynamic model and conduct quantitative analysis.

Each period, a consumer receives her income I in terms of the numeraire good (which could be interpreted as the payoff to her endowment of effective labor). The consumer, indexed by her income I, purchases a variety of goods from merchants, each of which is associated with a preference parameter α, where $\alpha \in (0, \overline{\alpha})$. The consumer has a Cobb-Douglas utility U_I, and her consumption of each good α is denoted as $x_{\alpha,I}$. Let $G(\alpha)$ be the cumulative distribution function of α and denote $E(\alpha) = \int_0^{\overline{\alpha}} \alpha dG(\alpha)$, so

$$\ln U_I = \int_0^{\overline{\alpha}} \frac{\alpha}{E(\alpha)} \ln x_{\alpha,I} dG(\alpha), \tag{2}$$

where $\int_0^{\overline{\alpha}} \frac{\alpha}{E(\alpha)} dG(\alpha) = 1$. The consumer's budget constraint is

$$\int_0^{\overline{\alpha}} (1 + \tau_c) p_{\alpha,h} x_{\alpha,I} dG(\alpha) \leqslant I, \tag{3}$$

where τ_c is the consumer's transaction cost of using cash. For simplicity, we assume there is no intertemporal saving technology, so income is fully consumed each period. Maximizing the utility (2) subject to the constraint (3) yields consumer I's demand for each good α:

$$x_{\alpha,I}^* = \frac{\alpha I}{E(\alpha)(1 + \tau_c) p_{\alpha,h}}. \tag{4}$$

As a result, consumer I enjoys the utility level $U_{I,h}$ such that

$$\ln U_{I,h} = \ln I + \int_0^{\overline{\alpha}} \frac{\alpha}{E(\alpha)} \ln \frac{\alpha}{E(\alpha)(1 + \tau_c) p_{\alpha,h}} dG(\alpha). \tag{5}$$

This implies that a consumer's utility is proportional to her income,

$$U_{I,h} = I \overline{U_h}, \tag{6}$$

where $\overline{U_h} = \exp\left\{ \int_0^{\overline{\alpha}} \frac{\alpha}{E(\alpha)} \ln \frac{\alpha}{E(\alpha)(1+\tau_c) p_{\alpha,h}} dG(\alpha) \right\}$ is the utility associated with one dollar of income.

Income I varies across consumers and is distributed according to a cumulative distribution function $F(I)$ on the support $(0, \overline{I})$. We denote $E(I) = \int_0^{\overline{I}} I dF(I)$ as the mean income and normalize the measure of consumers to be unity. At equilibrium, a merchant's sale of good α equals its total demand:

$$x_\alpha = \int_0^{\overline{I}} x_{\alpha,I} dF(I) = \frac{\alpha E(I)}{E(\alpha)(1 + \tau_c) p_{\alpha,h}}. \tag{7}$$

It follows from Eq (7) that the size of a merchant, measured by the value of sales $p_{\alpha,h} x_\alpha$, increases with the preference index α of the good that the merchant sells.

2.2   Introducing the Payment Card

We now introduce an electronic payment innovation, referred to as a payment card. The card service is provided by a monopoly network. The costs of providing the card service to merchants and consumers are d_m and d_c per-dollar transaction respectively, and we denote the sum $d = d_m + d_c$. It will become clear that in our two-sided market setting, only the sum d (but not its composition, d_m and d_c) matters for the analysis. In return, the card service provider charges merchants and consumers a percentage fee f_m and f_c, respectively.[4] Figure 3 describes the transaction flow in the card system in which consumers use a payment card to pay merchants. Merchants submit charges to the card network which then bills consumers.[5]

---

[4] Assuming payment cards charge percentage fees is consistent with reality. In most countries, credit cards charge fees proportional to the transaction value. In the U.S., most debit cards also charge percentage fees. Shy and Wang (2011) and Wang and Wright (2017) provide theoretical explanations as to why many platform industries favor percentage fees.

[5] For simplicity, we model a "three-party" system where the payment card network serves consumers and merchants directly, but our analysis can equally apply to a "four-party" system where the card network serves consumers and merchants indirectly through card issuers and merchant acquirers. In the literature, it is typically considered that merchant acquirers are competitive and the card network maximizes the joint profits of member issuers, so we can simply reinterpret the card network in our model to be the association of member issuers in a "four-party" system.
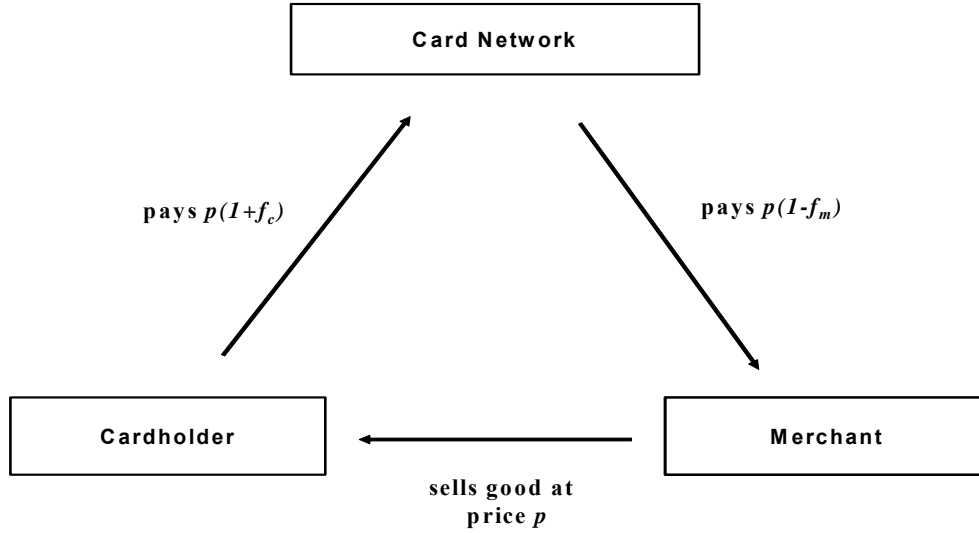
Fig. 3. Payment Card System.

We adopt the convention that the transaction costs of using payment cards are normalized to 0 for both merchants and consumers. Therefore, the avoided costs of handling cash, τ_m and τ_c, are the convenience benefits of card payments to merchants and consumers, respectively. For the card being a more efficient payment means, we impose the condition:

$$\tau_m + \tau_c > d. \tag{8}$$

In order to adopt the payment card, a merchant and a consumer each incur a fixed cost K_m and K_c every period, respectively. The fixed costs may arise for technological or institutional reasons. For example, a merchant may incur a fixed cost renting card-processing equipment and/or training employees, while a consumer may incur a fixed cost maintaining her bank balance or credit score. Therefore, in deciding whether to accept or hold the payment card, merchants and consumers need to weigh the benefit of avoiding handling cash against their card adoption costs and usage fees. For merchants who decide to accept payment cards, they still have to accept cash and charge a single price to card

and cash customers.[6]

To study the industry evolution, we proceed in two steps with the following analysis.

(i) Within-Period Analysis. At a point in time, with the card service cost d given, we solve for a subgame perfect Nash equilibrium of the following three-stage game:

Stage I. The payment card network sets the card fees f_m and f_c.

Stage II. After observing f_m and f_c, merchants make card acceptance decisions and post retail prices for the period. Meanwhile, consumers decide whether to hold cards.

Stage III. Consumers then decide whether to purchase, which merchants to purchase from, and what payment device to use.

In making the decisions, consumers and merchants each maximize their utility or profits, and the card network sets card fees to maximize profit.

(ii) Dynamic Analysis. The industry evolves over time. To study the industry evolution, we assume time is discrete and the payment card is introduced at time 0. We add back the time subscripts to variables. On the one hand, the mean consumer income $E(I_t)$ may grow exogenously, together with the changes of card adoption costs $K_{m,t}$ and $K_{c,t}$. On the other hand, the card network can make endogenous R&D investment R_t (in units of the numeraire good) to reduce the card service cost next period such that

$$d_{t+1} = \Gamma(d_t, R_t),$$

where $\partial \Gamma / \partial d_t > 0$ and $\partial \Gamma / \partial R_t < 0$. Given the initial value of d_0 and the laws of motion for $E(I_t)$, $K_{m,t}$ and $K_{c,t}$, the card network chooses a sequence of card fees $(f_{m,t}, f_{c,t})$ and R&D investment R_t to maximize the present value of profits.

---

[6]This is referred to as "price coherence" in the literature, and is commonly seen in reality. Price coherence may result from network rules or state regulation (e.g., In the U.S., card network rules and some state laws explicitly prohibit surcharging on payment cards). It can also result from high transaction costs for merchants to price discriminate based on payment means (e.g., even if surcharging payment cards were allowed, few merchants would choose to do so).

3 Market Equilibrium

In this section, we study the market equilibrium under the profit-maximizing network. We first study the industry equilibrium within a period, for which the industry has given values of d, $E(I)$, K_m and K_c. We then embed that in a fully dynamic setting to study the industry evolution.

3.1 Within-Period Analysis

At a point in time, the industry has given values of d, $E(I)$, K_m and K_c. The market equilibrium can be solved backward for the three-stage game. In Stages II and III, given card fees f_m and f_c set by the card network, merchants and consumers each decide their own card adoption and usage, taking others' decisions as given. In Stage I, anticipating consumers' and merchants' responses, the card network decides the profit-maximizing card fees.

Note that the interactions between the two sides of the market can easily yield multiple equilibria. For example, an equilibrium may exist at which neither merchants nor consumers adopt cards. In the following analysis, we will focus on an equilibrium with positive card adoption and usage. We show that there exist threshold values of merchant size and consumer income, above which merchants and consumers accept and hold cards. We will first construct such an equilibrium and then characterize its properties.

3.1.1 Merchants' Choices

Merchants take card fees as given and expect that consumers $I \geq I_0$ will hold cards when making their card acceptance decision. Recall that the index α indicates merchant size. It can be shown that there exist two threshold values α_1 and α_0, where $\alpha_1 > \alpha_0$, so that merchants fall into three categories: (1) Large merchants ($\alpha \geq \alpha_1$) accept both cash and cards, and charge price $p_{\alpha,d} \leq p_{\alpha,h}$ so they are patronized by both cash and

card customers. (2) Intermediate merchants ($\alpha_0 \leq \alpha < \alpha_1$) specialize. Some accept both cash and cards, and charge $p_{\alpha,d}$ where $\frac{1+\tau_c}{1+f_c}p_{\alpha,h} \geq p_{\alpha,d} > p_{\alpha,h}$, so they are patronized only by card customers. The others do not accept cards and charge $p_{\alpha,h}$, so they only serve cash customers. (3) Small merchants ($\alpha < \alpha_0$) only accept cash, so all customers make purchases there using cash regardless of whether they have adopted a card.

As we will show next, the thresholds α_1 and α_0 are endogenously determined under card fees f_m and f_c. Note that because merchants who accept payment cards still have to accept cash and charge a single price to card and cash users, the consumer card fee has to be below the consumer cost of handling cash, i.e., $f_c \leq \tau_c$. This condition can be thought of as a card pricing constraint to provide consumers "incentive at the counter" to use payment cards. As we show in the following analysis, card fees also need to satisfy additional pricing constraints that make card-accepting stores attractive to consumers, in other words, providing consumers "incentive at the door."

Large Merchants: $\alpha \geq \alpha_1$. Merchants in this category charge $p_{\alpha,d} \leq p_{\alpha,h}$ and receive revenues from both card customers ($I \geq I_0$) and cash customers ($I < I_0$):

$$p_{\alpha,d}x_{\alpha,d}^{card} = \frac{\alpha[E_{I \geq I_0}(I - K_c)]}{E(\alpha)(1+f_c)}, \quad p_{\alpha,d}x_{\alpha,d}^{cash} = \frac{\alpha[E_{I<I_0}(I)]}{E(\alpha)(1+\tau_c)}, \tag{9}$$

where $E_{I \geq I_0}(I) \equiv \int_{I_0}^{\overline{I}} I dF(I)$.

Because merchants need to post their prices and cannot make changes within the same period, they each set a lowest possible price to prevent being undercut by potential entrants. This contestable market assumption leads to zero profit so that a merchant's total revenue equals total cost,

$$(1 - f_m)p_{\alpha,d}x_{\alpha,d}^{card} + (1 - \tau_m)p_{\alpha,d}x_{\alpha,d}^{cash} = \mu_\alpha x_{\alpha,d}^{card} + \mu_\alpha x_{\alpha,d}^{cash} + K_m. \tag{10}$$

Equations (9) and (10) pin down the price $p_{\alpha,d}$:

$$p_{\alpha,d} = \frac{\mu_\alpha \frac{\alpha[E_{I \geq I_0}(I-K_c)]}{(1+f_c)} + \mu_\alpha \frac{\alpha[E_{I<I_0}(I)]}{(1+\tau_c)}}{(1-f_m)\frac{\alpha[E_{I \geq I_0}(I-K_c)]}{1+f_c} + (1-\tau_m)\frac{\alpha[E_{I<I_0}(I)]}{1+\tau_c} - K_m E(\alpha)}. \tag{11}$$

Recall from Eq. (1) that $p_{\alpha,h} = \mu_\alpha/(1-\tau_m)$. Hence, $p_{\alpha,d} \leq p_{\alpha,h}$ implies that there exists a threshold

$$\alpha_1 = \frac{K_m E(\alpha)}{[E_{I \geq I_0}(I-K_c)](\frac{1-f_m}{1+f_c} - \frac{1-\tau_m}{1+f_c})} \quad \text{if} \quad f_m \leq \tau_m, \tag{12}$$

so that merchants whose size $\alpha \geq \alpha_1$ fall into this category. Note Eq. (12) suggests that no merchant would belong to this category if $f_m > \tau_m$, because the prices they offer would not attract cash users.

Medium-size Merchants: $\alpha_0 \leq \alpha < \alpha_1$. Merchants in this category specialize. For each good, there are two merchants. One accepts both cash and card, and charges $p_{\alpha,d}$ where $\frac{1+\tau_c}{1+f_c}p_{\alpha,h} \geq p_{\alpha,d} > p_{\alpha,h}$, so it is patronized only by card customers (i.e., $I \geq I_0$).[7] The other does not accept cards and charges $p_{\alpha,h}$, so it only serves cash customers (i.e., $I < I_0$).

A card-accepting merchant in this category receives revenues only from card customers and earns zero profit, which implies

$$p_{\alpha,d} = \frac{\mu_\alpha \frac{\alpha[E_{I \geq I_0}(I-K_c)]}{(1+f_c)}}{(1-f_m)\frac{\alpha[E_{I \geq I_0}(I-K_c)]}{1+f_c} - K_m E(\alpha)}. \tag{13}$$

Therefore, $\frac{1+\tau_c}{1+f_c}p_{\alpha,h} \geq p_{\alpha,d} > p_{\alpha,h}$ implies that merchants $\alpha_0 \leq \alpha < \alpha_1$ are in this group, where α_1 is given in Eq. (12) and α_0 is determined by

$$\alpha_0 = \frac{K_m E(\alpha)}{[E_{I \geq I_0}(I-K_c)](\frac{1-f_m}{1+f_c} - \frac{1-\tau_m}{1+\tau_c})}. \tag{14}$$

---

[7]Because $p_{\alpha,d} > p_{\alpha,h}$, the merchant would not attract cash customers. But given $(1+f_c)p_{\alpha,d} \leq (1+\tau_c)p_{\alpha,h}$, the merchant would attract card users because cards are cost-saving to them overall.

Equations (12) and (14) suggest that $\frac{1-f_m}{1+f_c} \geq \frac{1-\tau_m}{1+\tau_c}$ has to hold in order for any merchants in the market to accept cards. In fact, if $\frac{1-f_m}{1+f_c} < \frac{1-\tau_m}{1+\tau_c}$, there would be no merchant in either category (1) or (2).[8]

Small Merchants: $\alpha < \alpha_0$. Given $\frac{1-f_m}{1+f_c} \geq \frac{1-\tau_m}{1+\tau_c}$, small merchants $\alpha < \alpha_0$ are in the third category. Due to small transaction values, these merchants cannot afford paying the fixed adoption cost to accept cards. Otherwise, it would result in $p_{\alpha,d} > \frac{1+\tau_c}{1+f_c} p_{\alpha,h}$. Therefore, they only accept cash and all consumers, regardless of holding a card or not, make purchases there using cash.

3.1.2 Consumers' Choices

An individual consumer takes market prices, card fees and merchants' card acceptance as given and decides whether to hold a payment card or not. Recall that merchants in the market fall into three categories according to their sizes. A consumer I, if not adopting the card, may use cash and derive utility $U_{I,d}^h$ from shopping at card-acceptance stores in category (1) and at cash-only stores in categories (2) and (3). Accordingly,

$$\ln U_{I,d}^h = \int_0^{\min(\alpha_1,\overline{\alpha})} \frac{\alpha}{E(\alpha)} \ln \frac{\alpha I}{(1+\tau_c)p_{\alpha,h}E(\alpha)} dG(\alpha) + \int_{\min(\alpha_1,\overline{\alpha})}^{\overline{\alpha}} \frac{\alpha}{E(\alpha)} \ln \frac{\alpha I}{(1+\tau_c)p_{\alpha,d}E(\alpha)} dG(\alpha).$$

In contrast, if she holds a payment card, she can use the card to shop at card-accepting stores in categories (1) and (2), though she still needs to use cash at stores in category (3). As a result, she derives utility $U_{I,d}^d$ so that

$$\ln U_{I,d}^d = \int_0^{\alpha_0} \frac{\alpha}{E(\alpha)} \ln \frac{\alpha(I-K_c)}{(1+\tau_c)p_{\alpha,h}E(\alpha)} dG(\alpha) + \int_{\alpha_0}^{\overline{\alpha}} \frac{\alpha}{E(\alpha)} \ln \frac{\alpha(I-K_c)}{(1+f_c)p_{\alpha,d}E(\alpha)} dG(\alpha).$$

---

[8]Because $\tau_c \geq f_c$ has to hold for any consumers to use cards, $\frac{1-f_m}{1+f_c} < \frac{1-\tau_m}{1+\tau_c}$ implies $f_m > \tau_m$.

Therefore, consumer card adoption requires $\ln U_{I,d}^d \geq \ln U_{I,d}^h$, which implies

$$E_{\alpha \geq \alpha_0}(\alpha) \ln(\frac{1 + \tau_c}{1 + f_c}) \geq E(\alpha) \ln(\frac{I}{I - K_c}) + \int_{\alpha_0}^{\min(\alpha_1, \overline{\alpha})} \alpha \ln(\frac{p_{\alpha,d}}{p_{\alpha,h}}) dG(\alpha), \qquad (15)$$

where $E_{\alpha \geq \alpha_0}(\alpha) \equiv \int_{\alpha_0}^{\overline{\alpha}} \alpha dG(\alpha)$.

Equation (15) suggests that an adopter's income has to satisfy $I \geq I_0$:

$$I_0 = \frac{(\frac{1 + \tau_c}{1 + f_c})^{E_{\alpha \geq \alpha_0}(\alpha)/E(\alpha)} K_c}{(\frac{1 + \tau_c}{1 + f_c})^{E_{\alpha \geq \alpha_0}(\alpha)/E(\alpha)} - \exp(\int_{\alpha_0}^{\min(\alpha_1, \overline{\alpha})} \frac{\alpha}{E(\alpha)} \ln(\frac{p_{\alpha,d}}{p_{\alpha,h}}) dG(\alpha))}, \qquad (16)$$

where

$$(\frac{p_{\alpha,d}}{p_{\alpha,h}})_{\alpha_0 \leq \alpha < \min(\alpha_1, \overline{\alpha})} = \frac{\frac{1 - \tau_m}{1 + f_c}[E_{I \geq I_0}(I - K_c)]}{\frac{1 - f_m}{1 + f_c}[E_{I \geq I_0}(I - K_c)] - \frac{E(\alpha)}{\alpha} K_m} \qquad (17)$$

follows Eqs. (1) and (13).

3.1.3 Two-sided Market Interactions

As shown above, the interactions between consumer and merchant card adoption are described by the threshold equations (12), (14), (16), and (17).

To simplify the notations, we denote:

$$Z_1 = (\frac{1 - f_m}{1 + f_c} - \frac{1 - \tau_m}{1 + f_c}), \quad Z_0 = (\frac{1 - f_m}{1 + f_c} - \frac{1 - \tau_m}{1 + \tau_c}). \qquad (18)$$

And we can summarize our findings in the following proposition.

Proposition 1 *Given card fees (f_c and f_m) that satisfy*

$$\tau_c \geq f_c \quad and \quad \frac{1 - f_m}{1 + f_c} \geq \frac{1 - \tau_m}{1 + \tau_c}, \qquad (19)$$

there exist threshold values of merchant size (α_0 and α_1) and consumer income(I_0), above

which merchants and consumers accept and hold cards, where

$$\alpha_0 = \frac{K_m E(\alpha)}{[E_{I \geq I_0}(I - K_c)]Z_0}, \tag{20}$$

$$\alpha_1 = \frac{Z_0}{Z_1}\alpha_0 \quad if \quad f_m \leq \tau_m, \tag{21}$$

$$I_0 = \frac{(\frac{1+\tau_c}{1+f_c})^{E_{\alpha \geq \alpha_0}(\alpha)/E(\alpha)} K_c}{(\frac{1+\tau_c}{1+f_c})^{E_{\alpha \geq \alpha_0}(\alpha)/E(\alpha)} - \exp(\int_{\alpha_0}^{\min(\alpha_1, \overline{\alpha})} \frac{\alpha}{E(\alpha)} \ln \frac{(1-\tau_m)\alpha}{(1-f_m)\alpha - (1+f_c)\alpha_0 Z_0} dG(\alpha))}. \tag{22}$$

3.1.4 Within-Period Equilibrium

The card network, anticipating merchants' and consumers' card adoption and usage decisions in Stages II and III, would set card fees (f_c, f_m) in Stage I to maximize its profit. Accordingly, the card network solves the following problem:

$$\pi(d; E(I), K_m, K_c) = \max_{f_c, f_m} \frac{E_{\alpha \geq \alpha_0}(\alpha) E_{I \geq I_0}(I - K_c)}{E(\alpha)(1 + f_c)}(f_c + f_m - d) \tag{23}$$

$$s.t. \quad Eqs. \ (19), (20), (21), (22).$$

As shown in (23), the network's profit equals consumers' total spending on cards times its markup. In order to maximize profit, the network internalizes the two-sided market externalities by setting the optimal card fees (f_c^*, f_m^*) which determine the card adoption thresholds for merchants and consumers. Given the optimal card fees (f_c^*, f_m^*), the market achieves a within-period equilibrium, at which the card network maximizes profit, consumers maximize utility, merchants break even, and goods and payments markets clear.

3.2 Dynamic Analysis

For the dynamic analysis, we add back time subscripts to variables. Over time, the market may evolve due to exogenous changes (e.g., in mean consumer income $E(I_t)$ and card

adoption costs $K_{m,t}$, $K_{c,t}$) and the network's endogenous R&D investment R_t. Embedding the within-period equilibrium into the dynamic setting, the value function of the card network is defined as follows:

$$V(d_t; E(I_t), K_{m,t}, K_{c,t}) = \max_{R_t} \pi(d_t; E(I_t), K_{m,t}, K_{c,t}) - R_t \tag{24}$$
$$+ \beta V(d_{t+1}; E(I_{t+1}), K_{m,t+1}, K_{c,t+1})$$

$$s.t. \ d_{t+1} = \Gamma(d_t, R_t), \tag{25}$$

$$\pi(d_t; E(I_t), K_{m,t}, K_{c,t}) \geq R_t, \tag{26}$$

where $\pi(d_t; E(I_t), K_{m,t}, K_{c,t})$ is the card network's within-period profit, β is the discount rate, and Γ is the R&D function.

To solve the dynamic optimization problem, we can rewrite the constraint (25) as

$$R_t = \Psi(d_t, d_{t+1}), \tag{27}$$

where Ψ is the inverse function of Γ. As a result, if the budget constraint (26) never binds (which is verified in our calibrated model), the problem (24) is equivalent to

$$V(d_t; E(I_t), K_{m,t}, K_{c,t}) = \max_{d_{t+1}} \pi(d_t; E(I_t), K_{m,t}, K_{c,t}) - \Psi(d_t, d_{t+1}) \tag{28}$$
$$+ \beta V(d_{t+1}; E(I_{t+1}), K_{m,t+1}, K_{c,t+1}).$$

As long as the functions π and Ψ and the laws of motion for $E(I_t)$, $K_{m,t}$ and $K_{c,t}$ satisfy standard regularity conditions, the optimal solution to (28) can be pinned down by the following first-order and envelope conditions:

$$\Psi_2(d_t, d_{t+1}) = \beta V'(d_{t+1}; E(I_{t+1}), K_{m,t+1}, K_{c,t+1}), \tag{29}$$

$$V'(d_t; E(I_t), K_{m,t}, K_{c,t}) = \pi'(d_t; E(I_t), K_{m,t}, K_{c,t}) - \Psi_1(d_t, d_{t+1}). \tag{30}$$

Combining (29) and (30) yields a second-order difference equation

$$\Psi_2(d_t, d_{t+1}) = \beta[\pi'(d_{t+1}; E(I_{t+1}), K_{m,t+1}, K_{c,t+1}) - \Psi_1(d_{t+1}, d_{t+2})],$$

which, together with boundary conditions of d_t (i.e., d_0 is given and $d_\infty = 0$) and the laws of motion for $E(I_t)$, $K_{m,t}$ and $K_{c,t}$, pins down the dynamic equilibrium path.

4 Model Calibration

In this section, we calibrate the model to U.S. payment card data. We first lay out functional form assumptions and characterize the model properties, and then choose parameter values to match the model with U.S. credit and debit card pricing, adoption and usage data.

4.1 Functional Forms

To take our model to data, we make the following functional form assumptions F1-F4.

F1. Merchant size and consumer income follow explicit distributions: $\alpha \in (0, 1)$ is uniformly distributed where $G(\alpha) = \alpha$ and $E(\alpha) = 1/2$; and $I_t \in (0, \infty)$ is exponentially distributed where $F(I_t) = 1 - e^{(-\lambda_t I_t)}$ and $E(I_t) = 1/\lambda_t$.

F2. The mean consumer income has a constant growth rate g_I such that $\lambda_{t+1} = \lambda_t/(1 + g_I)$.

F3. The card adoption costs are proportional to the mean consumer income: $K_{m,t} = k_m E(I_t) = k_m/\lambda_t$ and $K_{c,t} = k_c E(I_t) = k_c/\lambda_t$, where k_m and k_c are fixed parameters.

F4. The R&D function Γ takes the following form:

$$\frac{1}{d_{t+1}} - \frac{1}{d_t} = (\frac{R_t \lambda_t}{\phi})^\gamma d_t^{-\gamma-1} \quad \text{with } 1 > \gamma > 0, \tag{31}$$

19

where R&D investment (scaled by income), $R_t/E(I_t) = R_t\lambda_t$, enters the function and ϕ is a fixed parameter measuring the efficiency of R&D.

The functional form assumptions are made mainly for tractability. Note that one assumption in F1, that merchant size α follows a uniform distribution, is inessential, and our model findings continue to hold even if α follows a positively skewed distribution (such as an exponential distribution). Another assumption, that consumer income I_t follows an exponential distribution, matches U.S. data well, as shown by Dragulescu and Yakovenko (2001). Moreover, the exponential distribution has a constant Gini coefficient at 0.5, which also agrees with U.S. data (according to Dragulescu and Yakovenko 2001, the Gini coefficient for U.S. personal income was between 0.47 and 0.56 from 1979-1997). Therefore, this assumption allows our analysis to focus on the change in the mean income while holding the income inequality fixed.

F2-F3 are simplifying assumptions that reduce free parameters of the model so that we can draw sharp results for the dynamic analysis. Particularly, by tying card adoption costs $K_{m,t}$ and $K_{c,t}$ to the mean consumer income $1/\lambda_t$, the exogenous growth of income would affect card transaction values but not the adoption thresholds, with the latter being fully driven by the endogenous decline in card service costs d_t. F4 models the decline in card service costs d_t as a result of R&D, and because R&D may entail opportunity costs of labor hours, we scale it by the mean consumer income (i.e., $R_t/E(I_t) = R_t\lambda_t$). As will become clear, the R&D functional form (31) allows us to derive a balanced growth path in the dynamic analysis.

4.2 Model Characterization

Before we take the model to data, we characterize some important properties of the model equilibrium given the functional form assumptions we made.

4.2.1 Within-Period Equilibrium

We start with the within-period equilibrium. First, given adoption costs, two-sided market interactions may yield multiple equilibria. Assume $\alpha_1 < \bar{\alpha} = 1$. Under our functional form assumptions, Eqs. (20) and (22) can be rewritten into

$$\alpha_0 = \frac{k_m}{2e^{(-\lambda I_0)}(1 + \lambda I_0 - k_c)Z_0} \quad \text{(L1)}, \qquad I_0 = \frac{(\frac{1+\tau_c}{1+f_c})^{1-\alpha_0^2}k_c/\lambda}{(\frac{1+\tau_c}{1+f_c})^{1-\alpha_0^2} - \exp(s\alpha_0^2)} \quad \text{(L2)},$$

where s is a scalar determined by the model parameters (see the proof in Appendix A2).

For plausible parameter values, we plot Eqs. (L1) and (L2) in Fig. 4, which illustrates the interactions between merchant and consumer card adoption. For a given pair of card fees (f_m, f_c), there exist two equilibria with positive levels of card adoption (Note that no adoption is an equilibrium as well): a high-adoption equilibrium (I_0^*, α_0^*) and a low-adoption equilibrium (I_0', α_0'). The high equilibrium is stable while the low equilibrium is not, so we always select the stable high-equilibrium in our following analysis.[9]

Second, under Assumption F3, the optimal card adoption thresholds $(\lambda I_0^*, \alpha_0^*)$ depend on d but not λ. As a result, changing the mean consumer income $1/\lambda$ just shifts the profit function $\pi(d; \lambda)$ proportionally. We can also verify that for a given λ, the network's optimal profit $\pi(d; \lambda)$ is approximately linear in d. Therefore, the profit function can be simplified as

$$\pi(d; \lambda) = \frac{1}{\lambda}(a_0 - a_1 d), \tag{32}$$

where a_0 and a_1 are scalars determined by the model parameters (see Appendix A3 for the detailed derivation).

---

[9]The idea of using stability to select an equilibrium goes back to Cournot (1838), which provided a dynamic adjustment argument by imagining sequential play by each agent myopically best-responding to all rivals. This is referred to as best-reply dynamics, and when the process converges, the solution is termed stable (Fudenberg and Tirole, 1991). In the context of our model, the existence of multiple equilibria implies that when the payment card is initially introduced, the card network may need to push the card adoption to overcome the unstable (low-adoption) equilibrium to achieve the stable (high-adoption) equilibrium. This could be interpreted as the adoption hurdles due to merchant and/or consumer learning.
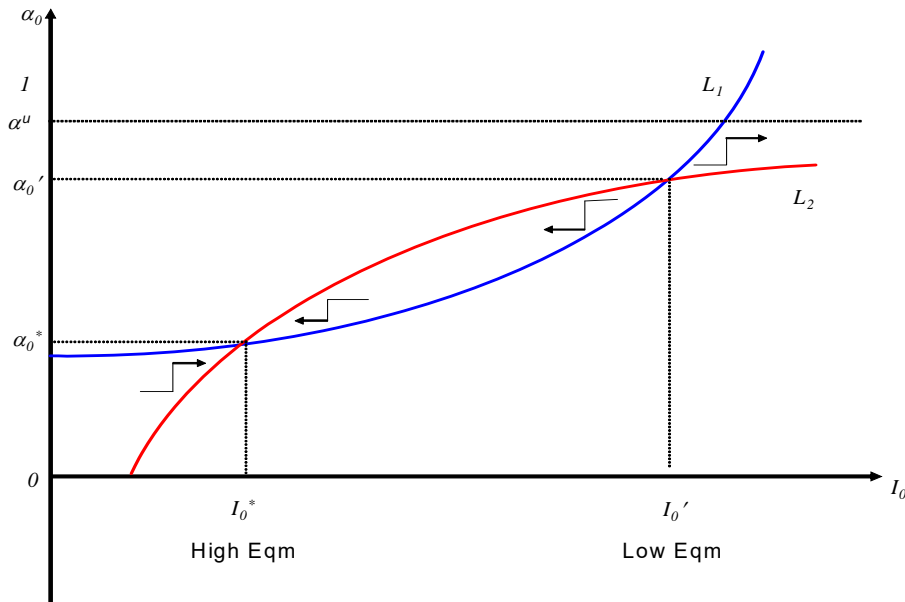
Fig. 4. Interaction of Merchants and Consumers in Card Adoption.

4.2.2 Dynamic Equilibrium

The within-period equilibrium can then be embedded in the dynamic analysis. Note that Assumption F4 implies a R&D investment function such that

$$R_t = \phi \frac{d_t}{\lambda_t} \left[\frac{d_t}{d_{t+1}} - 1 \right]^{\frac{1}{\gamma}} \text{ with } 1 > \gamma > 0.$$

The function, in the spirit of adjustment costs studied in the macro/investment models (e.g., Hall, 2001, 2004), is strictly increasing and convex in technological progress (d_t/d_{t+1}) and constant returns to scale in (d_t, d_{t+1}).[10]

We can then formulate the dynamic problem of the card network as follows:

$$V(d_t; \lambda_t) = \max_{d_{t+1}} \pi(d_t; \lambda_t) - \phi \frac{d_t}{\lambda_t} \left[\frac{d_t}{d_{t+1}} - 1 \right]^{\frac{1}{\gamma}} + \beta V(d_{t+1}; \lambda_{t+1}). \tag{33}$$

---

[10]See Khan and Thomas (2008) for a comprehensive review of the adjustment cost literature. Our modeling of R&D allows us to solve for a balanced growth path at the industry level. This is also related to the endogenous growth theories (e.g., see Aghion and Howitt 1998), though the focus there is to study how R&D could drive balanced growth at the aggregate economy level.

22

Given the profit function $\pi(d_t; \lambda_t)$ represented by Eq. (32), the dynamic optimization problem (33) can be explicitly solved for a balanced-growth path (see Appendix A3 for the detailed proof).

4.3 Calibration Results

4.3.1 Industry Background

Given the functional forms of the model, we choose parameter values to match U.S. payment card data from 1997-2008.[11] During this period, credit and debit cards became the most popular electronic payment means for retail transactions in the U.S. economy. The early history of general purpose credit cards in the U.S. can be traced back to 1950s when the Diners Club card was first launched, but it was not until the 1970s that credit cards eventually picked up, with the market eventually becoming highly concentrated (Visa and MasterCard became the two largest credit card networks, followed by Amex and Discover). The introduction of debit cards was more recent. ATM and ATM cards were invented in 1970s. By adding the debit function, some of the ATM cards became debit cards in early 1980s, which were used to pay at the point of sale. The adoption and usage of debit cards started to gain steam in the mid-1990s, and Visa, MasterCard and several ATM networks became dominant players.[12]

Since the late 1990s, with the wide adoption of credit cards and rapid expansion of debit cards, the fees associated with these cards have raised great controversies. The key issue is the asymmetric fees that card networks charge to merchants and consumers. In a card transaction, merchants are typically charged a much higher fee than consumers, with the major component being the interchange fee. Despite rapid technological progress in the card industry, the fees charged to merchants continue to increase while the fees

---

[11] Our sample starts from 1997 when the card fee data became available. The sample ends in 2008 because the adoption of credit and debit cards had almost saturated by then and we also want to avoid the disruption of the Great Recession.

[12] For the historical development of credit and debit cards in the U.S. market, see Evans and Schmalensee (2005), Hayashi, Sullivan and Weiner (2006), and Hayashi, Li and Wang (2017).

charged to consumers decline (and in some cases consumers pay a negative fee by receiving rewards). The controversy of payment card fees led to numerous lawsuits and Congress hearings, and eventually Congress mandated the Federal Reserve to regulate debit card interchange fees in 2011.

While our model is about electronic payments in general, credit and debit cards are particularly good applications. Note that while our model does not consider credit functionality, it is still very relevant for credit cards. In the late 1990s, about 78% of U.S. consumers adopted credit cards, but among them about 50% cardholders only used the payment function of the cards (i.e., by paying off all purchases made with the card at the end of each month, not utilizing any revolving credit), so-called "convenience users." However, a consumer's spending on a credit card is limited by the card's credit line, so credit cards may not cover the consumer's entire payment needs. This made way for the fast adoption of debit cards since the mid-1990s, especially among lower-income consumers who are constrained by their credit limits (Zinman, 2009). Because debit cards do not provide the credit function, they serve purely as a payment device. In 2008, credit card were used in 26.5 billion transactions worth about $2.1 trillion in the U.S. economy, while debit cards were used in 34 billion transactions worth about $1.3 trillion.

4.3.2 Model Fit

To calibrate the model, we first collect data on the pricing, adoption and usage for credit and debit cards. The pricing data we collect include the interchange fees charged by Visa, MasterCard and major debit card networks between 1997-2008 and we convert them into merchant card fees by adding a reasonable merchant acquiring fee. We also collect data on consumers' adoption of credit and debit cards in the same period, as well as annual transaction values for cash, credit card and debit card.[13] We then calibrate our model to

---

[13]Data sources: Interchange fee data are from the *American Banker* (various issues). We then calculate the average credit card interchange fee based on credit card data for Visa and MasterCard. We also calculate the value weighted average debit card interchange fee based on debit card data for Visa, MasterCard, and top PIN debit networks. We then add a merchant acquiring fee on top of the interchange fee so that

match these patterns. Our calibrated model also delivers additional patterns, including the decline in consumer card fees and card operation costs, which are consistent with industry observations.

Table 1. Parameter Values for Model Calibration

Parameter Definition		Value	Parameter Definition		Value
Merchant cost of handing cash	τ_m	4.0%	Initial value of card service costs	d_0	2.25%
Consumer cost of handing cash	τ_c	2.5%	R&D function curvature	γ	0.5
Merchant cost of adopting card	k_m	2.5%	R&D efficiency parameter	ϕ	10
Consumer cost of adopting card	k_c	0.3%	Initial value of mean income	$1/\lambda_0$	21,215
Merchant cost of goods	μ_α	1	Growth rate of mean income	g_I	2%

Parameter values used for the calibration are reported in Table 1. We set merchant cost of handling cash $\tau_m = 4\%$, which gives the upper limit of the merchant card fee implied by the model. We set consumer cost of handling cash $\tau_c = 2.5\%$. We calibrate merchants' card adoption cost $k_m = 2.5\%$ of the consumer mean income, which implies an adoption cost about \$500 in 1997. Consumer card adoption cost is set as $k_c = 0.3\%$, which implies an adoption cost about \$60 in 1997. These numbers allow the model to match the card adoption and usage pattern in the data.[14] We set the mean consumer income to be \$21,215 in 1997 with an average growth rate 2%. This is derived from U.S. Census data on deflated annual Personal Consumption Expenditures.[15] Because merchant cost of

---

the total merchant discount rate for accepting credit cards is around 3% in the late 2000s, which is about the industry standard (e.g., Square Inc. charges merchants 2.75% for each swiped credit card transaction and 3.5% per manually-entered credit card transaction). The data on consumers' adoption of credit and debit cards are from the *Survey of Consumer Finance* (various years). The data on annual transaction values for cash, credit card and debit card are from the *Nilson Report* (various issues).

[14] Note that if we set symmetric costs for merchants and consumers $\tau_c = \tau_m$ and $k_m = k_c$, the model would deliver the same qualitative pattern of card pricing that the merchant card fee continues to increase over time while consumer card fee continues to decrease.

[15] In our model, a consumer's income equals her total purchases. Therefore, we use U.S. Census data on deflated annual Personal Consumption Expenditure (PCE) from 1997-2008 to calibrate the consumer income/purchase in our model. Using the estimates from the *Nilson Report*, we count purchases of goods and services as 78% of the PCE (the rest are nonpurchases – transactions for which no payment was made, such as food and lodging received by employees).
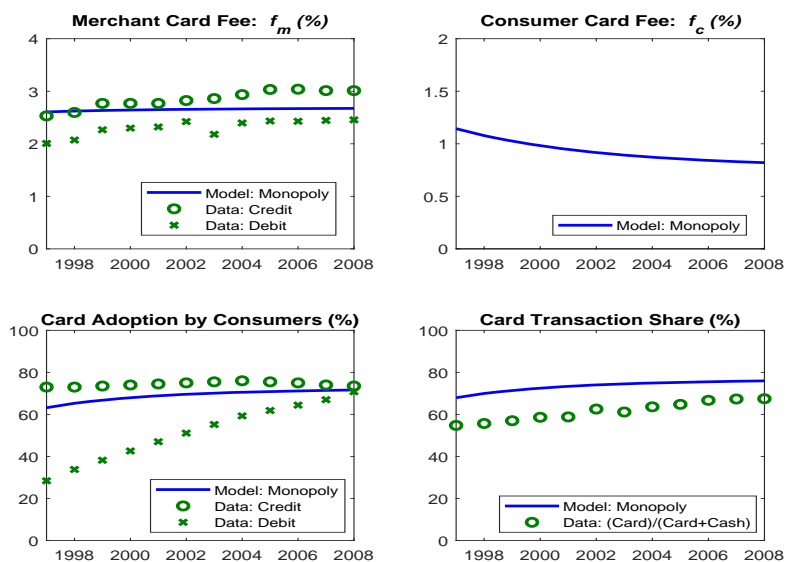
goods μ_α plays no role in our analysis except for scaling utility, we normalize that $\mu_\alpha = 1$. We set initial value of card service cost $d_0 = 2.25\%$, below the costs of handing cash. We set R&D function parameters $\gamma = 0.5$ and $\phi = 10$ to match the dynamic pattern of data.

The model calibration yields patterns that are consistent with data, as shown in Fig. 5A. The model generates a rising merchant fee over time and the level falls into the range between the average merchant fees charged by credit and debit cards in the data. The model generates a declining consumer fee over time, which is consistent with an increase of consumer card rewards during the period.[16] The model also generates increasing adoption of cards by consumers similar to the data, and the share of card transaction rises in parallel with the data (for the sum of credit and debit cards) though at a slightly higher level.

Beyond comparison with the data, Fig. 5B reports additional patterns of interest generated by the model. First, the overall card adoption by merchants increases over time, with the shares of large merchants (i.e., those who accept both cash and card but charge prices lower than cash-only stores) and small merchants (who accept cash only) declining. Second, the card network invests in R&D to reduce card service costs d, with the R&D expenditure to mean income ratio declining over time. Third, with the decline in card costs, the card network charges an increasing markup, which together with a rising card spending share indicates that the card network earns an increasing profit. Finally, consumer welfare in each period, scaled by the mean income, continues to rise, while in a cash economy it would be constant (which we normalize to 100 in the figure). The detailed derivation of consumer welfare is explained next.

---

[16]It is hard to estimate the average consumer card fees. Credit card users, if not rolling over any balances on their cards, may not need to pay a fee (or even receive rewards) for each card transaction. However, there is a chance that some of those users may end up borrowing from their cards, in which case they then need to pay a very high interest rate for every transaction made. For debit cards, consumers often need to pay a PIN fee for using online debit cards in our sample period, but some of them may also receive rewards. On the other hand, industry studies show that card rewards have gained increasing popularity over time. In 2001, less than a quarter of credit card offers included the promise of a rewards program. But by 2005, the share was 58 percent, according to Mail Monitor, a unit of consumer research company Synovate. A similar trend also happened to debit cards, as shown in 2005/2006 Study of Consumer Payment Preferences conducted by the American Bankers Association and Dove Consulting.
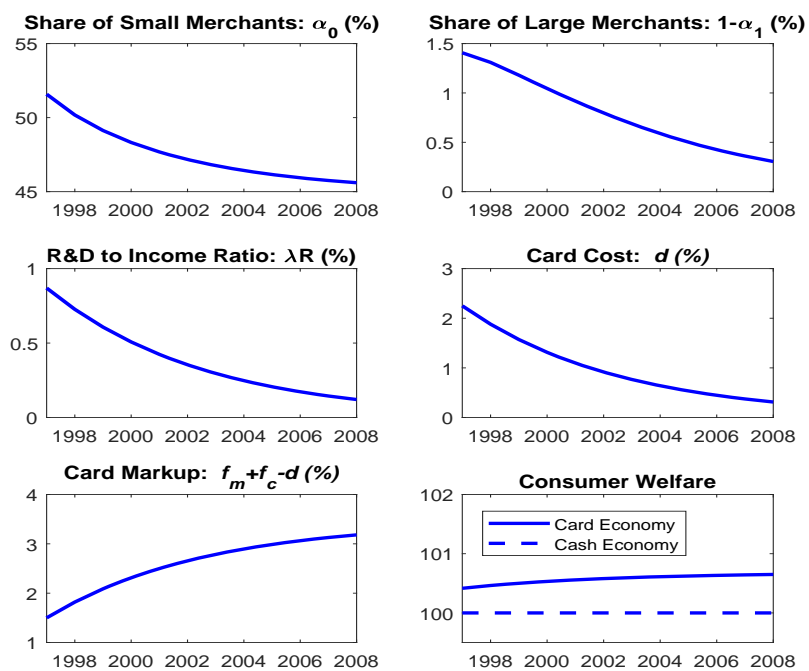
A.

Merchant Card Fee: f_m (%)

Model: Monopoly
Data: Credit
Data: Debit

Consumer Card Fee: f_c (%)

Model: Monopoly

Card Adoption by Consumers (%)

Model: Monopoly
Data: Credit
Data: Debit

Card Transaction Share (%)

Model: Monopoly
Data: (Card)/(Card+Cash)

B.

Share of Small Merchants: α_0 (%)

Share of Large Merchants: $1-\alpha_1$ (%)

R&D to Income Ratio: λR (%)

Card Cost: d (%)

Card Markup: $f_m + f_c - d$ (%)

Consumer Welfare

Card Economy
Cash Economy

Fig. 5. Model Calibration.

27

5 Welfare and Policy Analysis

This section provides a normative analysis of the card payment system. First, we show introducing payment cards increases welfare for both card adopters and nonadopters. Second, we show a Ramsey social planner who maximizes consumer and social welfare using the same set of pricing and R&D instruments would behave differently than a profit-maximizing platform. Finally, we discuss welfare implications of policy interventions.

5.1 Payment Card and Welfare Improvement

In our model framework, introducing the payment card improves every consumer's welfare, even for those who do not adopt the card. This can be shown in the following welfare comparison between a cash economy and a card economy.

Recall in a cash economy, an individual consumer I enjoys the utility level $U_{I,h}$ such that

$$\ln U_{I,h} = \int_0^{\overline{\alpha}} \frac{\alpha}{E(\alpha)} \ln \frac{\alpha I}{E(\alpha)(1+\tau_c)p_{\alpha,h}} dG(\alpha).$$

Consider the same economy but with the introduction of the payment card. A consumer decides whether to adopt the card based on her income. For a consumer who adopts the card (i.e., $I \geq I_0$), her utility $U_{I,d}^d$ implies that

$$\ln U_{I,d}^d = \int_0^{\alpha_0} \frac{\alpha}{E(\alpha)} \ln \frac{\alpha(I - K_c)}{E(\alpha)(1+\tau_c)p_{\alpha,h}} dG(\alpha) + \int_{\alpha_0}^{\overline{\alpha}} \frac{\alpha}{E(\alpha)} \ln \frac{\alpha(I - K_c)}{E(\alpha)(1+f_c)p_{\alpha,d}} dG(\alpha),$$

while a nonadopter (i.e. $I < I_0$) derives the utility level $U_{I,d}^h$ such that

$$\ln U_{I,d}^h = \int_0^{\min(\alpha_1,\overline{\alpha})} \frac{\alpha}{E(\alpha)} \ln \frac{\alpha I}{(1+\tau_c)p_{\alpha,h}E(\alpha)} dG(\alpha) + \int_{\min(\alpha_1,\overline{\alpha})}^{\overline{\alpha}} \frac{\alpha}{E(\alpha)} \ln \frac{\alpha I}{(1+\tau_c)p_{\alpha,d}E(\alpha)} dG(\alpha).$$

Therefore, a consumer receives different welfare gain depending on her income. For a card

28

adopter, her welfare gain is

$$
\begin{aligned}
\ln U_{I,d}^d - \ln U_{I,h} &= \int_{\alpha_0}^{\min(\alpha_1,\overline{\alpha})} \frac{\alpha}{E(\alpha)} \ln(\frac{p_{\alpha,h}}{p_{\alpha,d}}) dG(\alpha) + \int_{\min(\alpha_1,\overline{\alpha})}^{\overline{\alpha}} \frac{\alpha}{E(\alpha)} \ln(\frac{p_{\alpha,h}}{p_{\alpha,d}}) dG(\alpha) \\
&\quad + \frac{E_{\alpha \geqslant \alpha_0}(\alpha)}{E(\alpha)} \ln(\frac{1+\tau_c}{1+f_c}) + \ln(\frac{I - K_c}{I}),
\end{aligned}
\tag{34}
$$

while for a nonadopter, her welfare gain is

$$
\ln U_{I,d}^h - \ln U_{I,h} = \int_{\min(\alpha_1,\overline{\alpha})}^{\overline{\alpha}} \frac{\alpha}{E(\alpha)} \ln(\frac{p_{\alpha,h}}{p_{\alpha,d}}) dG(\alpha).
\tag{35}
$$

Equations (34) and (35) are intuitive: A card consumer enjoys utility gains from shopping at card-accepting stores in both categories (1) and (2), subject to the card adoption and usage costs, while a cash consumer only benefits from lower retail prices charged by merchants in category (1). In fact, we can derive the following proposition.

Proposition 2 *No consumer is worse off given any positive measure of card adoption.*

Proof. Equation (35) shows that cash consumers' welfare cannot be lower in a card economy than in a cash economy, given that $p_{\alpha,d} \leq p_{\alpha,h}$ in merchant category (1). Equation (34) shows this also holds true for the marginal card consumers $I = I_0$, while all other card consumers $I > I_0$ are strictly better off in the card economy. ∎

The findings of Proposition 2 stand in contrast to the previous studies (e.g., Rochet and Tirole 2002, 2011), which argue that cash consumers subsidize card consumers. In our setting, because cards are a more efficient payment means, cash consumers are indeed subsidized by card consumers. However, this does not mean that cash consumers' welfare cannot be further improved in a card economy, as we will study next.

5.2   Ramsey Social Planner

Given the utility measures described above, the Ramsey social planner would choose a sequence of card fees $(f_{c,t}, f_{m,t})$ and R&D investment R_t to maximize the present value

of consumer welfare subject to the adoption incentive constraints of merchants and consumers as well as the network balanced-budget constraint each period. Because the Ramsey social planner makes zero profit for running the card network, it also maximizes the social welfare under those constraints. The Ramsey social planner's problem is analyzed as follows.

Within-Period Decision Within each period, for given values of card costs d and R&D investment R, the Ramsey social planner who runs the card network would choose (f_c, f_m) to maximize consumer welfare with the revenue covering operation and R&D expenditures. This means that

$$U(d, R; E(I), K_m, K) = \underset{f_c, f_m}{Max} \int_{I_0}^{\infty} U_{I,d}^d dG(I) \ + \int_0^{I_0} U_{I,d}^h dG(I) \tag{36}$$

$$s.t. \ Eqs \ (11), (17), (19), (20), (21), (22), (34), (35), \ \text{and}$$

$$\pi(d; E(I), K_m, K_c) = \frac{E_{\alpha \geq \alpha_0}(\alpha) E_{I \geq I_0}(I - K_c)}{E(\alpha)(1 + f_c)}(f_c + f_m - d) \geq R. \tag{37}$$

To minimize distortion, the Ramsey social planner would always keep the constraint (37) binding.

Dynamic Decision To describe the dynamic problem, we add back the time subscript to each variable. The value function of the Ramsey social planner is as follows.

$$V(d_t; E(I_t), K_{m,t}, K_{c,t}) \ = \ \max_{R_t} U(d_t, R_t; E(I_t), K_{m,t}, K_{c,t}) \tag{38}$$
$$+ \beta V(d_{t+1}; E(I_{t+1}), K_{m,t+1}, K_{c,t+1})$$

$$s.t. \ d_{t+1} = \Gamma(d_t, R_t), \tag{39}$$

where Γ is the R&D function.

Based on the same set of functional form assumptions and parameter values as above, we can solve the Ramsey social planner's problem and compare the result with that of the profit-maximizing network (see Appendix A4 for the detailed proof).

For illustrative purposes, we consider a hypothetical scenario where the Ramsey social planner takes over and runs the card network at the beginning of our sample period. Figures 6A-6B compare the Ramsey outcome with the monopoly network case and show several important findings. First, both merchants and consumers enjoy lower fees under the Ramsey social planner, and unlike the monopoly case, both the merchant fee and the consumer fee decrease over time. These lead to higher card adoption and usage, which suggests that the market power of card networks does contribute to a slow diffusion of electronic payments. Second, under the Ramsey social planner, the share of large merchants whose sizes are large enough to attract both card and cash consumers is much higher and increasing over time. This suggests that cash users would benefit much more from electronic payments than under the monopoly network case. Third, the Ramsey social planner would conduct more R&D, which results in a faster decline in card service cost. Finally, all things together, consumer welfare in each period increases faster under the Ramsey social planner, which is due to the fact that the Ramsey social planner charges a decreasing markup over time, while the monopoly charges an increasing one.

A few reasons explain why the monopoly network and the Ramsey social planner would behave differently. For the monopoly network, imposing a high merchant fee leads to high retail prices of goods and allows the network to extract more rents produced by replacing costly cash payments. Moreover, because the network does not earn profits from cash users, charging a high merchant fee reduces cross subsidies from card users to cash users through large merchants who serve both card and cash users. In contrast, the Ramsey social planner cares about consumers' real purchases rather than their nominal card spending, and cares about the welfare of both card and cash users. In terms of R&D

31

decisions, the monopoly only sees the benefit of increased profit, which is a subset of the social welfare that the social planner would value, so the monopoly makes less R&D investment than the Ramsey social planner.
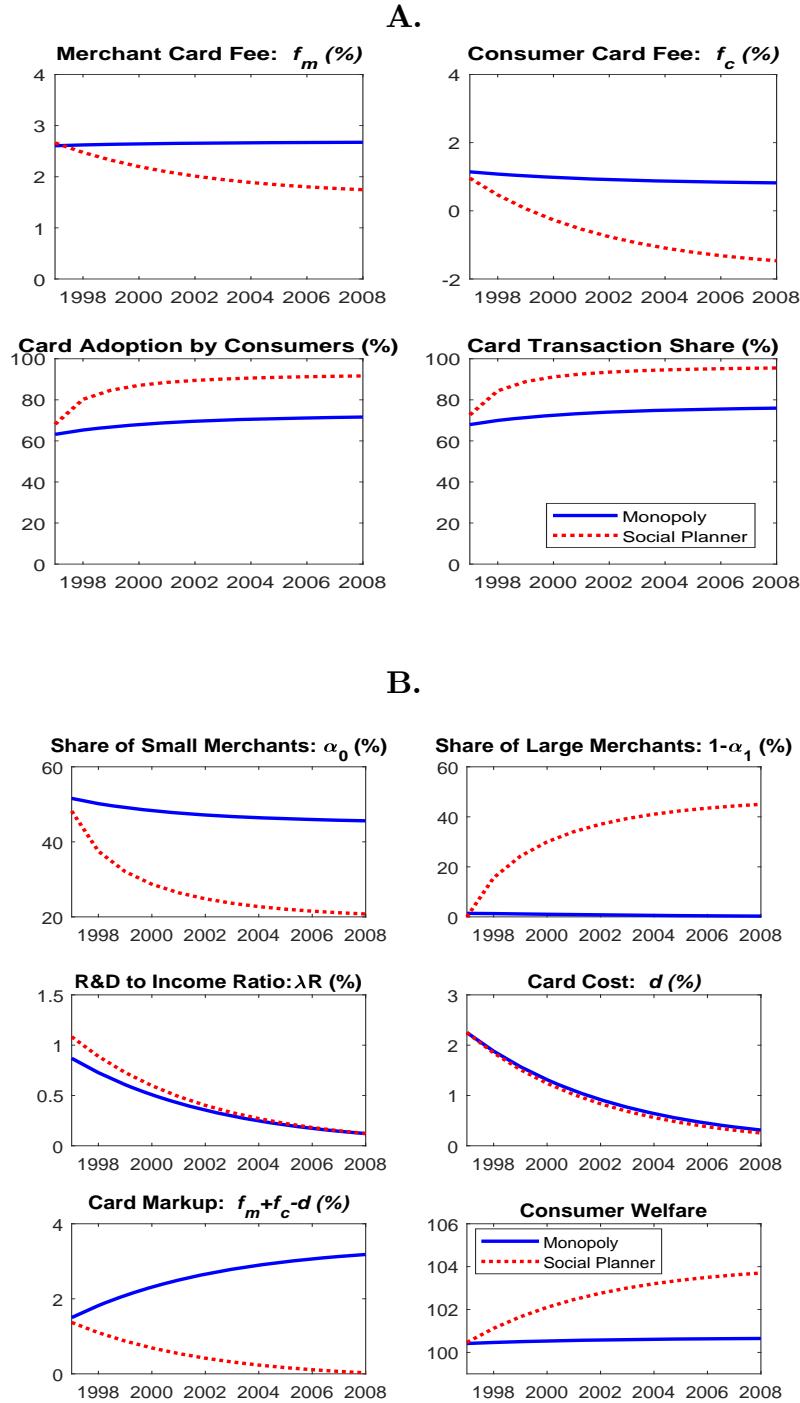
A.

Merchant Card Fee: f_m (%)

Consumer Card Fee: f_c (%)

Card Adoption by Consumers (%)

Card Transaction Share (%)

— Monopoly
····· Social Planner

B.

Share of Small Merchants: α_0 (%)

Share of Large Merchants: $1-\alpha_1$ (%)

R&D to Income Ratio: λR (%)

Card Cost: d (%)

Card Markup: f_m+f_c-d (%)

Consumer Welfare

— Monopoly
····· Social Planner

Fig. 6. Monopoly Network vs. Ramsey Social Planner

Note that in the Ramsey social planner case, because the network earns zero profit, consumer welfare equals social welfare. However, in the monopoly network case, evaluating social welfare needs to take into account both consumer welfare and network profits by assigning appropriate weights to each. Here, we consider a natural approach by assuming that the monopoly network rebates its profits to consumers proportionally to consumers' income (e.g., consumers own shares of the network in proportion to their income). In so doing, we can assign consumption values to the network profits to calculate the total social welfare. Of course, there are other ways of assigning welfare weights to profits, and we keep the distinction between "consumer welfare" (without profit rebates) and "social welfare" (with profit rebates) in our following analysis.
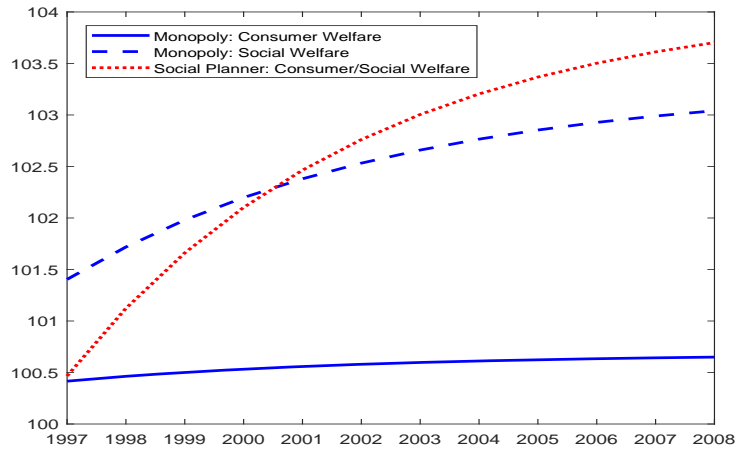
Fig. 7. Social Welfare Comparison.

Based on our calibrated model, consumers' income with profit rebates continues to follows an exponential distribution but just has a higher mean. Because the card adoption costs are proportional to the mean income, card adoption thresholds would remain unchanged for merchants and consumers. Figure 7 compares social welfare under the monopoly network versus the Ramsey social planner. The results show that in the beginning years, because the social planner invests more in R&D, social welfare in a given period is actually lower than that under the monopoly, even though consumer welfare (without profit rebates) is higher. However, social welfare in each period grows faster

under the Ramsey social planner and surpasses that under the monopoly in a few years and the gap becomes wider over time.

5.3 Policy Experiments

In this section, we use our calibrated model to study two regulatory approaches on payment cards. One is the marginal-cost pricing regulation, and the other is the merchant fee cap regulation. These two approaches are the most popular ones in policy debates, and can each be justified by some existing theories.

The marginal-cost pricing regulation can find its root in traditional one-sided markets, and a naive argument is to require the card network to set fees to merchants and consumers equal to the marginal cost of serving each side, which implies $f_m = d_m$ and $f_c = d_c$. However, as Baxter (1983) pointed out, because payment card markets are two-sided, it would be decidedly inefficient to block side payments between merchants and consumers. Instead, the socially optimal card pricing should be $f_m + f_c = d_m + d_c$. While this approach considers the two-sided nature of card markets, it is based on a static analysis and ignores the endogenous R&D of card networks.

On the other hand, the merchant fee cap regulation has been adopted in many countries. Compared with the marginal-cost pricing regulation, it is easier to implement because it regulates only the fee on the merchant side. In practice, the merchant (interchange) fee cap is rationalized either by an issuer-cost argument (e.g., in Australia and the U.S.) or by a merchant-benefit argument (e.g., in the EU).

We simulate our calibrated model for each policy experiment by assuming the regulation is implemented at the beginning year of our sample period. For the marginal-cost pricing regulation, we require the card network to set card fees $(f_{m,t}, f_{c,t})$ to maximize consumer welfare subject to the zero markup constraint $f_{m,t} + f_{c,t} = d_t$. Figures 8A-8B show the results. Compared with the unregulated monopoly case, the regulated network lowers card fees to both merchants and consumers. This boosts card adoption by both

sides and the fraction of large merchants who serve both card and cash users increases substantially. All these result in a higher level of consumer welfare. However, this regulation deprives the card network of R&D incentive and resources, so the card service cost d_t stays at the initial level d_0. Figure 9 shows that compared with the unregulated monopoly and the Ramsey social planner, this regulation yields a lower level of social welfare in each period except for the initial few years.

We also simulate alternative regulations that cap the merchant fee at various levels. Figures 8A-8B show the results for the cap set at 1%. Given the binding cap on the merchant fee, the card network turns to a higher consumer fee to make up the lost revenue. As a result, consumer card adoption is lower but the adoption by merchants (including large merchants who serve both card and cash users) becomes higher compared with the unregulated monopoly case. The card transaction share does not change much but the markup is lower, so the network profit is reduced. As a result, the card network's R&D spending is constrained in the first few years, which leads to a slower decline in card service costs until the late stage of the sample period.[17] Consumer welfare in each period is higher under the merchant fee cap regulation compared with the unregulated monopoly case, but as shown in Fig. 9, social welfare gets lower in the early years before it becomes higher in the longer run.

Figure 10 provides welfare comparison between different scenarios by computing the present values of consumer and social welfare at the beginning year of our sample period, with the present value of the cash economy being normalized to 100. The marginal-cost pricing regulation maximizes consumer welfare in a static setting, but it leaves no profit for the card network to conduct R&D. As a result, it yields a higher present value of consumer welfare than the unregulated monopoly, but the present value of social welfare

---

[17]Our calibrated model shows that the merchant fee cap regulation reduces the card network's profit level but not the marginal return of doing R&D. Recall the monopoly network's profit function (32): $\pi(d; \lambda) = \frac{1}{\lambda}(a_0 - a_1 d)$. The merchant fee cap reduces the value of a_0 but not the value of a_1 (and even increases it slightly), and a_1 determines the rate of cost decline at the balanced growth path. Therefore, the decline in card service cost d_t does not slow down except for the first few years when the network's optimal R&D spending is constrained by its period profit.

is lower. In comparison, the merchant fee cap regulation redistributes between network profit and consumer welfare but without hurting the network's R&D much. We do see that the lower the cap, the higher the consumer welfare, but the social welfare (which slightly increases in the cap value) changes little compared with the unregulated monopoly.

A.
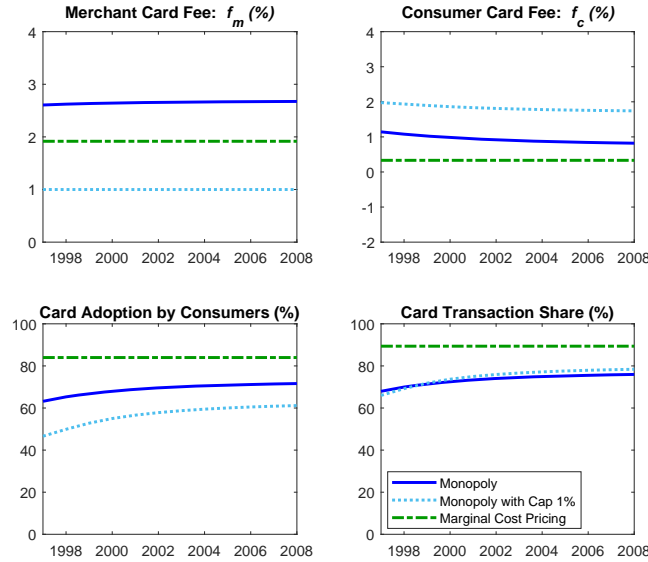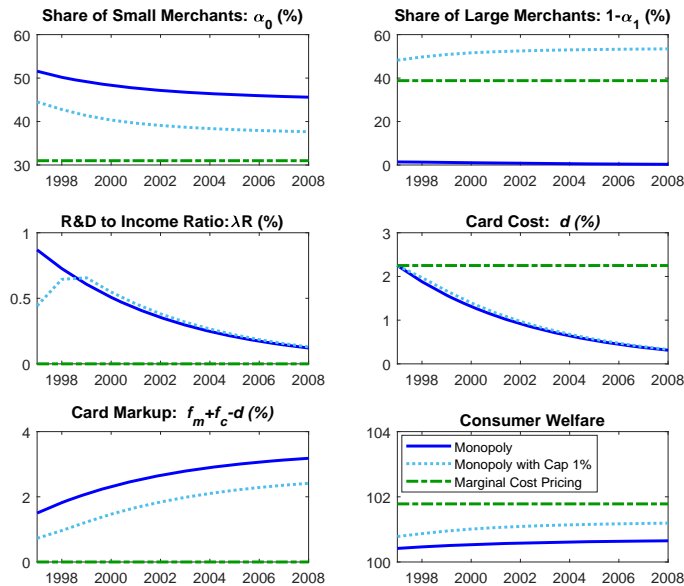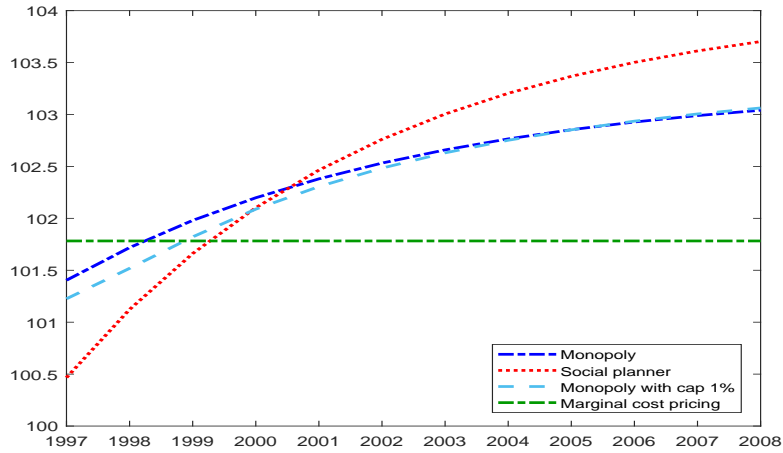

B.

Fig. 8. Policy Experiments
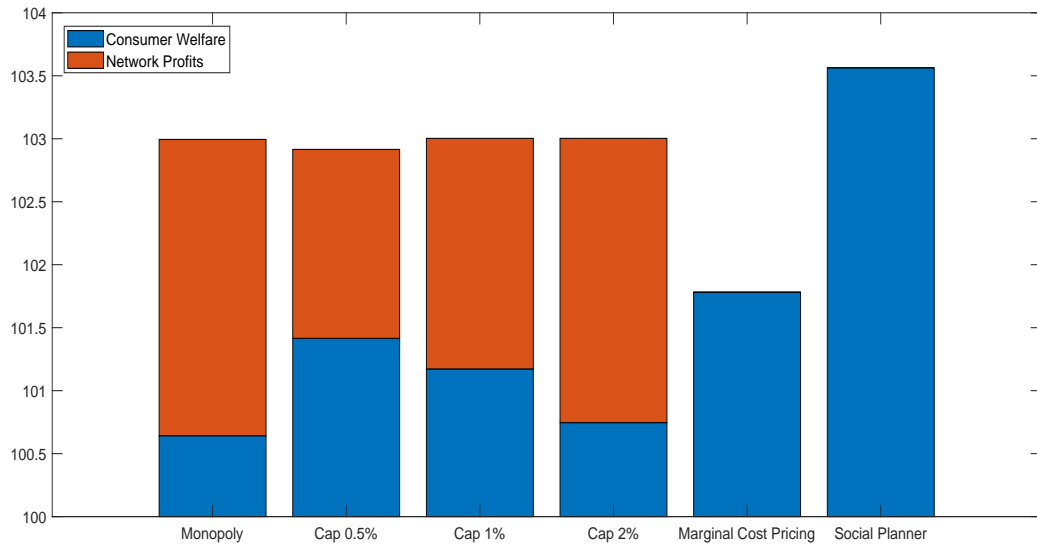
Fig. 9. Social Welfare Comparison of Policy Experiments.

Fig. 10. Present Value Comparison of Consumer and Social Welfare.

6   Conclusion

In this paper, we provide a new analysis of payments system evolution and the accompa-
nying competitive efficiency issues. Our theory lays out a two-sided market environment
where consumers and merchants make adoption and usage decisions for electronic pay-
ments. The economics of these payment choices, specifically the fixed costs of adoption

and marginal benefits of usage intertwined with the heterogeneity of consumer income and merchant size, together with price coherence, generate nontrivial network externalities. This setting is embedded in a fully dynamic model, in which a monopoly electronic payment network sets usage fees and conducts R&D to lower costs. We calibrate the model to U.S. payment card pricing, adoption and usage data, and conduct welfare and policy analysis.

Our analysis suggest that the market power of electronic payment networks plays important roles in explaining the slow adoption of electronic payments. In contrast, a Ramsey social planner who aims to maximize consumer and social welfare would set lower usage fees and conduct more R&D, thus achieving higher levels of adoption and usage of electronic payments. We also consider several regulatory interventions. Our findings suggest that while marginal-cost pricing regulation could maximize consumer welfare in a static setting, it shuts off the card network's R&D and may reduce social welfare in a dynamic setting. In comparison, the merchant fee cap regulation may improve consumer welfare without causing much dynamic inefficiency.

Our analysis can be extended in several directions. First, our model assumes that the fixed costs of adopting electronic payments move together with the consumer mean income. This is a simplifying assumption that allows our analysis to focus on cost-reducing technological progress. In case data becomes available, it would be useful to study how the adoption costs actually change over time and how that affects diffusion of electronic payments. Second, by assuming consumer income follows an exponential distribution, our analysis focuses on changes in the average income while holding the income inequality fixed. For future research, it might also be interesting to study the effect of income inequality, perhaps by choosing more flexible distributions. Third, the competitive merchants that we consider provides a first approximation to modeling retail market competition. It would be interesting to study alternative merchant market structures and their effects on payment system evolution. Finally, one may extend the framework to consider multiple

payments networks. It would be interesting to study how cooperation and competition among networks affect the market outcome.

References

[1] Aghion, P., Howitt, P., 1998. Endogenous Growth Theory. MIT Press, Cambridge, MA.

[2] Baxter, W., 1983. Bank interchange of transactional paper: Legal perspectives. Journal of Law and Economics 26, 541-588.

[3] Dragulescu, A., Yakovenko, V.M., 2001. Evidence for the exponential distribution of income in the USA. European Physical Journal B 20, 585-589.

[4] Evans, D., Schmalensee, R., 2005. Paying with Plastic: The Digital Revolution in Buying and Borrowing, 2nd Edition. MIT Press, Cambridge, MA.

[5] Fudenberg, D., Tirole, J., 1991. Game Theory. MIT Press, Cambridge, MA.

[6] Hall, R., 2001. The stock market and capital accumulation. American Economic Review 91(5), 1185-1202.

[7] Hall, R., 2004. Measuring factor adjustment costs. Quarterly Journal of Economics 119(3), 899-927.

[8] Hayashi, F., Sullivan, R., Weiner, S., 2006. A Guide to the ATM and Debit Card Industry: 2006 Update. Federal Reserve Bank of Kansas City.

[9] Hayashi, F., Li, B., Wang, Z., 2017. Innovation, deregulation, and the life cycle of a financial service industry. Review of Economic Dynamics 26, 180–203.

[10] Kahn, C., Roberds, W., 2009. Why pay? An introduction to payments economics. Journal of Financial Intermediation 18, 1-23.

[11] Khan, A., Thomas, J., 2008. Adjustment costs. In: The New Palgrave Dictionary of Economics. Palgrave Macmillan, London.

[12] Kiyotaki, N., Wright, R., 1989. On money as a medium of exchange. Journal of Political Economy 97, 927-954.

[13] Lagos, R., Wright, R., 2005. A unified framework for monetary theory and policy analysis. Journal of Political Economy 113, 463-484.

[14] McAndrews, J., Wang, Z., 2012. The economics of two-sided payment card markets: Pricing, adoption and usage. Federal Reserve Bank of Richmond Working Paper No. 12-06.

[15] Mester, L., 2012. Changes in the use of electronic means of payment: 1995-2010. Federal Reserve Bank of Philadelphia Business Review, Third Quarter, 25-36.

[16] Nosal, E., Rocheteau, G., 2011. Money, Payments, and Liquidity, MIT Press, Cambridge, MA.

[17] Rochet, J., Tirole, J., 2002. Cooperation among competitors: Some economics of payment card associations. RAND Journal of Economics 33, 549-570.

[18] Rochet, J., Tirole, J., 2011. Must-take cards: Merchant discounts and avoided costs. Journal of the European Economic Association 9(3), 462–495.

[19] Rysman, M., Wright, J., 2014. The economics of payment cards. Review of Network Economics 13, 303-353.

[20] Shy, O., Wang, Z., 2011. Why do payment card networks charge proportional fees? American Economic Review 101(4), 1575-1590.

[21] Wang, Z., Wright, J., 2017. Ad valorem platform fees, indirect taxes, and efficient price discrimination. RAND Journal of Economics 48, 467–84.

[22] Wright, J., 2003. Optimal card payment systems. European Economic Review 47, 587-612.

[23] Zinman, J., 2009. Debit or credit? Journal of Banking and Finance 33(2), 358-366.

Appendix

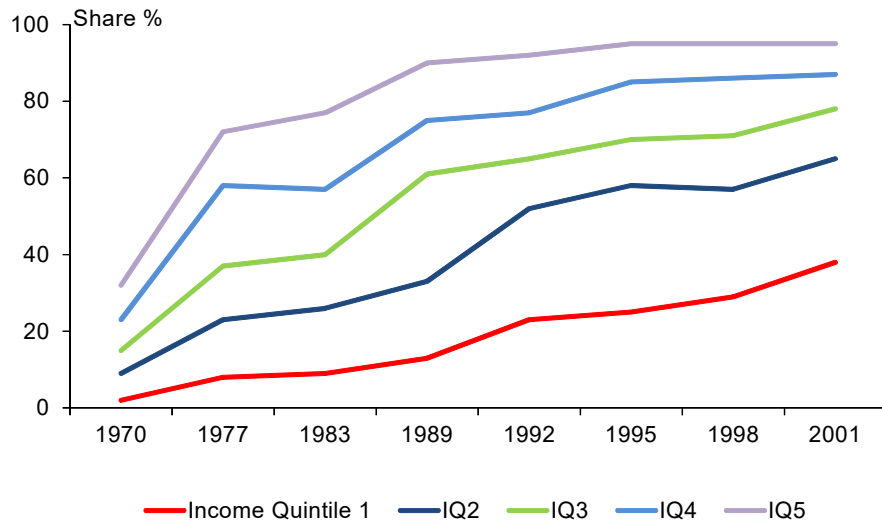## A1. Card Adoption by Consumer Income and Merchant Size[18]

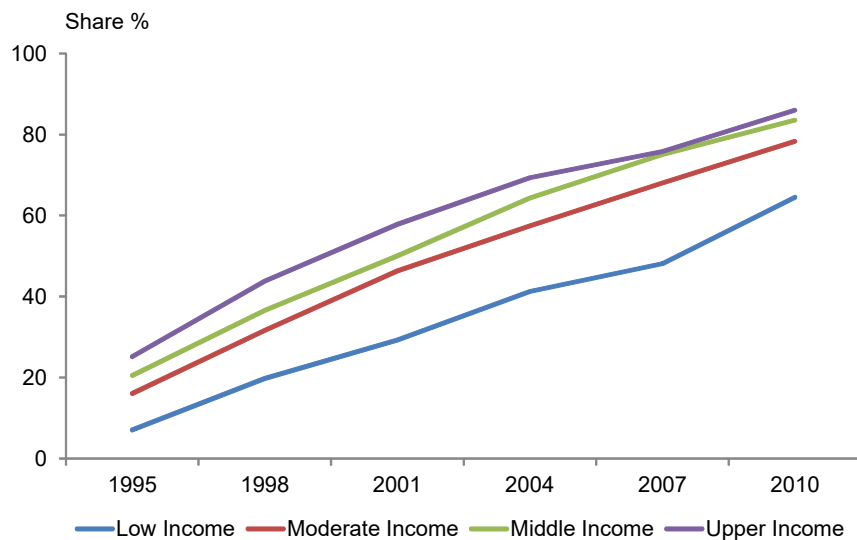Fig. A1(a). Percentage of U.S. Households Holding Credit Cards

Fig. A1(b). Percentage of U.S. Households Using Debit Cards
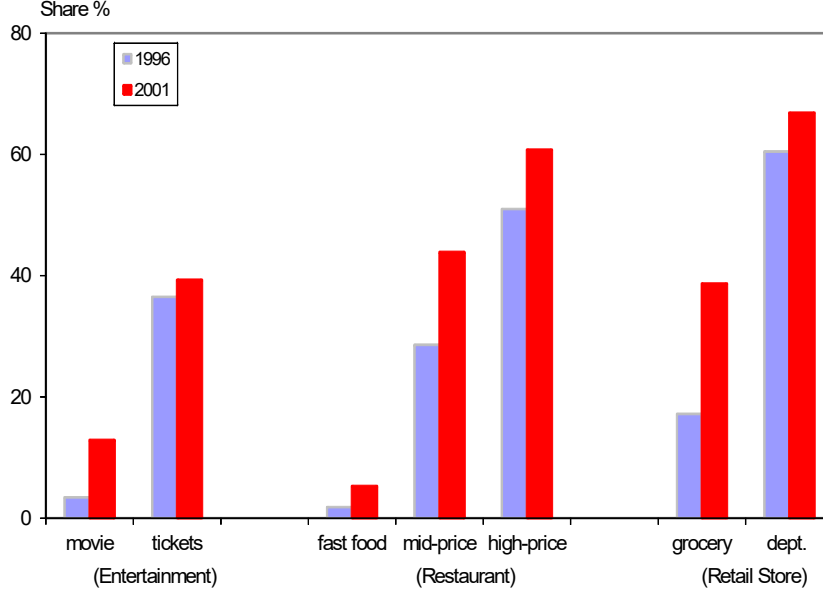
Fig. A1(c). Percent of Transactions Using Payment Cards

A2. Derivation of Eqs. (L1) and (L2).

Under Assumptions F1 and F3, we can rewrite Eqs. (20) and (22) in the main paper into

$$\alpha_0 = \frac{k_m}{2e^{(-\lambda I_0)}(1 + \lambda I_0 - k_c)Z_0}, \tag{40}$$

$$\lambda I_0 = \frac{(\frac{1+\tau_c}{1+f_c})^{(1-\alpha_0^2)}k_c}{(\frac{1+\tau_c}{1+f_c})^{(1-\alpha_0^2)} - \exp(2\int_{\alpha_0}^{\min(\alpha_1,1)} \alpha \ln(\frac{(1-\tau_m)\alpha}{(1-f_m)\alpha-(1+f_c)\alpha_0 Z_0})d\alpha)}. \tag{41}$$

Equation (40) is exactly Eq. (L1). We can further rewrite Eq. (41) by noting that

$$
\begin{aligned}
&\int \alpha \ln\left(\frac{(1-\tau_m)\alpha}{(1-f_m)\alpha-(1+f_c)\alpha_0 Z_0}\right)d\alpha \\
= \ &\frac{\alpha^2}{2}\ln(1-\tau_m) - \frac{1}{2}\alpha^2\ln\left((1-f_m) - \frac{\alpha_0 Z_0(1+f_c)}{\alpha}\right) \\
&+ \frac{\alpha_0 Z_0(1+f_c)}{2(1-f_m)}\alpha + \frac{\alpha_0^2 Z_0^2(1+f_c)^2}{2(1-f_m)^2}\ln\left(\alpha - \frac{\alpha_0 Z_0(1+f_c)}{(1-f_m)}\right).
\end{aligned}
$$

42

Hence, we derive

$$2\int_{\alpha_0}^{\frac{Z_0}{Z_1}\alpha_0} \alpha \ln\Big(\frac{(1-\tau_m)\alpha}{(1-f_m)\alpha - (1+f_c)\alpha_0 Z_0}\Big)d\alpha$$
$$= \alpha_0^2 \left\{ \ln\frac{1+f_c}{1+\tau_c} + \frac{Z_0(1+f_c)}{(1-f_m)}\Big(\frac{Z_0}{Z_1}-1\Big) + \frac{Z_0^2(1+f_c)^2}{(1-f_m)^2}\ln\Big(\frac{\frac{Z_0}{Z_1}-\frac{Z_0(1+f_c)}{(1-f_m)}}{1-\frac{Z_0(1+f_c)}{(1-f_m)}}\Big) \right\}.$$

Equation (41) can then be rewritten into Eq. (L2) where

$$s = \ln\frac{1+f_c}{1+\tau_c} + \frac{Z_0(1+f_c)}{(1-f_m)}\Big(\frac{Z_0}{Z_1}-1\Big) + \frac{Z_0^2(1+f_c)^2}{(1-f_m)^2}\ln\Big(\frac{\frac{Z_0}{Z_1}-\frac{Z_0(1+f_c)}{(1-f_m)}}{1-\frac{Z_0(1+f_c)}{(1-f_m)}}\Big).$$

A3. Model Solution: Monopoly Network

Given functional form assumptions F1-F4, we can solve the model equilibrium under the monopoly network.

Within-Period Decision Within each period, the monopoly card network maximizes its profit as follows.

$$\pi(d;\lambda) = \max_{f_c,f_m} \frac{e^{(-\lambda I_0)}}{\lambda}(1+\lambda I_0 - k_c)\Big(\frac{1-\alpha_0^2}{1+f_c}\Big)(f_c + f_m - d) \tag{42}$$

$$s.t. \quad \alpha_0 = \frac{k_m}{2e^{(-\lambda I_0)}(1+\lambda I_0 - k_c)Z_0}, \tag{43}$$

$$\alpha_1 = \frac{Z_0}{Z_1}\alpha_0 \quad \text{if} \quad f_m \leq \tau_m, \tag{44}$$

$$\lambda I_0 = \frac{(\frac{1+\tau_c}{1+f_c})^{(1-\alpha_0^2)}k_c}{(\frac{1+\tau_c}{1+f_c})^{(1-\alpha_0^2)} - \exp(2\int_{\alpha_0}^{\min(\alpha_1,1)}\alpha\ln(\frac{(1-\tau_m)\alpha}{(1-f_m)\alpha - (1+f_c)\alpha_0 Z_0})d\alpha)}, \tag{45}$$

$$\tau_c \geq f_c, \quad \frac{1-f_m}{1+f_c} \geq \frac{1-\tau_m}{1+\tau_c}. \tag{46}$$

43

For a given marginal cost d and mean income $1/\lambda$, we can derive f_c^*, f_m^*, λI_0^*, α_0^*. The corresponding maximum profit is

$$\pi(d;\lambda) = \frac{e^{(-\lambda I_0^*)}}{\lambda}(1 + \lambda I_0^* - k_c)(\frac{1 - \alpha_0^{*2}}{1 + f_c^*})(f_c^* + f_m^* - d). \tag{47}$$

Conditioning on d, Eqs. (43-45) imply that the optimal values $(f_m^*, f_c^*, \lambda I_0^*, \alpha_0^*)$ are invariant to λ. Therefore, according to Eq. (47), the profit function is proportional to the mean income $\frac{1}{\lambda}$. Moreover, we can verify that $\lambda\pi$ is approximately linear in d (as shown in Fig. A2 based on the model calibration), so that

$$\pi(d;\lambda) = \frac{1}{\lambda}(a_0 - a_1 d),$$

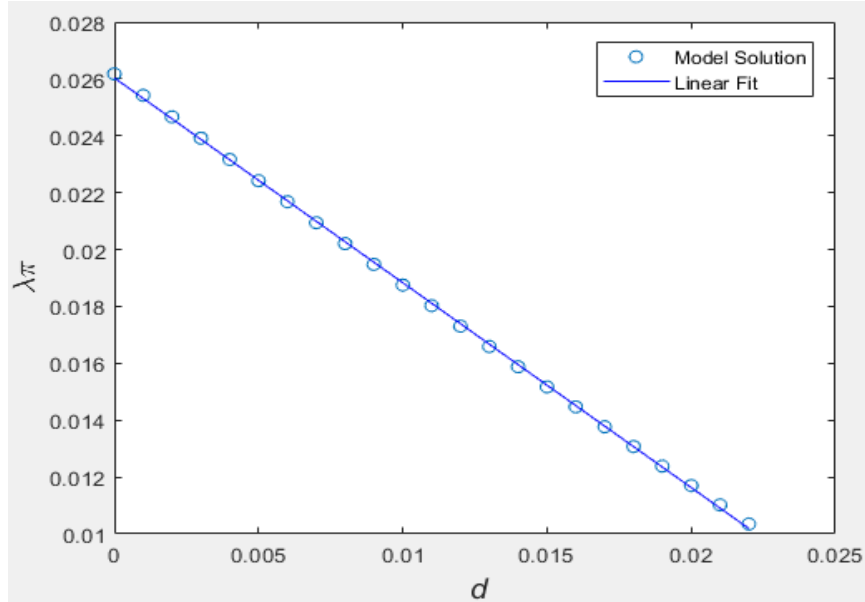where a_0 and a_1 are scalars determined by the model parameter values.

Fig. A2. Network Profit Function and Linear Fitting

Dynamic Decision   We now add back the time subscript to each variable. The value function of the monopoly network is as follows.

$$V(d_t;\lambda_t) = \max_{R_t} \pi(d_t;\lambda_t) - R_t + \beta V(d_{t+1};\lambda_{t+1}), \tag{48}$$

$$s.t. \quad \frac{1}{d_{t+1}} - \frac{1}{d_t} = (\frac{R_t \lambda_t}{\phi})^\gamma d_t^{-\gamma-1}, \tag{49}$$

$$\pi(d_t; \lambda_t) \geq R_t. \tag{50}$$

Assuming the budget constraint (50) never binds, the dynamic problem can be rewritten as

$$V(d_t; \lambda_t) = \max_{d_{t+1}} \frac{1}{\lambda_t}(a_0 - a_1 d_t) - \phi \frac{d_t}{\lambda_t} \left[\frac{d_t}{d_{t+1}} - 1\right]^{\frac{1}{\gamma}} + \beta V(d_{t+1}; \lambda_{t+1}).$$

The first-order condition implies that

$$\frac{\phi}{\gamma \lambda_t} \left[\frac{d_t}{d_{t+1}} - 1\right]^{\frac{1}{\gamma}-1} \left[\frac{d_t}{d_{t+1}}\right]^2 + \beta V'(d_{t+1}; \lambda_{t+1}) = 0. \tag{51}$$

The envelope condition implies that

$$V'(d_t; \lambda_t) = -\frac{a_1}{\lambda_t} - \frac{\phi}{\lambda_t} \left[\frac{d_t}{d_{t+1}} - 1\right]^{\frac{1}{\gamma}} - \frac{\phi}{\gamma \lambda_t} \left[\frac{d_t}{d_{t+1}} - 1\right]^{\frac{1}{\gamma}-1} \frac{d_t}{d_{t+1}},$$

which suggests that

$$V'_{t+1}(d_{t+1}; \lambda_{t+1}) = -\frac{a_1}{\lambda_{t+1}} - \frac{\phi}{\lambda_{t+1}} \left[\frac{d_{t+1}}{d_{t+2}} - 1\right]^{\frac{1}{\gamma}} - \frac{\phi}{\gamma \lambda_{t+1}} \left[\frac{d_{t+1}}{d_{t+2}} - 1\right]^{\frac{1}{\gamma}-1} \frac{d_{t+1}}{d_{t+2}}. \tag{52}$$

We can combine Eqs. (51) and (52) to get

$$\frac{\phi}{\gamma \lambda_t} \left[\frac{d_t}{d_{t+1}} - 1\right]^{\frac{1}{\gamma}-1} \left[\frac{d_t}{d_{t+1}}\right]^2 = \frac{\beta a_1}{\lambda_{t+1}} + \frac{\beta \phi}{\lambda_{t+1}} \left[\frac{d_{t+1}}{d_{t+2}} - 1\right]^{\frac{1}{\gamma}} + \frac{\beta \phi}{\gamma \lambda_{t+1}} \left[\frac{d_{t+1}}{d_{t+2}} - 1\right]^{\frac{1}{\gamma}-1} \frac{d_{t+1}}{d_{t+2}}.$$

Calibrating $\gamma = 1/2$, we can rewrite the above equation as

$$\frac{2\phi}{\lambda_t} \left[\frac{d_t}{d_{t+1}} - 1\right] \left[\frac{d_t}{d_{t+1}}\right]^2 = \frac{\beta a_1}{\lambda_{t+1}} + \frac{\beta \phi}{\lambda_{t+1}} \left[\frac{d_{t+1}}{d_{t+2}} - 1\right]^2 + \frac{2\beta \phi}{\lambda_{t+1}} \left[\frac{d_{t+1}}{d_{t+2}} - 1\right] \frac{d_{t+1}}{d_{t+2}}. \tag{53}$$

Given the constant income growth rate $\frac{\lambda_t}{\lambda_{t+1}} = g_\lambda$, we can solve for the balance growth

path where $\frac{d_t}{d_{t+1}} = \frac{d_{t+1}}{d_{t+2}} = g_d$. Note that we can rewrite Eq. (53) as

$$\frac{2\phi}{g_\lambda}\left[g_d - 1\right]g_d^2 = \beta a_1 + \beta\phi\left[g_d - 1\right]^2 + 2\beta\phi\left[g_d - 1\right]g_d,$$

and solve for a constant cost declining rate g_d. Based on g_d, we can then solve dynamic paths of all the endogenous variables in the model. We can also verify that the budget constraint (50) never binds.

A4. Model Solution: Ramsey Social Planner

Given functional form assumptions F1-F4, we can solve the model equilibrium under the Ramsey social planner.

Within-Period Decision

Within each period, for a given income-scaled R&D expenditure λR, the Ramsey social planner maximizes consumer welfare as follows.

$$U(d, \lambda R; \lambda) = \underset{f_c, f_m}{Max} \int_{I_0}^{\infty} U_{I,d}^d \lambda e^{-\lambda I} dI \ \ + \int_{0}^{I_0} U_{I,d}^h \lambda e^{-\lambda I} dI$$

$$s.t. \ \ U_{I,d}^d = U_{I,h} \exp(\ln U_{I,d}^d - \ln U_{I,h}) \tag{54}$$

$$= I\overline{U_h} \exp\left\{ \begin{array}{c} \int_{\alpha_0}^{\min(\alpha_1,1)} 2\alpha \ln(\frac{p_{\alpha,h}}{p_{\alpha,d}})d\alpha + \int_{\min(\alpha_1,1)}^{1} 2\alpha \ln(\frac{p_{\alpha,h}}{p_{\alpha,d}})d\alpha \\ +(1 - \alpha_0^2)\ln(\frac{1+\tau_c}{1+f_c}) + \ln(\frac{\lambda I - k_c}{\lambda I}) \end{array} \right\},$$

$$U_{I,d}^h = U_{I,h}\exp(\ln U_{I,d}^h - \ln U_{I,h}) = I\overline{U_h}\exp\left\{ \int_{\min(\alpha_1,1)}^{1} 2\alpha \ln(\frac{p_{\alpha,h}}{p_{\alpha,d}})d\alpha \right\}, \tag{55}$$

$$(\frac{p_{\alpha,h}}{p_{\alpha,d}})_{1\geq\alpha\geq\min(\alpha_1,1)} = \frac{\frac{1-f_m}{1+f_c}e^{-\lambda I_0}(1 + \lambda I_0 - k_c) + \frac{1-\tau_m}{1+\tau_c}[1 - e^{-\lambda I_0}(1 + \lambda I_0)] - \frac{k_m}{2\alpha}}{\frac{1-\tau_m}{1+f_c}e^{-\lambda I_0}(1 + \lambda I_0 - k_c) + \frac{1-\tau_m}{(1+\tau_c)}[1 - e^{-\lambda I_0}(1 + \lambda I_0)]},$$

$$\left(\frac{p_{\alpha,h}}{p_{\alpha,d}}\right)_{\min(\alpha_1,1)>\alpha\geqslant\alpha_0} = \frac{(\frac{1-f_m}{1+f_c})e^{-\lambda I_0}(1+\lambda I_0 - k_c) - \frac{k_m}{2\alpha}}{\frac{1-\tau_m}{1+f_c}e^{-\lambda I_0}(1+\lambda I_0 - k_c)},$$

$$\alpha_0 = \frac{k_m}{2e^{(-\lambda I_0)}(1+\lambda I_0 - k_c)Z_0},$$

$$\alpha_1 = \frac{Z_0}{Z_1}\alpha_0 \quad \text{if} \quad f_m \leq \tau_m,$$

$$\lambda I_0 = \frac{(\frac{1+\tau_c}{1+f_c})^{1-\alpha_0^2}k_c}{(\frac{1+\tau_c}{1+f_c})^{1-\alpha_0^2} - \exp(2\int_{\alpha_0}^{\min(\alpha_1,1)}\alpha\ln(\frac{(1-\tau_m)\alpha}{(1-f_m)\alpha-(1+f_c)\alpha_0 Z_0})d\alpha)},$$

$$\tau_c \geq f_c, \quad \frac{1-f_m}{1+f_c} \geq \frac{1-\tau_m}{1+\tau_c},$$

$$\lambda\pi = e^{(-\lambda I_0)}(1+\lambda I_0 - k_c)(\frac{1-\alpha_0^2}{1+f_c})(f_c + f_m - d) \geq \lambda R.$$

For a given marginal cost d and the profit target λR, we can derive f_c^*, f_m^*, λI_0^*, α_0^*. The corresponding maximum consumer welfare is

$$U(d, \lambda R; \lambda) = \int_{I_0}^{\infty} U_{I,d}^d \lambda e^{-\lambda I} dI \quad + \int_0^{I_0} U_{I,d}^h \lambda e^{-\lambda I} dI \tag{56}$$
$$= \frac{\overline{U_h}}{\lambda}\left\{\int_{\lambda I_0}^{\infty}\lambda I\exp\{\cdot\}e^{-\lambda I}d\lambda I \quad + \exp\{*\}\int_0^{\lambda I_0}\lambda I e^{-\lambda I}d\lambda I\right\},$$

where $\exp\{\cdot\}$ and $\exp\{*\}$ are given in Eqs. (54) and (55), respectively.

We can again verify that conditioning on d and λR, the optimal values $(f_c^*, f_m^*, \lambda I_0^*, \alpha_0^*)$ are invariant to λ, so according to Eq. (56) the welfare function is proportional to the mean income $\frac{1}{\lambda}$. Moreover, we can verify that λU is approximately linear in d and λR (as shown in Fig. A3 based on the model calibration). Therefore, the welfare function

can be simplified as

$$U(d, \lambda R; \lambda) = \frac{1}{\lambda}(b_0 - b_1 d - b_2 \lambda R),$$

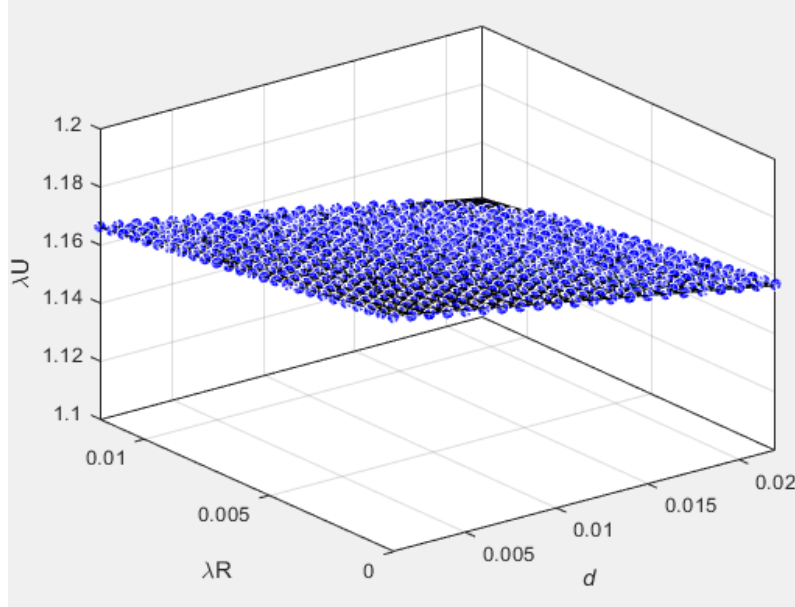where b_0, b_1 and b_2 are scalars determined by the model parameter values.

Fig. A3. Consumer Welfare Function and Linear Fitting

Dynamic Decision We now add back the time subscript to each variable. The value function of the Ramsey social planner is as follows.

$$V(d_t; \lambda_t) = \max_{R_t} U(d_t, \lambda_t R_t; \lambda_t) + \beta V(d_{t+1}; \lambda_{t+1}),$$

$$s.t. \ \frac{1}{d_{t+1}} - \frac{1}{d_t} = (\frac{R_t \lambda_t}{\phi})^\gamma d_t^{-\gamma-1}.$$

The dynamic problem can be rewritten as

$$V(d_t; \lambda_t) = \max_{d_{t+1}} \frac{1}{\lambda_t}(b_0 - b_1 d_t - b_2 \lambda_t R_t) - b_2 \phi \frac{d_t}{\lambda_t}\left[\frac{d_t}{d_{t+1}} - 1\right]^{\frac{1}{\gamma}} + \beta V(d_{t+1}; \lambda_{t+1}).$$

The first-order condition implies that

$$b_2 \phi \frac{d_t^2}{\gamma \lambda_t d_{t+1}^2}\left[\frac{d_t}{d_{t+1}} - 1\right]^{\frac{1}{\gamma}-1} + \beta V'(d_{t+1}; \lambda_{t+1}) = 0. \qquad (57)$$

48

The envelope condition implies that

$$V'(d_t; \lambda_t) = \frac{1}{\lambda_t}(-b_1) - \frac{b_2\phi}{\lambda_t}\left[\frac{d_t}{d_{t+1}} - 1\right]^{\frac{1}{\gamma}} - b_2\phi\frac{d_t}{d_{t+1}\gamma\lambda_t}\left[\frac{d_t}{d_{t+1}} - 1\right]^{\frac{1}{\gamma}-1},$$

which suggests that

$$V'(d_{t+1}; \lambda_{t+1}) = \frac{1}{\lambda_{t+1}}(-b_1) - \frac{b_2\phi}{\lambda_{t+1}}\left[\frac{d_{t+1}}{d_{t+2}} - 1\right]^{\frac{1}{\gamma}} - b_2\phi\frac{d_{t+1}}{d_{t+2}\gamma\lambda_{t+1}}\left[\frac{d_{t+1}}{d_{t+2}} - 1\right]^{\frac{1}{\gamma}-1}. \quad (58)$$

Combining Eqs. (57) and (58), we get

$$b_2\phi\frac{d_t^2}{\gamma\lambda_t d_{t+1}^2}\left[\frac{d_t}{d_{t+1}} - 1\right]^{\frac{1}{\gamma}-1} = \frac{\beta}{\lambda_{t+1}}b_1 + \frac{\beta b_2\phi}{\lambda_{t+1}}\left[\frac{d_{t+1}}{d_{t+2}} - 1\right]^{\frac{1}{\gamma}} + b_2\phi\frac{\beta d_{t+1}}{d_{t+2}\gamma\lambda_{t+1}}\left[\frac{d_{t+1}}{d_{t+2}} - 1\right]^{\frac{1}{\gamma}-1}.$$

Calibrating $\gamma = 1/2$, we can rewrite the above equation as

$$2b_2\phi\frac{d_t^2}{\lambda_t d_{t+1}^2}\left[\frac{d_t}{d_{t+1}} - 1\right] = \frac{\beta}{\lambda_{t+1}}b_1 + \frac{\beta b_2\phi}{\lambda_{t+1}}\left[\frac{d_{t+1}}{d_{t+2}} - 1\right]^2 + 2b_2\phi\frac{\beta d_{t+1}}{d_{t+2}\lambda_{t+1}}\left[\frac{d_{t+1}}{d_{t+2}} - 1\right]. \quad (59)$$

Given the constant income growth rate $\frac{\lambda_t}{\lambda_{t+1}} = g_\lambda$, we can solve for the balance growth path $\frac{d_t}{d_{t+1}} = \frac{d_{t+1}}{d_{t+2}} = g_d$. Note that Eq. (59) can be rewritten as

$$\frac{2\phi b_2}{g_\lambda}[g_d - 1]g_d^2 = \beta b_1 + \beta\phi b_2[g_d - 1]^2 + 2b_2\beta\phi[g_d - 1]g_d,$$

and we can solve for the constant cost declining rate g_d. Based on g_d, we can then solve dynamic paths of all the endogenous variables in the model.