ISSN 1936-5330

THE FEDERAL RESERVE BANK of KANSAS CITY
ECONOMIC RESEARCH DEPARTMENT

The Economics of Payment Card Fee Structure: What is the Optimal Balance Between Merchant Fee and Payment Card Rewards?

Fumiko Hayashi

November 2008
RWP 08-06

RESEARCH WORKING PAPERS

THE ECONOMICS OF PAYMENT CARD FEE STRUCTURE: WHAT IS THE OPTIMAL BALANCE BETWEEN MERCHANT FEE AND PAYMENT CARD REWARDS?

Fumiko Hayashi[1]

November 2008

RWP 08-06

Abstract: This paper theoretically considers the optimal balance between the merchant fee and the cardholder fee (rewards) from both efficiency and equity perspectives. First, the paper constructs the models that can be used by the U.S. policymakers. Because theoretical results are very sensitive to the assumptions of the models, it is important to construct models that reflect the reality of the market. Second, the most efficient fee structure and product price are considered under the various combinations of the assumptions. And finally, the paper considers welfare consequences of the most efficient fee structure.

Keywords: Payment card rewards, optimal fee structure, efficiency, equity

JEL Classification: D61, D63

[1] Payments System Research Function, Economic Research Department, Federal Reserve Bank of Kansas City. E-mail: fumiko.hayashi@kc.frb.org. The views expressed in this article are those of the author and do not necessarily reflect those of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

1. Introduction

As card payments have become more prevalent, the fee structures of payment cards have been attracting more policy debates and public policy interventions. In many countries, public policy interventions have reduced the level of fees paid by merchants and as a result either payment card rewards received by card users have been reduced or fees paid by card users have been raised. In contrast to the trend in many other countries, the U.S. public authorities and legislature have not taken actions until very recently and both the merchant fees and payment card rewards have continued to increase. In 2008, the U.S. legislature has introduced two bills in Congress, which are aiming to change the balance between the merchant fee and the cardholder fee (or rewards).

Although policymakers may be urged to take actions to resolve the heating policy debates between the merchants and the card networks, before taking actions they should consider three key questions, regarding payment card fee structures. First, what is the optimal balance between the merchant fee and cardholder fee (or payment card rewards)? Second, if the market cannot reach the optimal balance, what market forces cause the equilibrium fee structure to deviate from the optimal fee structure? And third, what are policy options? This paper is the first of a series of three papers, each of which addresses each of the three questions above.

Knowing the optimal balance between the merchant fee and cardholder fee is important for policymakers. Policymakers cannot evaluate the current fee structure determined at the market place—whether the current fee structure is close enough to the optimal fee structure—without knowing the optimal fee structure. The optimal fee structure can also be the target of policies that policymakers would potentially implement.

There are two commonly used criteria to consider the "optimal" fee structures or price levels. One is efficiency and the other is equity. Efficiency is often measured by social welfare, which aggregates welfares of all parties involved in the market. The most efficient card fee structure, therefore, can be defined as the fee structure that maximizes social welfare of all parties involved in the payment card market. Equity considers the distribution of social welfare among different parties. In contrast to efficiency, there is no clear way to measure equity. Furthermore, sometimes equity and efficiency contradict: It is possible that the most efficient card fee structure results in a much skewed welfare distribution. With a given card fee structure, if only one entity or party absorbs all welfare gains and the rest incur welfare losses, then the fee structure has a problem from the equity point of view, even if the fee structure maximizes social welfare. Thus, political decision is required to define the "optimal" fee structure. This paper does not define the "optimal" fee structure; rather, the paper examines how the most efficient card fee structure affects welfare distribution among different parties.

The rest of the paper is organized as follows. Section 2 constructs theoretical models, taking into account institutional features of the U.S. payment card market. Section 3 examines the most efficient fee structures. Section 4 examines the welfare distribution under the most efficient fee structure. Section 5 concludes.

2. Models

This section constructs theoretical models that can be used by the U.S. policymakers when considering potential public policies. There is a large body of theoretical literature on payment card fee structure.[2] We have learned that the theoretical results are very sensitive to the

[2] See, for example, Chakravorti (2003).

assumptions of the models.[3] Therefore, constructing models that reflect the reality of the market is especially important for policymakers who implement actual policies.

There are several models that are used to analyze the most efficient fee structure. Among them, we use the model constructed by Rochet and Tirole (2002) (hereafter RT) as our base model, because the model reflects some of the institutional features in the U.S. market.[4] We extend their model, allowing some of the assumptions to vary. In the first half of this section, we consider key assumptions that reflect the U.S. payment card market. Then, in the second half, we construct our models.

2.1 Assumptions

RT model

The RT model assumes that the fee structure does not affect the number of merchants who accept the cards. In the model, the consumer's cardholding behavior and card usage behavior are not separated; however, one can interpret the model that all consumers hold a card and their usage decision is affected by the cardholder fee. Consumers ultimately determine which payment method to use: consumers use cards when their transactional benefit exceeds the cardholder fee. Transactional benefit of a card transaction is defined as the saving of opportunity cost of using the alternative payment method, such as cash. Consumers are assumed to be heterogeneous in terms of their transactional benefit from a card transaction, while merchants are assumed to be homogeneous. Merchants are required not to set different prices according to the payment methods their customers choose. In addition, their model (implicitly) assumes that i)

[3] Hayashi and Weiner (2006).
[4] RT model focuses on payment methods' payment function, and thus it does not consider the benefits/costs of credit card's credit function. This paper also focuses on credit card's payment function, rather than its credit function. In this sense, the model considers credit cards as "charge" cards.

consumers make a fixed number of transactions (i.e., they purchase a fixed number of goods regardless of the price) and ii) per transaction costs and fees for the payment methods are fixed regardless of the transaction value.

The U.S. payment card markets are matured

In the RT model, the number of merchants who accept the cards is fixed regardless of the merchant fee level. As mentioned, their model can be interpreted as if all consumers hold a card and their usage is affected by the cardholder fee. In this sense, the model assumes there is usage externality but no membership externality in the market.

Payment card markets potentially have two-types of externalities—membership externality and usage externality. Membership externality (or positive feedback) arises from membership decisions: A consumer's cardholding of a particular network's card depends on how many merchants accept that network's cards. Similarly, a merchant's card acceptance of a particular network's cards depends on how many current and potential customers use that network's cards. Usage externality arises from usage decision. In a payment market, consumers choose a payment method from a set of payment methods the merchant accepts. The consumer's choice of payment methods affects the merchant benefits/costs. However, the consumer's private incentive typically does not reflect the merchant benefits/costs.

Once the market matured, the positive feedback becomes almost negligible. That means, additional cardholders do not influence merchant card acceptance and additional merchants do not influence consumer cardholding. In contrast, usage externality exists, even after the market matured. Because the U.S. payment card markets can be considered to be matured, it may be appropriate to assume usage externality only.

Consumers

The RT model assumes that consumers are heterogeneous in terms of the transactional benefit from the card and use the card when their transactional benefit from the card exceeds the cardholder fee. A consumer's transactional benefit from cards may consist of three parts. One is resource cost saving by using a card as opposed to using an alternative payment method, such as cash or checks. For example, time spent at the casher may be much shorter if consumers use a card rather than checks. The second part is differences in benefits. Consumers may receive more benefits by using a card, such as record keeping, security, etc., but consumers may lose some benefits, such as anonymity or privacy. The last part is a saving from not paying fees to the banks or payment service providers for the alternative payment methods. How to value resource cost saving or difference in benefits may vary by individual consumers significantly. How the fees paid by consumers affect the consumers' choice of payment methods has not been fully uncovered. However, assuming the cardholder fee affects the consumer's card usage is more flexible than the other way. The model can treat it as a special case if the cardholder fee does not affect the consumer's payment choice.

In the RT model, consumers are assumed to make a fixed number of transactions. This assumption is likely to be true for some products and services but consumer demand for other products may increase as the product prices decrease or the rewards received by card users increase. The rewards received by card users affect the card users' effective price of product. A card-using consumer's effective price is the product price plus the cardholder fee (or minus rewards) minus the transactional benefit from cards, while a non-card-using consumer's effective price is the product price itself. Because there is little empirical evidence about which assumptions are more realistic, this paper considers both cases.

Merchants

Although some merchants are possibly monopolistic, many U.S. merchants are considered to be quite competitive. While monopolistic merchants are likely very sensitive to the merchant fee, competitive or oligopolistic merchants are less sensitive to the merchant fee, because of the strategic motives. Even when the merchant fees are quite high compared with the merchant transactional benefit from cards, competition among merchants may keep the merchants from rejecting the cards, given their rival merchants are accepting the cards.[5] Therefore, the assumption that merchants are less sensitive to the merchant fee may reflect the U.S. merchant behavior well.

A merchant's transactional benefit from cards is assumed to consist of two parts.[6] One is the resource cost saving by a card transaction as opposed to an alternative payment method transaction. The other part is saving from not paying fees to the banks or payment service providers for the alternative payment methods.

As the RT model assumes, this paper also assumes merchants are homogeneous in terms of the transactional benefit from cards and the cost of selling a unit of goods. Although merchants are quite different across industries, they are more homogeneous within an industry, in terms of transactional benefit, costs of selling and markup per unit of goods. Because in the United States, interchange fees (and thus merchant fees) are typically industry specific, this assumption can be justifiable within an industry.

Card networks

Payment card schemes take one of two principal organizational forms. One is four-party schemes: Four-parties are cardholders, merchants, card issuers, and merchant acquirers. Both

[5] Hayashi (2006).
[6] A merchant transactional benefit from a card may include another part. According to a report by Government Accountability Office (2008), accepting cards improves internal operations at merchants.

card issuers and merchant acquirers should be members of a payment card network. In a four-party scheme network, an interchange fee is set by the card network and paid by the merchant acquirer to the card issuer.[7] The merchant pays a merchant (discount) fee to the merchant acquirer. The merchant fee is set by each merchant acquirer, and a typical merchant acquirer entirely passes through the interchange fee to its merchants and charges other fees, such as an acquirer processing fee, association dues, and a switch fee. The cardholder either receives rewards or pays a cardholder fee to its card issuer. Each card issuer sets its own rewards or cardholder fees, and typically rewards are mostly financed by the issuer's interchange fee revenue.[8] The other organizational form is three-party schemes: Three parties are cardholders, merchants and a card network, such as American Express and Discover. In contrast to the four-party schemes, there is no interchange fee, because the card network acts as card issuer and merchant acquirer. The merchant pays a merchant fee to the card network and the cardholder either receives rewards or pays a cardholder fee to the network. The network sets both merchant fees and cardholder fees.

If transferring funds between the acquirer and the issuer (i.e., interchange fee) is allowed in a four-party scheme network, a three-party scheme network and a four-party scheme network are almost equivalent when social welfare is considered. In a social welfare function, a card network's profit is considered to be the joint profits of the card network and its member acquirers and issuers, in the case of four-party scheme network.

[7] In some networks, the interchange fee flows from the card issuer to the merchant acquirers. However, in the United States, the interchange fee flows from the merchant acquirer to the card issuer.
[8] The author obtained this information from industry experts at the "Consumer Behavior and Payment Choice Conference" held at the Federal Reserve Bank of Boston in July, 2006.

Nature of per transaction costs and fees

As the majority of models assume, the RT model assumes that per transaction costs, fees, and benefits for payment methods are fixed regardless of the transaction value (flat per transaction costs, fees, and benefits). Only a few models assume per transaction costs, fees, and benefits are proportional to the transaction value.[9] According to cost studies in the United States, some of the costs and fees are fixed and some are proportional to the transaction value. Typically, interchange fee structure (thus merchant fee structure) consists of a fixed portion and a proportional portion. For credit cards, a fixed portion is relatively small, while for debit cards, especially for PIN debit cards, the interchange fee is more or less a flat fee—the fee reaches its cap at the average transaction value. A bank's costs of processing a cash or a credit card transaction seem to be proportional to the transaction value, while their costs of processing a debit card or a check transaction seems not to be influenced by the transaction value. A merchant's resource costs seem to be proportional for credit cards and cash and to be flat for debit cards and checks. Some of the consumer's transactional benefits from payment methods, such as book keeping and anonymity might not be influenced by the transaction value, but other benefits might be proportional to transaction value. Thus, two extreme cases can be considered. In one case, all costs and fees (and benefits) per transaction are fixed, and in the other case, all of them are proportional to the transaction value. This paper considers both cases.

Merchant ability to set different prices according to the payment methods

Currently, many card networks have a rule that restricts merchants setting different prices for their customers according to the payment method they use. Although merchants are allowed to offer a cash discount for their customers, many of them do not do so. Therefore, the models assuming that merchants set the same price for both card-using and non-card-using consumers

[9] For example, McAndrews and Wang (2006).

are well aligned with the practice. However, abolishing this rule is often considered to be a viable public policy option. Therefore, this paper considers both cases.

2.2 Models

Base model

The base model is the following. The payment card markets are considered to be matured. All consumers hold at least one card and merchants accept cards as long as the merchant fees are lower than a certain threshold level, which is endogenously determined.

Consumers are heterogeneous in their transactional benefit from cards as opposed to the alternative payments. A consumer's transactional benefit from a card, b_B, consists of three parts. One is a gross benefit minus gross cost from using a card, B_B^C;[10] one is a gross benefit minus gross cost from using the alternative payment method, B_B^A; and one is the consumer fee paid for the alternative payment method, f^A. Thus, the transactional benefit from a card is defined as: $b_B = B_B^C - B_B^A + f^A$. b_B is assumed to be distributed over the interval $[\underline{b_B}, \overline{b_B}]$ with a density function of $h(b_B)$, and a cumulative distribution function of $H(b_B)$. Consumers pay the cardholder fee of f when they use a card.

Merchants are homogeneous (at least ex-ante) and their transactional benefit from cards, $\hat{b}_S$, is defined as the merchant cost for the alternative payment method, c_S^A, plus the merchant fee paid for the alternative payment method, m^A, minus the merchant cost for a card transaction, c_S^C (i.e., $\hat{b}_S = c_S^A + m^A - c_S^C$). Merchants pay the merchant fee of m when their customers use a card. Merchants also incur a cost of selling one unit of goods, d.

[10] Note that the fees for a card transaction are not included in the cost.

Assumptions that can be varied

The assumptions in the final models vary in terms of three categories: (i) Per transaction costs and fees; (ii) Consumer demand for goods; and (iii) Merchant ability to set different prices according to the payment method. There are two variations in each category. Per transaction costs and fees are either flat or proportional to the transaction value. Consumer demand for goods is either elastic (i.e., a consumer makes a fixed number of transactions) or downward-sloping (i.e., the number of transactions increases as the effective price of goods decreases). And finally, a merchant either sets the same price for all of its customers regardless of the payment method or sets the different prices according to the payment method its customers use.

3. The Most Efficient Fee Structure

The most efficient fee structure is defined as the fee structure that maximizes social welfare. Social welfare is defined as the aggregate surpluses of all parties involved in the payment card markets. Parties involved in the payment card markets are card-using consumers, non-card-using consumers, merchants, card networks, and payment service providers that provide the alternative payment method. Since payment cards are considered to be a substitute for the alternative payment methods, such as cash and checks, the surplus of a consumer who uses an alternative payment method and the surplus of alternative payment method service providers are also counted in social welfare. Payment cards can be provided by four-party scheme networks. Therefore, a card network's surplus can be interpreted as the joint surpluses of the card network and its member acquirers and issuers.

Policymakers should be interested in the most efficient fee structures for two reasons. First, the most efficient fee structure can be used to examine whether the current fee structure is

efficient or not. Second, the most efficient fee structure can be the target of policies that policymakers would potentially implement.

Existing theoretical models consider the most efficient fee structures in rather limited circumstances. Most models assume 1) consumers make a fixed number of transactions; 2) costs and fees per transaction do not vary by the transaction value (flat per transaction costs and fees); and 3) merchants are not allowed to set different product prices according to the payment methods. However, as discussed in the previous section, it is possible that a consumer's quantity demanded for goods depends on the effective price the consumer faces and that costs and fees per transaction are proportional to the transaction value. Potential policies may allow merchants to set discriminatory pricing. Thus, it is important to examine how the differences in these assumptions affect the most efficient fee structures.

In the first four subsections, we examine how consumer demand for goods and per transaction costs and fees affect the most efficient payment card fee structures by assuming merchants set the same price for both card-using and non-card-using consumers. In the last subsection, we relax this assumption and examine how merchant's ability to set different prices across payment methods affects the most efficient fee structure.

3.1 Scenario I: Fixed Number of Transactions and Flat Per-Transaction Costs and Fees

Assume that each consumer receives gross benefit, υ , by purchasing one unit of product. Since merchants set the same price for card-using consumers and non-card-using consumers, consumers whose transactional benefit exceeds the cardholder fee (i.e., $f \le b_B = B_B^C - B_B^A + f^A$) use a card and consumers whose transactional benefit is below the cardholder fee use an alternative payment method. The social welfare function is defined as:

$$SW = \int_{f}^{\bar{b}_B} (\upsilon - p - f + B_B^C) h(b_B) db_B + \int_{\underline{b}_B}^{f} (\upsilon - p - f^A + B_B^A) h(b_B) db_B$$
$$+ \int_{f}^{\bar{b}_B} (p - d - m - c_S^C) h(b_B) db_B + \int_{\underline{b}_B}^{f} (p - d - m^A - c_S^A) h(b_B) db_B \quad (1)$$
$$+ \int_{f}^{\bar{b}_B} (f + m - c) h(b_B) db_B + \int_{\underline{b}_B}^{f} (f^A + m^A - c^A) h(b_B) db_B$$

The first two terms of equation 1 are the surpluses of card-using consumers and non-card-using consumers, respectively, and the third and fourth terms are the merchants' profits from card-using consumers and from non-card-using consumers, respectively, and the last two terms are the card networks' profits and the (joint) profits of alternative payment method service providers.

This social welfare function is essentially the same as the social welfare function defined in RT. Equation 1 can be rewritten as:

$$SW = (\upsilon - d + B_B^A - c_S^A - c^A) + \int_{f}^{\bar{b}_B} \{b_B + \hat{b}_S - c - (f^A + m^A - c^A)\} h(b_B) db_B, \quad (1')$$

and the first term is fixed regardless of the payment card fee structure. The most efficient cardholder fee, f^*, satisfies the first-order condition:

$$f^* = \{c + (f^A + m^A - c^A)\} - \hat{b}_S . \quad (2)$$

This condition implies that the most efficient cardholder fee is the difference between the payment service providers' (card networks and alternative payment service providers) net costs of processing a card transaction and the merchant transactional benefit from a card. To the extent that allowing the alternative payment method generates profits or losses (i.e., $f^A + m^A - c^A > 0$ or $f^A + m^A - c^A < 0$) to maximize social welfare depends on the policymakers' objectives. To simplify the model, we assume that the alternative payment method generates zero profit hereafter. Note that if payment service providers of the alternative payment method jointly earn zero profit from the alternative payment method, then the most efficient cardholder fee is the

difference between the card networks' costs of processing a card transaction and the merchant transactional benefit from a card.

$$f^* = c - \hat{b}_S . \tag{2'}$$

We should note that neither the product price nor merchant fee affects social welfare although they are likely to affect the welfare distribution among different parties.

3.2 Scenario II: Fixed Number of Transactions and Proportional Per-Transaction Costs and Fees

The social welfare function is the same as equation 1 except for the costs, fees, and benefits for payment methods are proportional to the price of the goods.[11]

$$\begin{aligned} SW = & \int_f^{\bar{b}_B} (\upsilon - p - fp + B_B^C p) h(b_B) db_B + \int_{\underline{b}_B}^{f} (\upsilon - p - f^A p + B_B^A p) h(b_B) db_B \\ & + \int_f^{\bar{b}_B} (p - d - mp - c_S^C p) h(b_B) db_B + \int_{\underline{b}_B}^{f} (p - d - m^A p - c_S^A p) h(b_B) db_B \\ & + \int_f^{\bar{b}_B} p(f + m - c) h(b_B) db_B . \end{aligned} \tag{3}$$

The first-order condition with respect to the cardholder fee gives the same condition as equation 2'. Thus, the most efficient cardholder fee is the same regardless of whether per transaction costs and fees are flat or proportional to the transaction value when consumers make a fixed number of purchases and transactions.

However, the product price affects the social welfare when costs and fees per transaction are proportional to the transaction value. The first-order condition with respect to the product price is:

$$\frac{\partial SW}{\partial p} = \int_{f^*}^{\bar{b}_B} (B_B^C - c_S^C - c) h(b_B) db_B - \int_{\underline{b}_B}^{f^*} (B_B^A - c_S^A - c^A) h(b_B) db_B , \tag{4}$$

[11] Note that we assume that payment service providers of the alternative payment method jointly earn zero profit from the alternative payment method.

This condition implies that if aggregate surpluses from using a card (the first term) exceed aggregate surpluses from using the alternative payment method (the second term), then the social welfare increases as the product price increases. The highest product price that satisfies all parties' incentive compatibility constraints is the one that makes the marginal card user's surplus from using a card zero.[12] That is $p = \upsilon/(1+ f - B_B^C(i_M))$, where $B_B^C(i_M)$ is the marginal card user i_M's gross benefit minus gross cost of using a card. If equation 4 is negative, the social welfare increases as the product price decreases. The lowest price can be achieved when merchants set the product price at the marginal cost and card networks set the merchant fee so that they earn zero profit from cards. The lowest price $p = \frac{d}{1-\hat{b}_S - c_S^C}$ can be achieved when the card networks set the merchant fee at $m = \hat{b}_S$.

The summary of the most efficient fee structure and product price is the following:

When $\int_{f^*}^{\bar{b}_B}(B_B^C - c_S^C - c)h(b_B)db_B - \int_{\underline{b}_B}^{f^*}(B_B^A - c_S^A - c^A)h(b_B)db_B > 0$,

$f^* = c - \hat{b}_S$; $p^* = \upsilon/(1+ f^* - B_B^C(i_M))$; and the merchant fee does not affect social welfare;

When $\int_{f^*}^{\bar{b}_B}(B_B^C - c_S^C - c)h(b_B)db_B - \int_{\underline{b}_B}^{f^*}(B_B^A - c_S^A - c^A)h(b_B)db_B < 0$,

$f^* = c - \hat{b}_S$; $m^* = \hat{b}_S$; $p^* = \frac{d}{1-\hat{b}_S - c_S^C}$; and

When $\int_{f^*}^{\bar{b}_B}(B_B^C - c_S^C - c)h(b_B)db_B - \int_{\underline{b}_B}^{f^*}(B_B^A - c_S^A - c^A)h(b_B)db_B = 0$,

$f^* = c - \hat{b}_S$, and the merchant fee and product price do not affect social welfare.

[12] Marginal card user is defined as the consumer who is indifferent between using a card and the alternative payment method.

3.3 Scenario III: Downward-sloping Consumer Demands and Flat Per-Transaction Costs and Fees

There are several variations when the market demand curve is downward sloping. Even though each consumer purchases a fixed number of products, if each consumer i has a different reservation utility, υ_i, then the number of consumers who purchase the product increases as the price of the goods decreases.[13] This section, however, assumes that each consumer has a downward-sloping demand curve for goods.

Assume that the alternative payment method (such as cash) is a "base" payment method for all consumers. That is, when product price is p, the quantity demanded by a consumer who uses the alternative payment method is $D(p) = a - bp$. The quantity demanded by a card user depends not only on the product price but also on the cardholder fee and the transactional benefit from cards. Two extreme cases can be considered. One case is where a consumer purchases goods on one transaction. The other case is where a consumer purchases one unit of goods per transaction.[14] In the former case, the effective price of a card-using consumer with transactional benefit of b_B is likely the same as that of a consumer who uses an alternative payment method; however, using a card effectively increases his income by $b_B - f$. In the latter case, the effective price of a card-using consumer with transactional benefit of b_B is $p + f - b_B$. Thus, his quantity demanded becomes $D(p) = a - b(p + f - b_B)$ when he uses a card. This paper focuses on the latter case: because in the former case, the (per transaction) fee structure of payment card and product price alone cannot maximize social welfare without violating some parties' incentive compatibility constraints.

[13] Schuh, Shy, and Stavins (2008).
[14] Schwartz and Vincent (2006).

The social welfare function is defined as: [15]

$$SW = \int_{f}^{\bar{b}_B} \frac{(a - bp - bf + bb_B)^2}{2b} h(b_B) db_B + \int_{\underline{b}_B}^{f} \frac{(a - bp)^2}{2b} h(b_B) db_B$$
$$+ \int_{f}^{\bar{b}_B} (p - d - m - c_S^C)(a - bp - bf + bb_B) h(b_B) db_B + \int_{\underline{b}_B}^{f} (p - d - m^A - c_S^A)(a - bp) h(b_B) db_B \quad (5)$$
$$+ \int_{f}^{\bar{b}_B} (f + m - c)(a - bp - bf + bb_B) h(b_B) db_B.$$

The first-order conditions with respect to the cardholder fee and with respect to the product price are, respectively:

$$\frac{\partial SW}{\partial f} = -b(p - d + f - c)(1 - H(f)) + (a - bp)(c - \hat{b}_S - f) h(f) = 0 \quad (6)$$

$$\frac{\partial SW}{\partial p} = b\{-p + d + \hat{b}_S - c_S^C + (c - \hat{b}_S - f)(1 - H(f))\} = 0 \quad (7)$$

Equation 7 implies that the product price that maximizes social welfare is:

$$p^* = d + \hat{b}_S - c_S^C + (c - \hat{b}_S - f)(1 - H(f)) \quad (7')$$

We obtain two f's from equations 6 and 7', but the second-order condition is satisfied only in one of the two, which is: $f^* = c - \hat{b}_S$.

Thus, the most efficient fee structure and product price are: $f^* = c - \hat{b}_S$ and $p^* = d + \hat{b}_S - c_S^C$. To achieve the most efficient product price without violating any parties' incentive compatibility constraints, the merchant fee needs to be set at $m^* = \hat{b}_S$. The most efficient cardholder fee is the same as the previous two scenarios and the most efficient product price can be achieved when merchants practice marginal cost pricing and card networks earn zero profit from the cards.

[15] Again, we assume that payment service providers of the alternative payment method jointly earn zero profit from the alternative payment method.

3.4 Scenario IV: Downward-sloping Consumer Demands and Proportional Per-Transaction Costs and Fees

Assume that each consumer has a downward-sloping demand curve. When per transaction costs and fees are proportional to the transaction value, whether the consumer purchases goods on one transaction or the consumer purchases one unit of goods per transaction does not affect the social welfare function. The social welfare function is defined as:[16]

$$SW = \int_{f}^{\bar{b}_B} \frac{\{a - bp(1 + f - b_B)\}^2}{2b} h(b_B) db_B + \int_{\underline{b}_B}^{f} \frac{(a - bp)^2}{2b} h(b_B) db_B$$
$$+ \int_{f}^{\bar{b}_B} \{p(1 - m - c_S^C) - d\}\{a - bp(1 + f - b_B)\} h(b_B) db_B + \int_{\underline{b}_B}^{f} \{p(1 - m^A - c_S^A) - d\}(a - bp) h(b_B) db_B$$
$$+ \int_{f}^{\bar{b}_B} p(m + f - c)\{a - bp(1 + f - b_B)\} h(b_B) db_B$$

The first best solution violates merchant and/or card network's incentive compatibility constraints: That means either merchants or card networks or both make losses at the first best solution. The second best solution is, therefore, to maximize consumer surplus subject to the incentive compatibility constraints of merchants and card networks. The problem is defined as:

$$\text{Max } CS = \int_{f}^{\bar{b}_B} \frac{\{a - bp(1 + f - b_B)\}^2}{2b} h(b_B) db_B + \int_{\underline{b}_B}^{f} \frac{(a - bp)^2}{2b} h(b_B) db_B$$

$$\text{s.t. } \int_{f}^{\bar{b}_B} \{p(1 - m - c_S^C) - d\}\{a - bp(1 + f - b_B)\} h(b_B) db_B$$
$$+ \int_{\underline{b}_B}^{f} \{p(1 - m^A - c_S^A) - d\}(a - bp) h(b_B) db_B = 0, \quad \text{and}$$
$$m + f - c = 0$$

[16] We assume that payment service providers of the alternative payment method jointly earn zero profit from the alternative payment method.

The incentive compatibility constraints imply that when $f = c - \hat{b}_S$, $p = \frac{d}{1 - \hat{b}_S - c_S^C}$. Although it is difficult to analytically solve the most efficient cardholder fee and product price, we are able to show providing rewards is unlikely the most efficient in this case.[17]

3.5 When Merchants Practice Discriminatory Pricing

The previous subsections examined the most efficient fee structure and product price when merchants are not allowed to price discriminate their customers. This subsection considers how the change in this assumption affects the most efficient fee structure and product price.

In each scenario, the maximum social welfare cannot be lowered when merchants set different prices according to their customers' payment methods, because for social planners the merchants' ability to price discriminate their customers means an additional variable they can control. It is easy to show that the merchant's discriminatory price setting does not affect the maximum social welfare when per transaction costs and fees are fixed (Scenarios I and III). In contrast, when per transaction costs and fees are proportional to the transaction value (Scenarios II and IV), it is likely that the merchant's ability to set different prices increases the maximum social welfare: The maximum social welfare is more likely to be achieved when the product prices for card-using consumers and for non-card-using consumers, respectively, are set at the merchant's marginal costs, and the sum of a cardholder fee and a merchant fee equals the card network's cost. Depending on the relationship between the card network's cost and consumers' transactional benefit from cards ($[\underline{b_B}, \overline{b_B}]$), the marginal card user's transactional benefit from cards may be either higher or lower than $c - \hat{b}_S$, the most efficient marginal card user's

[17] See Appendix A.

transactional benefit from cards when merchants are *not* allowed to set different prices according to the payment methods.[18]

4. Welfare Distribution under the Most Efficient Fee Structure

This section considers how the most efficient card fee structure affects welfare distribution among different parties. Because we were not able to analytically obtain the most efficient card fee structure under Scenario IV, we consider only three Scenarios, I, II and III. First, we consider the case where merchants are not allowed to set different prices across payment methods, and then consider the case where merchants set different prices.

Under Scenario I, where consumers make a fixed number of transactions and per transaction costs and fees are fixed, only the cardholder fee determines social welfare. The product price and the merchant fee do not affect social welfare. To satisfy the card network's incentive compatibility, the merchant fee cannot be lower than the merchant transactional benefit from a card (i.e., $m \geq \hat{b}_S$). When the merchant fee is higher than the merchant transactional benefit from a card, then the product price set by the merchants is likely to be higher, compared with the level of the product price in the economy where no card products are available.[19] As a result, consumers who use the alternative payment method would likely be worse off, compared with the economy without cards. Some card using consumers, whose transactional benefit from a card is relatively low, would also likely be worse off, due to the higher product price. These consumers' payment choice between the card and the alternative payment method is very sensitive to the cardholder fee, because their transactional benefit from a card is very close to the cardholder fee. Some card-using consumers, whose transactional benefit from a card is relatively

[18] The most efficient fee structure and product price under Scenario II are solved in Appendix B.
[19] Unless merchants are monopoly.

high, would likely be better off, because their transactional benefit from cards would likely exceed the welfare loss due to the higher product price. These consumers' payment choice between the card and the alternative payment method is not sensitive to the cardholder fee because their transactional benefit from cards far exceeds the cardholder fee.[20] Thus, under the most efficient card fee structure, some consumers would likely be worse off, even when the surplus of consumers as a whole would increase. As long as merchants practice marginal cost pricing, their profits are not affected by the merchant fee set by the card networks.

Under Scenario II, where consumers make a fixed number of transactions and per transaction costs and fees are proportional to the transaction value, social welfare is affected by the cardholder fee and product price. When the aggregate surpluses from using a card are greater than the aggregate surpluses from using the alternative payment method, the product price set at the highest level maximizes the social welfare. This product price, however, would reduce the welfare of consumers who use the alternative payment method and who use a card but their transactional benefit from a card is relatively lower. If, on the other hand, the aggregate surpluses from using a card are smaller than the aggregate surpluses from using the alternative payment method, the product price set at the lowest level maximizes social welfare. Under this product price, consumers who use the alternative payment method would *unlikely* be worse off compared with the economy without cards at all. Depending on the most efficient product price, the most efficient card fee structure would possibly make some consumers worse off even if it would make the surplus of consumers as a whole increase.

Under Scenario III, where a consumer's demand function for goods is downward-sloping and per transaction costs and fees are fixed, the cardholder fee, the merchant fee and the product

[20] However, these consumers' choice among the card products, such as the brand of the card or the issuer of the card, may be very sensitive to the cardholder fees (rewards).

price, all affect social welfare. Under the most efficient fee structure and product price, the card networks and the merchants earn zero profits. The surplus of a consumer who uses the alternative payment method would not be reduced compared with the economy without cards. All card-using consumers would likely be better off. Unlike the previous two scenarios, under Scenario III, the most efficient fee structure and product price would not make any consumers worse off.

Finally, when the merchants set the different prices according to their customers' payment methods, the most efficient fee structure and product price would be less likely to affect the surplus of a consumer who uses the alternative payment method. Card-using consumers are better off with cards. Thus, the total consumer surpluses would be higher with cards than those without cards. The card networks would earn zero profits and the merchants would likely earn the same profits with and without cards.

5. Conclusion

This paper examined the most efficient fee structure and product price and their effects on the welfare distribution. Although it is not always the case, in most cases the most efficient cardholder fee is the difference between the card network's costs of processing a card transaction and the merchant transactional benefit from a card transaction. Therefore, in most cases, providing rewards to card-using consumers is the most efficient only when the merchant transactional benefit from a card transaction exceeds the card network's costs.

The most efficient fee structure and product price do not necessarily make all parties involved in the payment card market better off. Especially, consumers who use the alternative payment method, instead of a card, would likely be worse off, if the product price they face is higher. Since the product price is generally positively correlated with the merchant fee, the higher the merchant fee, the worse off the consumers who use the alternative payment method

are. Moreover, not all card-using consumers are better off with the most efficient fee structure and product price. Card-using consumers whose choice of payment method between the card and the alternative payment method is very sensitive to the cardholder fee (or rewards) would likely be worse off if the product price is higher. In contrast, card-using consumers whose transactional benefit from cards is high would likely be better off. Because of the incentive compatibility constraints, the merchants and card networks would be unlikely to incur losses under the most efficient fee structure and product price.

Whether the most efficient cardholder fee is positive or negative is an empirical question. Available existing cost studies, which used relatively old information on merchant costs, suggest that the most efficient cardholder fee may likely be positive, which implies providing rewards may not be the most efficient.[21] In order for policymakers to accurately evaluate whether currently provided payment card rewards are efficient or not, collecting comprehensive and updated information on costs and benefits of various parties is required.

References

Armstrong, Mark. 2006. "Competition in Two-Sided Markets," *RAND Journal of Economics*, 37(3): 668-691.

Baxter, William F. 1983. "Bank Interchange of Transactional Paper: Legal Perspectives," *Journal of Law and Economics*, 26: 541-588.

Carten, Margaret, Dan Littman, Scott Schuh, and Joanna Stavins. 2007. "Consumer Behavior and Payment Choice: 2006 Conference Summary," Federal Reserve Bank of Boston *Public Policy Discussion Paper*, No. 07-4.

Chakravorti, Sujit. 2003. "Theory of Credit Card Networks: A Survey of Literature," *Review of Network Economics*. 2(2): 50-68.

Food Marketing Institute. 1998. *EPS Costs: A Retailers Guide to Electronic Payment Systems Costs*.

[21] Hayashi (forthcoming 2009) calculates the most efficient cardholder fees using available empirical evidence.

Gans, Joshua and Stephen King. 2003. "The Neutrality of Interchange Fees in Payment Systems," *Topics in Economic Analysis & Policy*, 3(1).

Garcia-Swartz, Daniel, Robert Hahn, and Anne Layne-Farrar. 2006a. "The Move Toward a Cashless Society: A Closer Look at Payment Instrument Economics," *Review of Network Economics*. 5(2): 175-197.

Garcia-Swartz, Daniel, Robert Hahn, and Anne Layne-Farrar. 2006b. "The Move Toward a Cashless Society: Calculating the Costs and Benefits," *Review of Network Economics*. 5(2): 199-228.

Government Accountability Office. 2008. "Credit and Debit Cards. Federal Entities Are Taking Actions to Limit Their Interchange Fees, but Additional Revenue Collection Cost Savings May Exist," GAO-08-558.

Guthrie, Graeme and Julian Wright. 2007. "Competing Payment Schemes," *Journal of Industrial Economics*, 55(1): 37-67.

Hayashi, Fumiko. 2006. "A Puzzle of Card Payment Pricing: Why Are Merchants Still Accepting Card Payments?" *Review of Network Economics*, 5 (1): 144-174.

Hayashi, Fumiko. Forthcoming 2009. "Do Consumers Really Benefit from Payment Card Rewards?" Federal Reserve Bank of Kansas City, *Economic Review*.

Hayashi, Fumiko and Stuart E. Weiner. 2006. "Interchange Fees in Australia, the UK, and the United States: Matching Theory and Practice," Federal Reserve Bank of Kansas City *Economic Review*, 3rd Quarter: 75-111.

Katz, Michael. 2001. *Reform of Credit Card Schemes in Australia II*. Reserve Bank of Australia: Sydney.

McAndrews, James and Zhu Wang. 2006. "Microfoundations of Two-sided Markets: The Payment Card Example," Federal Reserve Bank of Kansas City Working Paper 06-04.

Reserve Bank of Australia. 2007. *Payment Costs in Australia: A Study of the Costs of Payments Methods*.

Rochet, Jean-Charles. 2007. "Competing Payment Systems: Key Insights From the Academic Literature," Reserve Bank of Australia, *Proceedings of Payments System Review Conference*, 5-18.

Rochet, Jean-Charles and Jean Tirole. 2002. "Cooperation among Competitors: Some Economics of Payment Card Associations," *Rand Journal of Economics*, 33(4): 549-570.

Rochet, Jean-Charles and Jean Tirole. 2006. "Two-Sided Markets: A Progress Report," *RAND Journal of Economics*, 37(3): 645-667.

Schmalensee, Richard. 2002. "Payment Systems and Interchange Fees," *Journal of Industrial Economics*, 50: 103-122.

Schuh, Scott, Oz Shy, and Joanna Stavins. 2008. "Payment Cards, Income, and Consumer Welfare," mimeo.

Schwartz, Marius and Daniel Vincent. 2006. "The No Surcharge Rule and Card User Rebates: Vertical Control by a Payment Network," *Review of Network Economics*, 5(1): 72-102.

Simes, Ric, Annette Lancy, and Ian Harper (2006) "Costs and Benefits of Alternative Payments Instruments in Austlaria," Melbourne Business School Working Paper No 8.

Wright, Julian. 2003 "Optimal Card Payment Systems," *European Economic Review*, 47: 587-612.

Wright, Julian. 2004. "The Determinant of Optimal Interchange Fees in Payment Systems," *Journal of Industrial Economics*, 52: 1-26.

Appendix

A. Providing rewards is unlikely the most efficient in Scenario IV.

We investigate whether $f = c - \hat{b}_S$ is higher than, lower than, or the most efficient level. If $\frac{\partial CS}{\partial f} > 0$ at $f = c - \hat{b}_S$, then $f = c - \hat{b}_S$ is lower than the most efficient cardholder fee. Suppose $c - \hat{b}_S \le 0$.

The incentive compatibility constraints imply, at $f = c - \hat{b}_S$:

$$\frac{dp}{df} = \frac{-p(1-H(f))\{a - bp + bp(\hat{b}_B - f)\}}{(1-\hat{b}_S)\{a - bp + bp(\hat{b}_B - f)(1-H(f))\}} < 0, \quad \text{(A1)}$$

and

$$\frac{\partial CS}{df} = \frac{dp}{df}[(a-bp)\{(\hat{b}_B - f)(1-H(f)) - \hat{b}_S\} + bp\int_f^{\bar{b}_B}(b_B - f)^2 h(b_B)db_B]. \quad \text{(A2)}$$

Since $\frac{dp}{df} < 0$ from equation A1, if [ ] in equation A2 is negative, then $\frac{\partial CS}{\partial f} > 0$. The second term in [ ] is always positive, however the first term in [ ] can be negative. In fact, unless the highest transactional benefit to consumers, $\bar{b}_B$, is quite high (i.e., $\bar{b}_B \ge c + \frac{1+H(c-\hat{b}_S)}{1-H(c-\hat{b}_S)}\hat{b}_S$), then the first term is negative. If a, the cash user's quantity demanded when the product price is zero, is high enough so that monopolistic merchants can set the positive product price, then $a - bp$ is always greater than bp when $p = \frac{d}{1-\hat{b}_S}$. Thus, [ ] is likely negative and $\frac{\partial CS}{\partial f}$ is likely positive when $f = c - \hat{b}_S \le 0$. This implies that providing rewards to card users is unlikely the most efficient.

B. The most efficient fee structure with merchants' discriminatory pricing in Scenario II.

The marginal card user is indifferent between using a card or an alternative payment method, which implies that:

$$p^A = p^C(1+ f - b_B^m),$$

where p^A and p^C are the prices for alternative payment methods users and for card users, respectively. The social welfare function is defined as:

$$\begin{aligned} SW = & \int_f^{\bar{b}_B} \{\upsilon - p^C(1+ f - b_B)\}h(b_B)db_B + \int_{\underline{b}_B}^{f} (\upsilon - p^A)h(b_B)db_B \\ & + \int_f^{\bar{b}_B} \{p^C(1-m) - d\}h(b_B)db_B + \int_{\underline{b}_B}^{f} \{p^A(1-\hat{b}_S) - d\}h(b_B)db_B \\ & + \int_f^{\bar{b}_B} p^C(f + m - c)h(b_B)db_B, \end{aligned} \tag{B2}$$

which is equivalent to:

$$SW = \upsilon - d - p^A\hat{b}_S H(b_B^m) + p^C(\hat{b}_B - c)(1 - H(b_B^m)), \tag{B2'}$$

where $\hat{b}_B$ is the average transactional benefit among card users.

The first-order conditions with respect to price for cash users, price for card users, and marginal card user are:

$$\frac{\partial SW}{\partial p^A} = -\hat{b}_S H(b_B^m) < 0, \tag{B3}$$

$$\frac{\partial SW}{\partial p^C} = (\hat{b}_B - c)(1 - H(b_B^m)) = 0 \tag{B4}$$

$$\frac{\partial SW}{\partial b_B^m} = \{p^C(c - b_B^m) - p^A\hat{b}_S\}h(b_B^m) = 0 \tag{B5}$$

Equation B3 implies social welfare increases as price for cash users decreases. The lowest price for cash users is $p^{A*} = \frac{d}{1-\hat{b}_S}$. An interior solution exists if $\bar{b}_B \geq c \geq \frac{\bar{b}_B + \underline{b}_B}{2}$.

$$b_B^{m^*} = 2c - \bar{b}_B,$$

$$p^{C^*} = \frac{d}{1-m} = \frac{\hat{b}_S d}{(1-\hat{b}_S)(\bar{b}_B - c)}, \text{ and}$$

$$p^{A^*} = \frac{d}{1-\hat{b}_S}.$$

If $\bar{b}_B < c$, then $b_B^{m^*} > \bar{b}_B$, which implies that social welfare is maximized when nobody uses the cards. If $c < \frac{\bar{b}_B + \underline{b}_B}{2}$, then $b_B^{m^*} < \underline{b}_B$, which implies that social welfare is maximized when everybody uses the cards.