Card versus cash: empirical evidence of the impact of payment card interchange fees on end users' choice of payment methods

Ardizzi, Guerino

Bank of Italy

29 May 2013

Journal of Financial Market Infrastructures **1**(4), 73–105

Card versus cash: empirical evidence of the impact of payment card interchange fees on end users' choice of payment methods

Guerino Ardizzi
Bank of Italy, Market and Payment System Oversight, Via Nazionale, 91 00184 Rome, Italy; email: guerino.ardizzi@bancaditalia.it

(Received February 4, 2013; revised May 25, 2013; accepted May 29, 2013)

Interchange fees in card payments are a mechanism to balance costs and revenues between banks for the joint provision of payment services. However, such fees represent a relevant input cost used as a reference price for the final fee charged to the merchants, who may be reluctant to accept cards and induce the cardholder to withdraw cash. In this paper, we empirically verify for the first time the effect of the interchange fee on the decision to withdraw cash and compare it with that of paying with payment cards, considering a balanced panel data set of Italian issuing banks. Finally, results show that there is a positive correlation between the cash usage and the level of the interchange fees. Accordingly, regulation of the multilateral interchange fee level may be an effective tool in reducing cash payments at the point of sale, although there is no clear evidence that a zero interchange fee rate (or a close-to-zero rate) would be optimal.

1 INTRODUCTION

The objective of this study is to perform a first empirical investigation on the impact of the interchange fee on the propensity to use automated teller machine (ATM) cash withdrawals versus card payments at the point of sale (POS). But first we have to take a step back and describe what interchange fees are and how they work the payment networks, in which payment services are jointly provided by intermediaries to the benefit of different end users in a widespread network. In the leading credit and debit card

I wish to thank Lucia Bazzucchi, Daniele Condorelli, Federico Crudu, Massimo Doria, Paola Giucca, Eva St. Once an anonymous referee, and all seminar participants at the lunch seminar held at the Bank of Italy (Rome, February 2013). The views expressed in this paper are those of the author and do not involve the responsibility of the Bank of Italy.

schemes, the merchant's bank (the acquirer) pays an interbank exchange fee (interchange fee) to the cardholder's bank (the issuer) for each payment card transaction, in compensation for services received. These fees may cover very simple interbank services or extremely complex ones (trademark, clearing, authorization, charge-back) that benefit different parties (banks, merchants and cardholders). Interchange fees are usually set multilaterally, that is, uniform fees (so-called multilateral interchange fees (MIFs)) are established by the governing bodies of the respective networks.

Moreover, heterogeneous types of customer in two-sided markets and network industries could justify the existence of interchange fee flows. In the case of payment card schemes, different own-price demand elasticities of cardholder and merchant can be noted. For instance, whenever the interchange fee is directly transferred (or surcharged) to the cardholder, network expansion could be hampered, as the cardholder would not be willing to bear high expenses on payment card transactions.

The economic literature generally endorses the maintenance of an MIF, but shows that, privately and socially, optimal interchange fees often diverge (Rochet and Tirole 2002, 2011). In addition, the utility functions of the public authorities dealing with the interchange fee problem are heterogeneous (Rochet 2007); for instance, competition authorities may particularly care about end-user surplus, and central banks may consider the payment system as a whole.

Competition authorities claim that economic theory is not sufficient to justify an MIF, and the causal link between MIFs and efficiencies needs to be demonstrated empirically, since the interchange fee is a cost for the acquiring bank and may become a *de facto* floor for merchant fees. Hence, setting a framework to determine which MIFs can be justified on a case-by-case basis is one way of creating more effective competition in payment card pricing.

After the adoption of the single currency in Europe, the issue of interchange fees has also characterized the long and difficult process underway to create a single euro payment area (SEPA). The European Commission recently launched a public dialogue on the present landscape of the cards network, including the possibility of regulating interchange fees (European Commission 2012).[1] The Commission also conducted several legal assessments on the two major international card payment schemes (Visa and Mastercard).

The central banks, and the Eurosystem, in pursuit of its mandate to promote the smooth operation of payment systems, generally claim that MIFs, if there are any,

[1] For instance, as regards SEPA Direct Debit payment instruments (which allow bank customers to give companies or other organizations authorizations to take money directly from their bank accounts to pay their bills in Europe) the European Commission has already issued an ad hoc Regulation (924/2009) aimed to cap the MIF and to prohibit it after a transitional regime.

should not lead to bad price signals toward payers and payees, distracting them from using more efficient payment instruments (Börestam and Schmiedel 2011).[2]

More recently, economists have further analyzed the effect of interchange fees on technological innovation and on consumers' choices of payment instruments, especially considering the substitution between cash and payment card. The debate has been directed toward the imposition of a "price cap" or "zero" MIF rule (Leinonen 2011).

However, there are still few empirical studies on the issue of interchange fees, which has been mostly addressed by the literature from a theoretical point of view. For instance, to the best of our knowledge, available empirical evidence does not include direct estimates on the effect of the interchange fee on the choice of cash, which is considered inefficient in several studies on the social cost of retail payment instruments (Schmiedel *et al* 2012).

So we can return to the main objective of this study: assessing the impact of the interchange fee on the propensity to use ATM cash versus card payments at the POS. Toward this aim, we consider a two-period balanced panel data set of about 300 Italian issuing banks. The test is particularly interesting to perform in Italy, a country characterized by a high propensity to use cash (Banca d'Italia 2012a).

In Section 2, we review the relevant literature on antitrust issues and the economics of interchange fees. In Section 3 we give details and descriptive statistics on the payment card market and the interchange fee both in Italy and the rest of Europe. Section 4 illustrates the model under analysis and the econometric approach. The results are discussed in Sections 5 and 6, while the conclusions and some policy indications are reported in Section 7.

2 ECONOMICS OF INTERCHANGE FEES

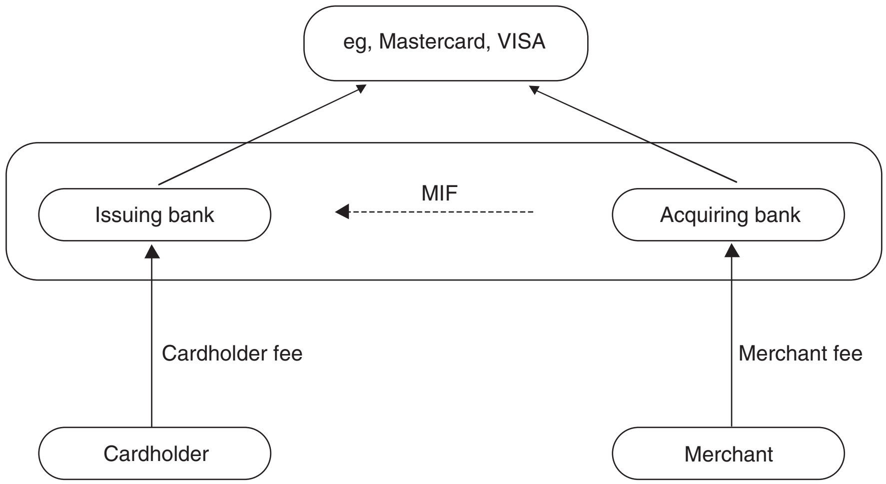
Interchange fees may be applied to any noncash retail payment network that involves more than one intermediary (such as checks, credit transfers, direct debits, etc). However, they are most typical of the payment card schemes. The credit and debit card networks are generally four-party schemes (Figure 1 on the next page), involving:

(1) the cardholder, who uses the card to purchase goods and services and generally pay an annual fee to the issuer;

(2) the issuer, the bank that issues the payment card to the cardholder;

[2] In Italy, from 1998 to 2005, the Banca d'Italia (as the competition authority and overseer) established rules and requirements on the level of fees in Italy, and its approval was required after every revision. Since 2006, such rules have been adopted and updated by the general Antitrust Authority.

FIGURE 1 Four-party scheme.



(3) the merchant, who accepts the card as payment for goods and services and pays the merchant fee to the acquirer;

(4) the acquirer, the merchant's processing agent who recruits merchants to the scheme, processes their card transactions, obtains funds from the card issuer and reimburses the merchant.

In the case of payment cards, costs are usually skewed to the issuing side and acquirers can have more revenues, provided that merchant fees are related to the turnover and the cardholder fees are fixed per payment card. The interchange fee paid by the acquirer to the issuer is intended to offset such cost/revenue imbalances, so the issuer can issue more cards and maximize the system output without surcharging the cardholder, who would be unwilling to make purchases by card.[3]

During the last twenty years many theoretical models have been developed, under different assumptions, to justify the existence of the interchange fee. For a long time

[3] The payment card associations still consider their interchange fee as "a financial adjustment to reduce the imbalance between the costs associated with issuing and acquiring, with a view to increasing demand for use of the payment services" (Visa International 2001). The governance authorities of the card payment schemes also argue that if each of the network's thousands of participants were to negotiate interchange fees bilaterally with each of the other thousands of network participants, the costs would be prohibitive. Furthermore, if any of the resulting agreements failed, some merchants would no longer accept certain issuers' cards, and the very notion of the network would be impaired.

there has been a general consensus (see Börestam and Schmiedel 2011) on the fact that payment card interchange fees may be useful in order to increase electronic payments. The two most frequently quoted articles on the economics of interchange fees are Baxter (1983) and Rochet and Tirole (2002). In his important early work, Baxter (1983) argues that under the assumption of perfect competition the socially optimal interchange fee is generally nonzero, that internal forces will drive the interchange fee to the socially optimal level and that the authorities should not consider interchange fees in payment card networks negatively. Twenty years later, Tirole and Rochet (2002) analyzed the cooperative determination of the interchange fee and concluded that raising the interchange fee will increase the use of credit cards so long as the fee does not exceed a level at which merchants no longer accept the card. A higher interchange fee lowers cardholders' fees, so that consumers who previously were not cardholders are induced to become cardholders. The optimal interchange fee for issuers is the highest level at which merchants continue to accept the card, and it is significantly different from zero.[4]

More recently, Rochet and Tirole (2003, 2011) have extended their analysis through applying the theory of the "two-sided market", in which, under certain conditions (eg, different demand elasticities), one side of the market will pay relatively less than the other side, to take into account some positive indirect network externalities in the market. In the payment card scheme, the side of the market that would pay relatively less is that of the cardholder, since the issuing bank receives the interchange fee revenues from the acquiring bank. On the acquiring side, the merchant fee, which includes both the interchange fee input costs and the acquiring internal costs, will be charged to the merchant. This solution should allow the merchant to increase their sales at the POS.[5]

---

[4] Such an outcome is valid under the no-surcharge rule (issued by the self-regulatory bodies in order not to crowd out their payment cards), which prohibits affiliated merchants from charging higher prices to customers who pay with credit cards or from offering discounts to those who use other payment instruments, such as cash. Some economists claim that if such rules are removed (and surcharges are allowed), the level of MIFs would not impact card usage (neutrality), as the cost and benefit are transferred efficiently to the end users (Gans and King 2003; Zenger 2011). Nevertheless, retailers may be reluctant to surcharge (Jonker 2011; ITM 2000; European Commission 2010; Börestam and Schmiedel 2011). Moreover, in some countries, including Italy, surcharge to electronic payments is prohibited by law in order to reduce the risk of promoting cash (see Coppola 2011; Doria 2010), as also pointed out by some empirical studies (Bolt *et al* 2008).

[5] In their contribution, Rochet and Tirole (2011) give a practical rule in order to internalize usage externalities in two-sided payment card markets. This is known as the "tourist test", or "merchant indifferent test" (MIT, as renamed by the competition authorities) and defines the optimal level of interchange fees. As a result the "tourist test" understates the threshold interchange fee (and corresponding merchant service commissions) at which a rational merchant would be indifferent between accepting cards or cash for a particular transaction.

With regards to the competition policy issues, antitrust authorities admit that MIFs are one way to internalize network externalities and thus to optimize card usage. However, MIFs also restrict competition and might be, therefore, prohibited to the extent that they do not generate sufficient efficiencies.[6] In the more relevant antitrust decisions (European Commission 2010), the European Commission has imposed caps or some kind of audited self-regulation to reduce interchange fees over the year, under the assumption that such fees might undermine competition and inflate final prices.

In this context, there is a common perception that on the acquiring side the merchant is charged according to an "interchange fee plus" model; accordingly, the merchant fee (mf) is calculated as follows:

$$\mathrm{mf} = \mathrm{if} + c + qc. \tag{2.1}$$

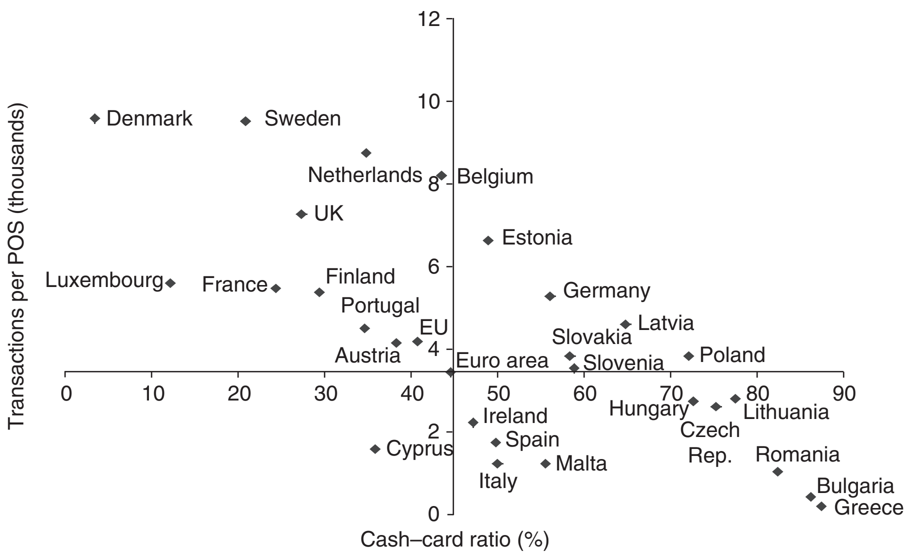
In the notation of (2.1), "if" is the unit interchange fee, c represents the direct unit internal cost for the acquirer and q is the margin in proportion to the direct costs to cover indirect costs together with the profit. The antitrust authorities fear that such a pricing mechanism sets a *de facto* floor limit on the price that merchants must pay to the acquirer for accepting payment cards. We consider such an assumption in our empirical model of analysis in Section 4.

The next step in the literature is then to relate the MIF with a bias toward the use of cash. A recent discussion paper published by the Bank of Finland (Leinonen 2011) has focused on the problem of the MIF and the "cash cross-subsidies" on the issuing side. According to Leinonen, "the higher the MIF, the larger the cross-subsidy for cash", provided that interchange fees increase payment costs for the merchants, who become reluctant to accept cards instead of cash, and thus reduce the possibility of passing on to customers the cost savings resulting from card efficiency.

Unfortunately, there is a scarcity of empirical contributions evaluating the effects in the real world of the interchange fee mechanism, due to the lack of reliable data on this matter. From the available studies in Europe, a decreasing trend in different European countries is associated with an increased usage of cards and competition between card schemes (Börestam and Schmiedel 2011; Bolt and Schmiedel 2013). Chakravorti *et al* (2009) analyze the positive effect on the growth of payment card transactions after the interchange fee regulation in Spain (ceilings on the MIF levels).

[6] The multilateral setting of the interchange fee among participants of a payment scheme, even where it is a default level with the possibility to negotiate a lower fee on a bilateral basis, may constitute a restriction of market competition pursuant to antitrust legislation. In allowing such a multilateral agreement, the antitrust authority must determine whether setting a multilateral fee "may improve the supply of payment services since banks avoid a large number of bilateral negotiations and transaction costs are reduced", with potential benefits for the final customer (Bank of Italy, Provision 23, October 8, 1998).

FIGURE 2 Cash card ratio and payment cards transactions.



Data for card payments per POS in 2011. *Sources:* European Central Bank, Blue Book statistics.

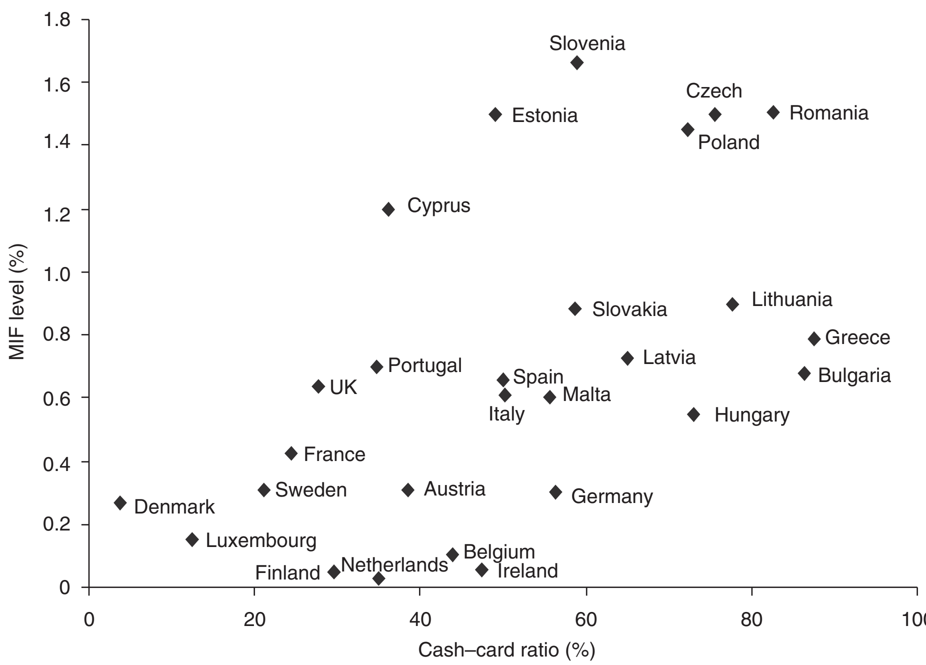
Nevertheless, the aforementioned authors test the impact of the interchange fee cap regulation on the merchant acceptance of cards in Spain but the issue of the impact of the interchange fee on the cardholder's inclination to shift to cash remains unsettled (Börestam and Schmiedel 2011).

Therefore, we shall test empirically the cash demand in relation to the interchange fee in Italy, after a brief description of the payment card landscape.

3 PAYMENT CARD MARKET, THE CARDHOLDERS' PROPENSITY TO USE CASH AND THE INTERCHANGE FEE

The payment card is the most prevalent noncash instrument in Europe. Card transactions accounted for 40% of cashless payments in 2010, and volumes are increasing at around 6–7% per year (European Central Bank statistics 2012). In Italy there were over 1.5 billion annual payment card operations in 2010, representing 38% of the total noncash payments that year (compared with 21% in 2000), but the "gap" with other industrialized countries is still large: only twenty-seven card transactions per capita annually compared with over seventy in the whole EU in 2010 (Banca d'Italia 2012b).

FIGURE 3 MIF and cash–card ratio.



Sources: European Central Bank, Visa EU, Mastercard.

The latest data confirms that cash is still the most widely prevalent payment instrument in Italy, to a greater extent than in other industrialized countries (over 90% of transactions at the POS, compared to around 80% in the rest of Europe (see Schmiedel *et al* 2012)). This is also confirmed by comparing the cash–card ratio per country. The cash–card ratio gives the value of cash acquired from ATMs divided by the total value of card turnover (the total value of ATM withdrawals plus the total value of debit POS expenditure by card (see Jones and Jones 2006)). In Figure 2 on the preceding page, the cash–card ratio per country is plotted in relation to the number of domestic card payments divided by the number of POSs: it is obviously a negative correlation, and Italy is on the side of the chart (the axes are centred on the mean values of the distribution) with countries that present a low level of POS utilization and a high propensity to use a card to acquire cash.

Figure 3 shows the average payment card interchange fee (as a percentage) by country as well as the cash–card ratio. The principal sources of data were the card payment schemes and the European Central Bank (Börestam and Schmiedel 2011).

There is a positive correlation between the use of debit cards at ATMs and interchange fee levels: in those countries where the interchange fee is the lowest, debit cards seem to be more widely used and accepted at the POS. Moreover, there is a significant variability between average levels of the MIF (on the horizontal axis): the average (transaction-weighted) MIF varies from a minimum of about 0.01% to over 1.55% in different member states (Börestam and Schmiedel 2011).

4 MODEL OF ANALYSIS

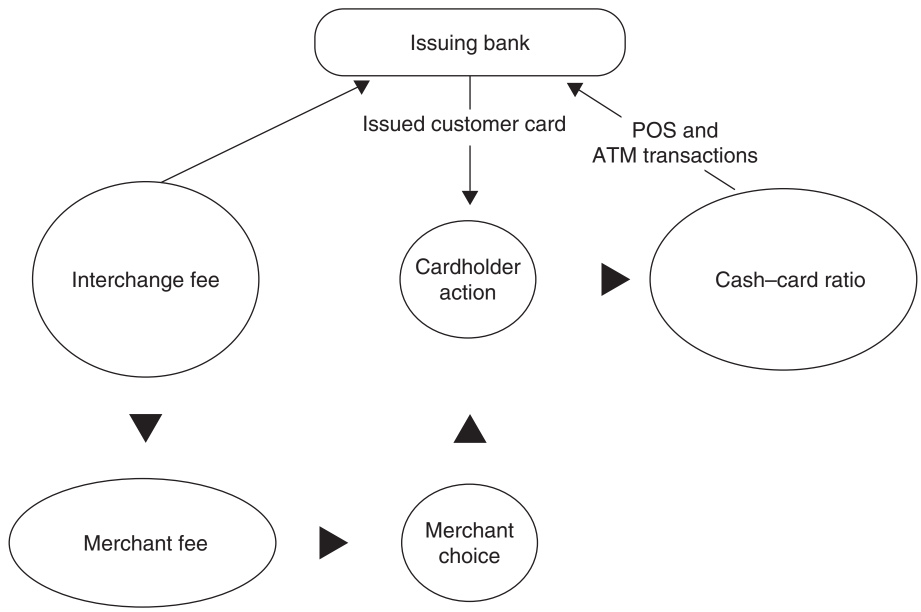
The model of analysis is based on the functioning of a payment card on the issuing side (Leinonen 2011): banks or other payment service providers issue payment cards in order to allow the cardholders to charge purchases or withdraw cash directly against funds on a transaction account at a deposit-taking institution (debit cards), prepaid account (prepaid cards) or according to credit facilities (credit cards). Payment cards (especially debit cards) may be used for ATM transactions as well as at POS terminals at retail locations.[7] When the payment card is used directly in shops, the issuer receives interchange fees revenues from the acquirers in the case of "not on us" transaction (when the issuer and the acquirer are not the same party). If the issuer and the acquirer are the same party (an "on us" transaction), the interchange fees are not applied but remain an implicit acquiring cost element in the pricing to the merchant. In every case, the cardholder does not usually pay variable fees at the POS.

When the card is used in the ATM card networks, the cardholder is charged a fee only if the transaction is performed at a "foreign ATM" (owned by an institution different from the issuing bank), because in the case of "foreign transactions" the issuer will pay an interbank service fee to the ATM owner. Nevertheless, in the case of cash withdrawals at the issuer's ATM, transactions are usually free of charge for cardholders. In fact, over 75% of ATM transactions are carried out at issuer's ATM terminals in Italy (Ardizzi and Coppola 2002).

In this work we jointly consider all types of cards (debit, credit, prepaid) for a variety of reasons. First of all, in this way we can remain aloof of any substitution effects between cards. Second, many merchants agree on a product bundle which is offered at a "blended" price that makes it difficult to compare different card types cost of acceptance (European Central Bank 2006). Finally, available revenue information on the cardholder is not always easily distinguishable between types of cards. For this reason, debit card networks usually show lower interchange fees than credit cards (European Commission 2010), and issuing banks specializing in credit cards receive higher average levels of interchange fee.

[7] However, in this work we consider all the payment card types (debit, credit, prepaid) issued by banks, so as to control for any substitution effect among cards and because of difficulties in distinguishing statistical information.

FIGURE 4 Transmission channel of the interchange fee mechanism.



On the issuing side we can correlate the POS interchange fee and the behavior of the cardholder interacting with the merchant. Given the "interchange fee plus" mechanism (Section 3), if interchange fees are too high, merchants might be reluctant to accept the card at the POS and induce the cardholder to shift to cash. Figure 4 illustrates the main transmission channels of the interchange fee mechanism on the issued customer card according to our model analysis.

This approach has the advantage that it does not require a simultaneous equation model (as in Chakravorti *et al* 2009 or Bolt *et al* 2008), which may impose several restrictions. We focus on the final information collected by the card-issuing bank, overcoming the problem of modeling the interaction between the merchant and the cardholder: the result of this interaction is given by the value of cash withdrawn and of POS payments carried out through the same cards, on which the issuer gains the interchange fee from the acquirer.[8]

[8] On the other hand, the acquiring bank does not collect information on the cash withdrawals with the same cards. Moreover, reliable information on the acquiring side is not available. For these reasons we do not consider a simultaneous equation model as in Chakravorti *et al* (2009).

Moreover, our model is valid even when merchants want (or are allowed) to apply a surcharge on card payments: if the interchange increases average unit transaction costs, a merchant will be less inclined to accept card payments or will be more likely to surcharge card payments (Jonker 2011; Bolt *et al* 2010) with the same effect: to induce the cardholder to withdraw cash at the ATM.

We can then assume a linear relationship between the cash–card ratio, the MIF and other control variables Z, with conditionally independent errors $E(\varepsilon_{it} \mid Z_{it}, \mathrm{MIF}_{it}) = 0$, i banks and t periods, and run a regression model such as the following:

$$\text{cash ratio}_{it} = \alpha_0 + a_1\mathrm{MIF}_{it} + \sum_h \alpha_h Z_{it} + u_{it}. \tag{4.1}$$

The dependent variable (cash ratio) is equal to the ratio of ATM cash withdrawals to the total amount of card turnover (at POS and ATM), which is consistent with the "cash–card ratio" described in Section 3.[9]

The first variable on the right-hand side of (4.1) is equal to the percentage of the average transaction-weighted (Chakravorti *et al* 2009) MIF received by the issuer on its cards. Its coefficient aims to capture the effect of the interchange fee on the cash–card ratio of the same cards issued by the bank.[10] To identify this effect, we follow a two-stage process. In the first stage we assume, as usual, linearity in the functional relationship. In this case, it is essential that we have linearity in the MIF parameter, but not necessarily essential in the predictor. Accordingly, in this section we shall determine the existence of a trend of the MIF impact with respect to the cash–card ratio in Italy. In other words, we verify whether an increase in the level of the MIF gives rise, on average, to more ATM cash withdrawals than card payments at the POS.

Thus, we put forward our first hypothesis:

(H0) under condition of linearity in the parameter, the higher the interchange fee, the higher the cash–card ratio on the issuing side, and vice versa.

In Section 6, as a second stage of the analysis, the linearity assumption will be removed in order to further investigate the issue "what about the optimal MIF?" Economic

---

[9] In this case, we did not include the ATM operations in the denominator so as to avoid the dependent variable being truncated between zero and one.

[10] Although the interchange fee levels are usually set by a self-regulatory body, various market conditions will, of course, affect the MIF variability at the level of the banks. This variability is remarkable and supposedly exogenous with respect to the payments and withdrawals of cash through ATMs. In other words, even if the selection of the card at the POS automatically affects the MIF received by the bank in our model, we are most interested in assessing how the variability of the interchange fee level affects the cash usage. Accordingly, self-selection problems related to the choice of the debit and credit cards may be omitted in our model.

theory (Rochet and Tirole 2002) affirms that interchange fees may increase the card usage at the POSs (accordingly, the cash ratio tends to decrease[11]) up to a threshold/optimal level, after which the acceptance costs surpass the benefit, the shift (return) to cash becomes relevant and the cash ratio rises. However, adding nonlinearity (or "second-order" effects) in the econometric model may be complex and requires further statistical checks; we focus on this issue in the next section.

The summation term among the covariates in (4.1) indicates the set of h environmental variables (Z), and that of the relative coefficients, which can influence the use of cash versus electronic payments at the POS. Control variables identify income components, which are also highly correlated with financial literacy, leading to less use of cash, and access point diffusion (ATMs, bank branches, POSs), which may affect the choice between cash and alternative payment instruments (Stix 2004; Humphrey *et al* 1996).

One of the control variables identifies the individual income component, which is also highly correlated with general education and financial literacy, leading to less use of cash and greater confidence in alternative payment instruments (Stix 2004; Humphrey *et al* 1996). In the model, such an income component is indirectly measured by the value of total turnover per card for transactions that are completed at POSs and ATMs.[12] Therefore, we consider $Z_1 =$ turnover. Even the expected effect of this variable on the cash–card ratio is positive, and the related hypothesis to be tested is the following:

(H1) the higher the average turnover per card, the lower the inclination toward the use of cash withdrawals as an alternative to electronic card payments.

A second control variable ($Z_2 =$ ATM) takes into account the relative size of the ATM card network managed by intermediaries, expressed as the ratio of the number of own ATMs and the number of own POSs.[13] A larger diffusion of the ATM card network may increase the probability of the bank having its own ATM cash withdrawals, which are free of charge for the cardholder (positive coefficient). We then formulate the following hypothesis:

---

[11] Actually, to the best of our knowledge the net impact on the cash–card ratio is not clearly investigated by the theoretical models.

[12] The turnover per card represents a proxy of the spending capacity of the cardholders, as we do not dispose of detailed information on the effective average income per cardholder.

[13] The latter standardization allows us to compare issuing banks characterized by different business strategies: ATM services versus POS or acquiring services. At this stage, we do not include separately the number of POSs in the equation, as the correlation between the number of ATMs and POSs is very high (0.82), increasing the risk of collinearity in the model. However, as a further test of stability we also include separately ATM (expected sign: positive) and POS (expected sign: negative) (see Section 5.3).

(H2) the bigger the ATM network, the higher the motivation toward the use of cash withdrawals as an alternative to electronic card payments.

Moreover, in order to control the over-the-counter (OTC) cash operations that could crowd out ATM cash withdrawals, our model includes an indicator of the relative incidence of the number of physical bank branches (OTC) to the number of automated cash machines or ATM ($Z_3 = \text{OTC}$). This control variable is expected to affect the cash–card ratio negatively, according to the following hypothesis:[14]

(H3) the higher the diffusion of the OTC network (number of branches to ATMs), the lower the motivation toward the use of cards to withdraw cash at ATMs.

As in the works of Chakravorti *et al* (2009) and Bolt *et al* (2008), we will not include the rate of interest in the model analysis: based on standard economic theory, the interest rate is expected to have a negative effect on the demand for money, via its role of opportunity cost of holding cash in alternatives to interest-bearing assets.[15]

However, since we want to further test the stability of our outcomes, in the robustness analysis we also include among the covariates a proxy of the interest rate levels calculated as the ratio of the total amount of interest expenses to the total amount of deposit liabilities (data is available for all the banks in the panel).

Finally, in the longitudinal models the term u_{it} in (4.1) can be broken down into an individual specific effect, a temporal effect and the stochastic disturbance (ε_{it}). In particular, the individual specific effect incorporates the unobservable elements[16] of "firm-specific" or "group-specific" heterogeneity, reducing the omitted variable bias

[14] As a further test of stability of the empirical results, we remove such a hypothesis from the equation model in Section 5.3 ("robustness"), by assuming that the effect of demand for OTC cash withdrawals is neutral for the dependent variable and homogenously affects both ATM (negatively) and POS (negatively) transactions.

[15] One reason to omit the interest rate at this first stage is that available banking data on interest rates is not consistent, as it does not consider only the cardholders' current accounts and does not fully match that on our balanced panel database. Moreover, the inclusion of the deposit interest rate may hamper the endogeneity problems, since it may be simultaneously affected with other variables (ie, size and institutional ones). Furthermore, several studies investigating the role of innovative payment systems in cash demand of Italian households (Lippi and Secchi 2009; Alvarez and Lippi 2009) point out that the progress in transaction technology may substantially reduce (or even eliminate) the impact of the interest rate on the cash demand of buyers, also considering that the period covered by our estimations was characterized by very low interest rates, which are likely to have strongly mitigated the speculative motive. Finally, our model deals with a ratio of cash-to-card payment flows rather than stocks of liquid assets, which implies the effect of the interest rate is ambiguous, which could in principle impact proportionally on both denominator and numerator of the ratio, leading to null overall effect.

[16] These elements may, for example, be linked to the internal payment procedures, to the type of customer and to the differences in the business strategies.

TABLE 1 Panel data set. [Continues on next page.]

(a) Definition of variables and data sources

Variable	Definition	Source
MIF	Average transaction-weighted interchange fee received by the issuing bank	BoI banking statistics and survey on the cost of payment instruments
Cash ratio	Ratio of the value of total ATM cash withdrawals accounts to the value of total POS payments with issued debit and credit cards	BoI banking statistics
ATM	Ratio of the number of owned ATMs to number of owned POSs by the issuing bank	BoI banking statistics
OTC	Ratio of the number of bank branches to the number of ATMs	BoI banking statistics
Turnover	Total value of card operations to total number of issued cards	BoI banking statistics
I rate	Ratio of the total amount of interest on deposit to the total value of bank deposits	BoI banking statistics

(b) Descriptive statistics

Variable		Mean	SD	Min	Max	Obs
MIF	overall	0.007	0.005	0.000	0.026	$N = 546$
	between		0.004	0.000	0.023	$n = 273$
	within		0.002	−0.004	0.018	$T = 2$
Cash–card ratio	overall	1.879	1.509	0.685	11.468	$N = 546$
	between		1.492	0.860	8.593	$n = 273$
	within		1.104	1.031	3.424	$T = 2$
ATM	overall	0.043	1.724	0.006	0.534	$N = 546$
	between		1.708	0.007	0.247	$n = 273$
	within		1.109	0.020	0.093	$T = 2$
Turnover	overall	1383.879	1.265	493.454	2915.985	$N = 546$
	between		1.245	727.352	2453.721	$n = 273$
	within		1.090	846.756	2261.716	$T = 2$
OTC	overall	0.921	1.357	0.333	3.000	$N = 546$
	between		1.348	0.412	2.500	$n = 273$
	within		1.067	0.582	1.456	$T = 2$

TABLE 1 Continued.

(c) Correlation matrix

Variable	Cash ratio	MIF	Turnover	ATM	OTC
Cash ratio	1.0000				
MIF	0.3928	1			
Turnover	–0.1787	–0.0089	1		
ATM	0.0778	–0.0038	0.0702	1	
OTC	–0.1789	–0.1108	–0.1308	–0.0691	1

in the estimates. We do not formulate the hypothesis here. The time-specific effect can be captured by providing a time dummy variable, which may be useful for considering the effects of business cycle influences or technological changes (Chakravorti *et al* 2009).

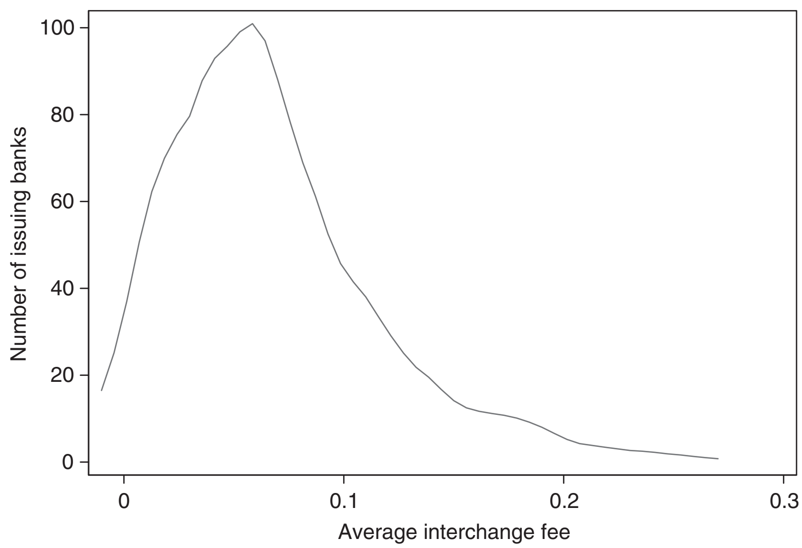
5 ESTIMATION OF THE LINEAR SINGLE EQUATION MODEL

5.1 Data set and estimation methodology

We use bank data drawn from the reports of the intermediaries on payment services collected by Banca d'Italia and from the survey on the costs of retail payments in Italy conducted in 2010 in close cooperation with the European System of Central Banks (Schmiedel *et al* 2012; Banca d'Italia 2012a). Combining the different sources of information, on the basis of the available data (accumulated at the bank level) it has been possible to construct a biannual (2009–2010) balanced panel of 273 issuing banks representing about 60% of the debit and credit card market in Italy. The database contains data (counted from the side of the issuing bank for 2009) concerning the interchange fee levels, cash and POS transactions, cards and accepting terminals, branches and other firm-specific variables. Table 1 on the facing page reports the definition, descriptive statistics and information about the different data sources for the whole sample.

As has been noted in the theoretical literature (Zenger 2011), payment networks typically differentiate their interchange fees by setting a variety of sector-specific MIFs for the same payment card. Figure 5 on the next page shows the density function of the average (transaction-weighted) percentage interchange fee for issuing banks of our sample. A significant variability of the average rates is evident at a bank level, not so different to that shown on Figure 3 on page 80 at a country level. This is due to different card payment networks (national debit card and international debit and credit cards), different pricing schemes (two-part tariffs, *ad valorem* fees, flat fees)

FIGURE 5 Interchange fee (*ad valorem*).



Kernel = Epanechnikov, bandwidth = 0.0011. *Sources:* balanced panel data, banking statistics (2009, 2010).

and different technologies of transaction (eg, base, chip and pin, enhanced electronic). Average MIF variability also affects the average merchant debit and credit discount fee variability.[17]

The parameters of (4.1) were estimated using the balanced panel data observed in 2009 and 2010. The dependent variable and the covariates described in the previous chapter are expressed in terms of logarithms in order to reduce the dispersion and the asymmetries. Robustness checks adopting further estimation techniques are conducted in Section 5.3.

Several methods have been proposed for the estimation of panel data models with a large number of cross-sectional units observed over a rather short period of time (in our case, $N = 273$ and $T = 2$). The estimated values of the static coefficients in

[17] Detailed information on the current nominal interchange fee rates in Italy are available through the websites of the card payments associations. See, for example, www.visaeurope.com/en/about_us/our_business/fees_and_interchange.aspx, www.mastercard.com/us/company/en/whatwedo/interchange/Intra-EEA.html and www.bancomat.it/it/infoutenti/esercenti.html.

(4.1) can be obtained by classic panel model estimators[18] with fixed, random, between effects and by the standard (pooled) least squares (OLS) estimator.

The Hausman test (fixed versus random effects) and the J-test for overidentifying restrictions robust to heteroscedasticity, strongly support ($\chi^2 > 22$; p-value $<$ 0.0002) the FE model and reject the assumption that the unobserved bank-specific effect is independent from the explanatory variables. Moreover, the Breusch and Pagan (1980) test refutes the hypothesis of "poolability" (cross-sectional model instead of panel model: $\chi^2 > 195$; p-value < 0.0001). However, Cameron and Trivedi (2005) argue the chi-square test and the "fixed effect" estimator may be subject to statistical problems when the "within" variability is dominated by the "between" variability of the panel (see Table 1 on page 86). Accordingly, in Section 5.2, we show the outcome for each estimation procedure (FE, RE, BE, OLS) to better evaluate the general performance of the model analysis.

First of all we estimate the base model, which considers only the percentage interchange fee (MIF), the card spending (Turnover), and the time dummy variable between the covariates,[19] Then we include the other control variables (full model) and test the stability of the results with respect to the disturbances affecting the initial model; a set of institutional ((1) commercial banks, (2) post office, (3) cooperative and rural banks) and size ((1) small bank, (2) medium bank, (3) large bank) dummy variables are also included in the OLS, BE and RE estimators. A further test of stability including the (proxy of) interest rate is reported in the robustness analysis.

5.2 Results

The results of the estimates are shown in Table 2 on the next page.

The estimated models (base and full) show coefficients that are all statistically significant and with signs consistent with our theoretical hypotheses (H0)–(H3), after controlling for the institutional and size dummies, whose coefficients are not reported

[18] The between effects (BE) estimator exploits exclusively the "between" dimension of the date by regressing the individual averages of the dependent variable on the individual averages of the covariates; the fixed effects (FE) estimate exploits solely the "within" dimension of the data by a regression in deviations from individual averages; the standard least squares estimator is applied to the pooled data, which can be shown to be an (inefficient) average of the between and within estimators; the random effects (RE) estimator, which is an efficient average of the between and within estimators, while the weighting is based on the ratio of the variances of the individual specific effect and the stochastic disturbance.

[19] A time dummy variable may be useful to take into account the effect of the business cycle influences or technological changes even in a biannual model.

TABLE 2 Estimation of the linear equation model (4.1) with dependent variable: cash ratio. [Table continues on next page.]

	Model							
	(1) Fixed effect		(2) Random effect		(3) Between effect		(4) Pooled OLS	
Regressor	**Base**	**Full**	**Base**	**Full**	**Base**	**Full**	**Base**	**Full**
MIF	0.0625*** (0.0204)	0.0624*** (0.0214)	0.0814*** (0.0171)	0.0788*** (0.0159)	0.117 (0.0245)	0.101*** (0.0245)	0.111*** (0.0171)	0.0943*** (0.0170)
Turnover	−0.181*** (0.684)	−0.181*** (0.0688)	−0.249*** (0.0569)	−0.274*** (0.0569)	−0.453*** (0.104)	−0.571*** (0.104)	−0.411*** (0.0704)	−0.508*** (0.0698)
ATM		−0.00912 (0.0650)		0.0608* (0.0346)		0.146*** (0.0430)		0.134*** (0.0308)
OTC		−0.0612 (0.103)		−0.196*** (0.0594)		−0.287*** (0.0790)		−0.274*** (0.0563)
Dummy time	−0.0259** (0.0117)	−0.0267** (0.0119)	−0.0226* (0.0126)	−0.0234** (0.0119)			−0.0168 (0.0331)	−0.0175 (0.0313)
Dummy size				Yes		Yes		Yes
Dummy institutional				Yes		Yes		Yes

TABLE 2 Continued.

	Model							
	(1) Fixed effect		(2) Random effect		(3) Between effect		(4) Pooled OLS	
Regressor	**Base**	**Full**	**Base**	**Full**	**Base**	**Full**	**Base**	**Full**
Constant	2.283**	2.251**	2.874***	2.879***	4.528***	5.456***	4.200***	4.926***
	(0.481)	(0.532)	(0.408)	(0.430)	(0.766)	(0.853)	(0.517)	(0.576)
Observations	546	546	546	546	546	546	546	546
R^2	0.101/0.894	0.103/0.894	0.123	0.223	0.131	0.254	0.124	0.235
AIC	−1022.906	−1026.117	116.6754	109.1199	241.1257	219.2909	514.8956	461.291
Number of banks	273	273	273	273	273	273	273	273

All variables enter the regressions in logarithms. Standard errors robust to heteroscedasticity are in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$. Consider R^2 within/overall adj-R^2 for the fixed effect model. Consider AIC the ML estimator for the random effect model.

for the sake of brevity.[20] The negative sign of the time dummy variable seems to be consistent with a linear trend of growth of card payments in Italy.

As expected, the coefficient of the interchange fee variable is consistently positive. The magnitude of the effect, moreover, is quite significant: a decrease of 10% in the MIF rate is associated with a reduction in the cash ratio of approximately 1%. Such an outcome also represents an indirect test that the "interchange fee plus" mechanism is present and that interchange fees are not neutral.[21] Such results seem to be consistent with those obtained by Chakravorti *et al* (2009).

The value of turnover per card ("Turnover") shows a significant negative impact upon the cash–card rate, as well as the higher relative diffusion of physical branches (OTC and ATM). The relevance of the ATM/card network dimension turns out to be significant and positive, given that such a variable may be a proxy of the probability that the intermediary intercepts the cards used at its own ATMs by providing free cash withdrawals.[22] In the "fixed effect" model the ATM and OTC coefficients lose significance. This can be partly explained by the fact that such (infrastructural) variables do not vary significantly in the biennium, and the inclusion of the firm-specific dummy variables capture unobserved bank heterogeneity that is constant over time.

5.3 Robustness checks

We conducted several robustness checks of the outcomes illustrated in the previous subsection, using alternative estimation methods that control for the presence of

(1) contemporary heteroscedasticity and autocorrelation of the residual terms,

(2) nonnormal distribution of the variables,

(3) endogeneity problems due to simultaneous causality or omitted variables.

Each of the above points highlights a violation of the assumptions underlying the linear regression models and can make the results inconsistent.

[20] The full model performs quite well also in term of fit: the F-statistics is significant at 1%; the overall adj-R^2 tends to 0.89 in the fixed effect specification (least squares dummy variable regression); we will deal again with this issue in the robustness checks. Moreover, main results are robust aggregating the information of the intermediaries that belong to the same banking group, in order to control for possible "group" specific effects. For the sake of brevity, we do not present the results of these tests, but they are available on request from the author.

[21] Based on empirical assessment of the Reserve Bank of Australia payment system reform, Hayes (2007) finds no significant impact of interchange fee level on the card usage. Nevertheless, this partly contradicts Chang *et al* (2005) and Chakravorti *et al* (2009).

[22] This result is confirmed if we include ATM (expected sign: positive) and POS (expected sign: negative) separately in the model (see Table 5 on page 97).

The method used to control the first distortion factor, indeed, relies on proper statistical tests,[23] we cannot reject the hypothesis of contemporary heteroscedasticity and cross-sectional and autocorrelation that can lead to biased statistical inference in the panel (Cameron and Trivedi 2005). Therefore, in order to adjust the standard error appropriately, we decide to apply the linear estimator with a panel-corrected standard errors (PCSE) estimator suggested by Beck and Katz (1995). In particular, we specify that, within groups, there is first-order autocorrelation and that the coefficient of the AR(1) process is specific to each group.[24] In addition, we consider a standard OLS regression robust to bank-specific clustered standard errors.[25]

Regarding the second distortion,[26] we consider a "quantile" regression estimator ("Quantile" estimator), where the relationship between y and x is not expressed by the variation of the conditional mean of y given x (classical linear model), but by the variation of one of its quantiles (eg, median). This approach is useful in the presence of nonnormal distributions of the dependent variable, or of high statistical dispersion, which may make the mean value less significant. Furthermore, it may be interesting to calculate the impact of the MIF rate at different levels of the cash–card ratio (ie, 25th or 75th percentile).[27] For this method we have also resorted to the nonparametric bootstrap to calculate the standard errors and test the significance of the estimated

---

[23] Specifically, we used the Wooldridge (2002) test for autocorrelation in panel data, the Greene (2000) test for groupwise heteroscedasticity and the Pesaran (2004) test for cross-sectional dependence in panel data.

[24] In particular, we consider the Prais–Winsten generalized least squares (GLS) estimator, derived for the AR(1) model for the error term, which represents a further innovation of the original OLS PCSE method proposed by Beck and Katz (Stata Technical Bulletin 1995).

[25] The usual OLS assumption is that standard errors are independently and identically distributed, but this assumption is clearly violated in many cases. A natural generalization is to assume "clustered errors", ie, that observations within groups (i banks) are correlated in some unknown way, inducing correlation in e_{it} within i, but that groups i and j do not have correlated errors. In the GLS-PCSE model we also remove the latter assumption.

[26] A standard Shapiro–Wilk test for normality rejects such an assumption (Racine 2008). The method most appreciated when addressing the problem arising from nonnormal distribution of the variables is nonparametric statistical techniques, which are robust to functional misspecification and do not require a researcher to specify functional forms for the objects being estimated. However, we leave this extension to future research.

[27] The estimation for quantiles is conducted on the "pooled" panel, in order to gain degrees of freedom. The quantile regression applied to panel models in fact requires a large sample size in order to unbundle the unobservable individual specific effects and produce consistent estimates (see Koenker 2004). Also, differencing (or de-meaning) the data, as we would do under an OLS framework, is not appropriate for quantile regressions: the quantiles of the sum of two random variables are not equal to the sum of the quantiles of each random variable. Moreover the interpretation given to individual fixed effects is less appealing in quantile regression models, as the quantile regression already accounts for unobserved heterogeneity and heterogeneous effects.

TABLE 3 Robustness checks against violations of the linear regression assumption. [Table continues on next page.]

	Model								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Least Sq				Quantile				
Regressor	**FE cluster**	**PCSE**	**IV°**	**IVFE°**	**q5**	**q25**	**q50**	**q75**	**q95**
MIF	0.0643***	0.0758***	0.1112***	0.2673***	0.0731	0.0735**	0.0644***	0.0849***	0.0992***
	(0.0207)	(0.0132)	(0.0278)	(0.0695)	(0.0460)	(0.0320)	(0.0160)	(0.0158)	(0.0246)
Turnover	−0.154	−0.357***	−0.460***	−0.022	−0.0285	−0.446***	−0.700	−0.889***	−0.910***
	(0.128)	(0.0487)	(0.148)	(0.763)	(0.0883)	(0.169)	(0.0723)	(0.151)	(0.166)
ATM	−0.0220	0.115	0.1378***	0.0826	0.147**	0.139***	0.115***	0.0814	0.0467
	(0.0991)	(0.0279)	(0.0420)	(0.0579)	(0.0641)	(0.0480)	(0.0316)	(0.0542)	(0.0625)
OTC	−0.0406	−0.196***	−0.289***	−0.0899	−0.156*	−0.247***	−0.285***	−0.377***	−0.367***
	(0.107)	(0.0240)	(0.0864)	(0.1199)	(0.0879)	(0.0539)	(0.0426)	(0.0928)	(0.123)
“I rate”	−0.134***	−0.107***			−0.0741	−0.106*	−0.123**	−0.267***	−0.430***
	(0.0422)	(0.0275)			(0.0554)	(0.0566)	(0.0620)	(0.0494)	(0.106)
Dummy time	−0.0826***	−0.0637***		−0.0055	−0.113**	−0.0872***	−0.0630*	−0.0898***	−0.0938*
	(0.0198)	(0.0121)		(0.1600)	(0.0521)	(0.0330)	(0.0332)	(0.0246)	(0.0556)
Dummy size		Yes	Yes		Yes	Yes	Yes	Yes	Yes
Dummy institutional		Yes	Yes		Yes	Yes	Yes	Yes	Yes

TABLE 3 Continued.

	Model								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Least Sq				Quantile				
Regressor	**FE cluster**	**PCSE**	**IV°**	**IVFE°**	**q5**	**q25**	**q50**	**q75**	**q95**
Constant	1.448	3.475***	4.406***	2.718***	0.687	3.716***	5.474***	6.365***	6.408***
	(1.049)	(0.512)	(1.1712)	(0.8125)	(0.805)	(1.236)	(0.569)	(1.295)	(1.636)
DWH χ^2 endogeneity			1.4503	12.446*	—	—	—	—	—
First stage-F relevance			201.599***	8.987*	—	—	—	—	—
Sargan χ^2 overidentifying restrictions			1.668	0.183	—	—	—	—	—
Observations	546	546	273	546	546	546	273	273	274
R^2	0.14/0.89	0.84	0.24	0.08/0.93	0.13	0.14	0.15	0.22	0.28
Banks	273	273	273	273	273	273	273	273	273

All variables enter the regressions in logarithms. Robust standard errors are in parentheses. For quantile regressions, standard errors are based on bootstrap with 399 replications. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. °Notes: instrumental variables are the following. 2SLS regression (*Source:* Bank of Italy, banking statistics); IV – random effect; instrumented: MIF; instruments: MIF_{t-1}; $\mathrm{Turnover}_{t-1}$; $\mathrm{Network}_{t-1}$; as a further robustness check, similar results are obtained if we consider the following alternative instruments: MIF_{t-1}, dIMP POS, dIM PATM; for the sake of brevity, we do not present such results in the table. IVFE – fixed effect. Instrumented: MIF, instruments: AVG_POS, Network. Legend instrumental variables are the following. Network = number of ATMs owned × number of cards issued. AVG_POS = mean value of the card transaction at the POSs.

TABLE 4 Further robustness check against the assumption of exogeneity of the MIF: GMM estimator.

	A	B
Variable°	GMM RE	GMM FE
MIF	0.1545***	0.3280**
N	273	546

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. °Full model; all other explanatory (Z_i) variables are omitted. All variables enter the regressions in logarithms. Robust standard errors are omitted. Generalized method of moments (GMM): A, size and institutional specific dummies included; B, individual fixed effect included.

coefficients without necessarily making assumptions about the probabilistic model and the reference distribution of the sample.[28]

The third factor of distortion (simultaneous causality) is the possibility that the relationship between the cash–card rate and the interchange fee (or other covariates) is bidirectional or that there is nonzero contemporaneous correlation between the regressor(s) and the error term. For instance, if another unobserved variable jointly determines both a high cash–card ratio and high levels of interchange fees, the econometric models will not give consistent estimates. As noted above, the interchange fee variables are set by the self-regulatory body. Nevertheless, if some immeasurable aspects of the environment in which banks operate are associated with the acceptance, issuance or usage of cards, the risk of endogeneity bias in payment instruments analysis may increase (Chakravorti *et al* 2009). Thus, the standard Durbin–Wu–Hausman (DWH) test does not allow us to confirm the strict exogeneity of the MIF variable if the "fixed effect" estimator is applied, but only in the absence of firm-specific intercepts this assumption is not refuted.[29] Accordingly, we also consider a two-stage least squares estimator (2SLS), using lagged values of the MIF and other selected variables as instruments (see Chakravorti *et al* 2009).[30]

---

[28] The results reported consider the regression on the 5th, 25th, 50th, 75th and 95th percentiles of the dependent variable.

[29] If the "firm-specific" intercepts are not included, the assumption of exogeneity is not refuted. However, we consider also the instrumental variable regression without individual "fixed effect" (see Table 3 on page 94).

[30] Other instrumental variables are the current value of the average transaction at POSs (see Table 3 on page 94), which may affect the average interchange fee levels (relevance condition) and should not necessarily be correlated with the error term in (4.1) (validity condition), as confirmed by "first stage" 2SLS test (Hausman) and the Sargan–Hansen of overidentifying restriction test (see Wooldridge 2002). Moreover, we also consider the (lagged) instrumental variable "network" (equal to the product of the number of ATMs and the number of cards managed by the issuing bank) and the lagged value of the "Turnover".

TABLE 5 Further robustness check of stability to perturbations in the parameters (alternative specification of the model and GMM estimator).

	Model			
Variable	(1) RE	(2) FE	(3) GMM IVRE	(4) GMM IVFE
MIF	0.0811***	0.0719***	0.1248***	0.1683*
Turnover	−0.2900***	−0.1724	−0.5998***	−0.1196
ATM	0.1162***	−0.0923	0.2044***	−0.0935
Pos	−0.0935***		−0.1522***	−0.0449
N	546	546	273	546

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. All variables enter the regressions in logarithms. Robust standard errors are omitted. Additional variables: ATM = (log) number of ATMs; Pos = (log) number of POSs. Models (1) and (3) have size and institutional specific dummies included. Models (2) and (4) have individual fixed effects included. Instrumental variables for GMM models: A, (3), random effect. Instrumented: MIF. Instruments: MIF_{t-1}, Turnover_{t-1}, Network_{t-1}. As a further robustness check, similar results are obtained if we consider the following alternative instruments: MIF_{t-1}, dIM PPOS, dIMP ATM; for the sake of brevity, we do not present such results in the table. B, (4), fixed effect. Instrumented: MIF. Instruments: AVG_POS, Network. Legend instrumental variables: network = number of ATMs owned × number of cards issued. AVG_POS = mean value of the card transaction at the POSs.

The outcomes relative to the different robust estimators[31] are reported in Table 3 on page 94.

As a result, the robustness checks seem to be more than satisfactory. Across all of the methods adopted, the significance and the intensity (positive) of the MIF effect on the cash–card rate (cash ratio) are confirmed.[32] Moreover, the inclusion of the (proxy) of the interest rate does not affect the main results.[33]

---

[31] Least absolute deviation methods (quantile regression) may be affected by endogeneity problem as well. However, this issue is not well developed in literature yet and beyond the scope of this first empirical investigation. As a further test of robustness, we limit to apply the (two-stage) instrumental variable quantile regression model proposed by Chernozhukov and Hansen (2008), who recently introduced an ad hoc command in the Stata Software, considering the same instrumental variables included in the instrumental variable least squares regression. Such an extension confirms the main conclusions of this paper and results are available upon request.

[32] The selection of the "full" model as the best one is supported by the Akaike information criterion (AIC). However, the AIC measure gives less support to the model including interest rates together with the time dummy. The time dummy intercept is indeed significantly correlated (Pearson's correlation coefficient: +0.46) with the interest rate variation across the two years.

[33] Recall that a proxy of the interest rate has been included just as a test of stability of the estimates shown in Section 5.2. However, all the estimates shown in Table 3 on page 94 have been replicated without the "interest rate" variable and no significant changes in the relevant coefficients have been found. For the sake of brevity, we do not present such results.

The magnitude of the marginal effect of the interchange fee rate is stronger and more significant for banks in the upper tail of the distribution (quantile regression)[34] and after controlling for endogeneity in the instrumental variable (IV) least squares estimation[35] (Table 3 on page 94). Finally, we replicated the regressions through the generalized method of moments (GMM) method developed by Hansen, which is less affected by the distributional assumptions (such as normality) than the 2SLS and may be more efficient in the presence of heteroscedasticity of unknown form and weak instruments (Wooldridge 2002). In addition, we further tested the sensitivity of the model to perturbations of the parameters by removing hypothesis (H3) and adopting an alternative specification to control for the network effects: we added the number of POSs together with the number of ATMs (without normalization) among the covariates. The main results are confirmed and summarized in Table 4 on page 96 and Table 5 on the preceding page[36]

6 WHAT ABOUT THE OPTIMAL MULTILATERAL INTERCHANGE FEE? FIRST EVIDENCE FROM A NONLINEAR APPROACH

As discussed in our review of the literature, one relevant policy issue is concerned with the determination (or regulation) of the optimal level of interchange fee. For instance, according to Rochet and Tirole (2008), an MIF reduction may be translated as an increase in the cardholder fees with a negative net impact on the "social welfare function", which is a single-peaked function including different components (consumer surplus, retailers' profit and banks' profit). This means (Rochet and Tirole 2008) that the optimal interchange fee (MIF*) that maximizes the social welfare function may be nonzero (Figure 6 on the facing page).

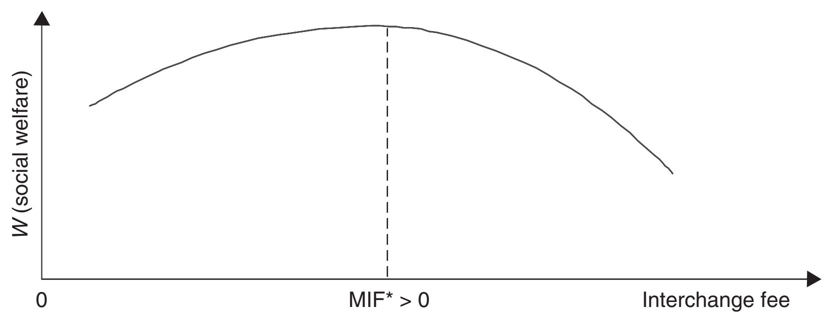
Since there was a risk that card payment schemes set excessively high interchange fees to increase banks' profits, the competition authorities started to limit exces-

[34] It is interesting also to note that the (negative) estimated coefficient for "Turnover" increases monotonically and considerably for the upper quantiles' regressions, suggesting that the income effect on the demand for cash is stronger for banks in the upper tail of the distribution.

[35] As in the standard FE estimations, the estimated coefficients of the control variables lose consistency in the IVFE specification, but we cannot exclude that this is due to their collinearity with the individual specific intercepts and the limited "within variability" in the biannual panel. We also excluded the interest rate in the IV regression models, to reduce endogeneity problems.

[36] The diagnostic tests for endogeneity and overidentifying restrictions in the GMM estimates (Table 4 and Table 5) are consistent with the ones obtain in the 2SLS estimates (Table 3). We also remove the "I rate" variable, which is not robust to the test for the orthogonality/exogeneity condition in the GMM model (Sargan). All these results are available upon request from the author.

FIGURE 6 Interchange fee and social welfare.



Source: Rochet and Tirole (2011).

sive interchange fees through the cost-based regulation or the implementation of the "tourist test", but always admitting positive levels of interchange fee in the system.[37]

Nevertheless, in recent years the debate has focused on the issue of whether the MIF on the modern card networks should be dismissed or maintained. Some economists claim that the interchange fee should be reduced to zero so as to remove direct and indirect regulatory costs to the market participants (Gans 2007) and to promote both competition and efficiency in retail payments, as compared to a situation with positive interchange fee (Leinonen 2009).

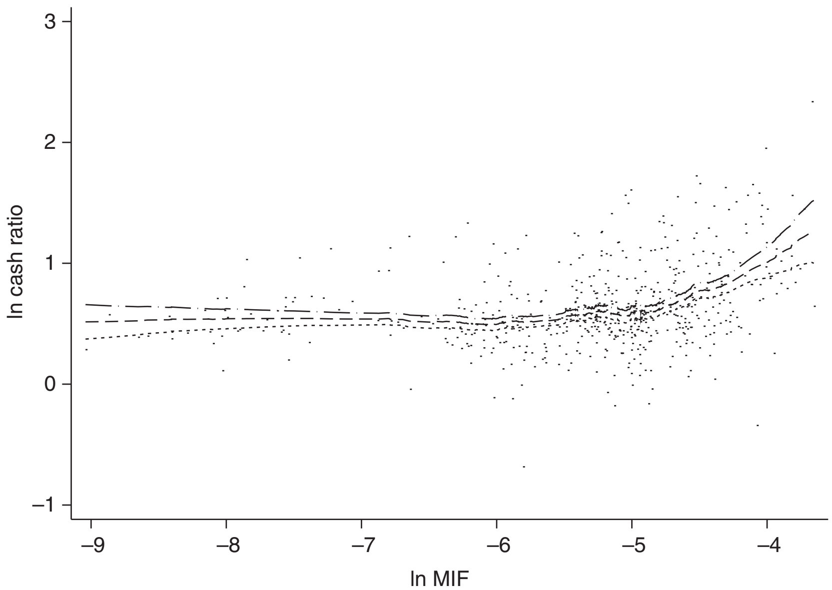
The theoretical problem of the optimal interchange fee determination is beyond the scope of the present paper. We can just find some clues through a more in-depth investigation of the relationship between the cash demand and the MIF according to our model analysis. Indeed, the cash–card ratio may be taken as a practical indicator for a social planner who is interested in shifting from cash to electronic payments.

In order to delve deeper into the effective impact of the interchange fee on the choice of payment instruments, as a first descriptive investigation we consider the multivariable scatter plot smoother (nonparametric approach),[38] which is a valuable

[37] Rochet (2007) states that "competition authorities often care only about user surplus and not about social welfare. This is justified if the profit of firms (banks) is completely dissipated. This is not justified if profit is reinvested to provide quality of service or attracts entry (lower prices, increased product variety)".

[38] The methods most appreciated when addressing the problem arising from nonlinear relationships are the nonparametric statistical techniques, which are robust to functional misspecification (ie, linear or polynomial). However, we leave this extension to a future research. In this case we adopted a nonparametric method called "k-nearest neighbours" or "k-nn", which is available through the software STATA 10 (Royston and Cox 2005). In this case, a simple estimator (corresponding to a uniform kernel) is to take the k observations nearest to x, and fit a linear regression of y_i on X_i

FIGURE 7 Multivariable scatter plot smoother.



Multivariate nearest-neighbour smoother; other explanatory variables contained in (4.1), including bank-specific and time dummy variables, are omitted in the figure. The upper and lower lines represent a pointwise confidence interval (95°) for the smoothed values of "cash ratio".

tool in exploratory data analysis (when the aim is to arrive at a parametric final model) and does not require us to specify functional forms for objects being estimated and strongly increases the goodness of the fit (Royston and Cox 2005). This graphical tool allows us to obtain a picture of the relationship between the "cash ratio" and MIF of (4.1) and each of the other explanatory variables simultaneously. The fitted (log) values of the cash–card ratio conditioned to the (log) average transaction-weighted MIF rate are shown in Figure 7.

Consistent with the findings reported in the previous sections, the MIF and the preference for cash withdrawal tend to be positively correlated. Nevertheless, this relationship is neither linear nor strictly monotonic: it is only beyond a certain level that the MIF impact is binding and the relationship with the cash–card ratio becomes strictly positive.

That being said, starting from Figure 7 we can try to estimate parametrically the nonlinear (second-order) effect of the MIF by adding a quadratic term in the original

---

using these observations. A smooth local linear k-nn estimator fits a weighted linear regression (see Hansen 2012).

TABLE 6 Estimation of the quadratic equation model (6.1).

	Model°					
	A	**B**	**C**	**D**	**E**	**F**
Regressor[1]	**PCSE[2]**	**FE[3]**	**RE[4]**	**IVRE[5]**	**OLS[6]**	**q50[7]**
MIF	0.1564***	0.0727***	0.1057***	0.4579***	0.1961***	0.1641***
	(0.0421)	(0.0253)	(0.0228)	(0.1304)	(0.0264)	(0.0248)
MIF2	0.0639***	0.0229*	0.0375***	0.1716***	0.0750***	0.0565***
	(0.0208)	(0.0121)	(0.011)	(0.0511)	(0.0126)	(0.0124)
Adj-R^2	0.8526	0.8953			0.2761	0.1708
Estimated turning points						
(ln) MIF*	−6.55	−6.91	−6.73	−6.66	−6.63	−6.78
MIF*	0.0014	0.0010	0.0012	0.0013	0.0013	0.0011

°"Full" model, MIF variable is log transformed and normalized on the sample mean; all other explanatory (Z_i) variables are omitted. All variables enter the regressions in logarithms. Significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Robust standard errors are in parentheses. [1]MIF variable is log transformed and normalized on the sample mean. Other explanatory variables (Z_i) contained in (6.1), including bank-specific and time dummy variables, are omitted in the table. [2]Panel corrected standard error estimator. [3]Fixed effect estimator. [4]Random effect estimator [5]Instrumented random effect estimator. Instrumented: MIF. Instruments: MIF_{t-1}, turnover_{t-1}, network_{t-1}. [6]OLS estimator. [7]Quantile regression (50° percentile).

equation model of (4.1), as follows:[39]

$$\text{cash ratio}_{it} = \alpha_0 + a_1\text{MIF}_{it} + a_2\text{MIF}^2 + \sum_h \alpha_h Z_{it} + u_{it}. \tag{6.1}$$

Whereas MIF and its squared term are strongly correlated ($r = |0.94|$), we have normalized the variable on its sample mean before taking its logarithm in order to reduce collinearity problems. Hence, we estimate (6.1) through the methods proposed in the previous sections.[40] For brevity, we just report estimated coefficients for the linear and quadratic MIF term (see (6.1)).

The outcomes in Table 6 show that the estimated regression coefficients for the first-order and second-order effects of the MIF are both positive and significant. This means that the cash ratio–MIF linear slope gets more positive as the MIF increases, the quadratic in MIF has a hump shape and the turning point (Wooldridge 2002) in

[39] We are indebted to an anonymous referee for this suggestion.

[40] We selected the following methods: ordinary least squares (OLS), quantile regression (q50), panel data random effect (RE) and fixed effect (FE), also controlling for autocorrelation and cross-sectional dependence (PCSE) and instrumental variable GMM random effect estimator to take into consideration endogeneity problems. As we show in Section 5.3 (see footnote 35 on page 98), also in this case we exclude the IVFE specification, because of a collinearity problem with the quadratic term and the individual specific intercepts and the limited "within variability" in the biannual panel.

the single peak response probability is equal to $[a_1]/(2a_2) > 0$.[41] Accordingly, the turning point, at which we have the minimum (fitted) value of cash ratio, is positive and different from zero MIF. This outcome seems to be consistent with the theoretical framework provided by Rochet and Tirole (2002, 2003). Although they do not consider the "cash–card ratio" explicitly in their model, our cash–card ratio may decrease until reaching the turning point. However, further empirical investigations are needed on this issue, such as in the case of the longer confidence interval for the smoothed values of "cash ratio" in Figure 7 on page 100, due to higher dispersion of the data at the lower tail of the MIF levels.[42]

Finally, the estimated turning point in the quadratic function (or "threshold level") is always much lower than the actual mean level of the MIF as reported in Table 6 on the preceding page and Table 1 on page 86, respectively. This means that the quadratic in MIF is positively skewed for a long stretch and that the positive linear trend, discussed in the previous section, is predominant.

7 CONCLUSION

The multilateral setting of interchange fees is usually justified because it reduces network transaction costs and increases payment card usage. Some economists and antitrust decisions also show how the interchange fee is an intermediary cost used as a reference price for the final fee charged to merchants (the so-called interchange fee plus mechanism). Nevertheless, few empirical works have measured the net impact of the interchange fee on the end user's decision to pay by card or withdraw cash.

Starting from our model of analysis, we consider the "cash–card ratio" as a practical reference indicator for a social planner who is interested in shifting from cash to electronic transactions. This would be consistent with the position expressed by several central banks, affirming that a strong shift to payment card transactions in lieu of cash would reduce the overall social cost of the payment industry and increase savings for payment service providers, firms and consumers. In Italy the cash–card ratio is higher than the average level in Europe. Thus, the main objective of this paper is to conduct a first empirical investigation on the impact of the interchange fee on the choice of cash withdrawals over card purchases for POS payments. This allows us to confirm that a regulation of the MIF level may be an effective tool in reducing the cash payments at the POSs.

Our results and robustness checks show that there is a positive relationship between cash usage and interchange fee mechanism, and that current mean MIF levels are still

---

[41] As the MIF variable is logarithmically transformed and normalized on the sample mean, the turning point is calculated as follows: $\exp[(a_1/2a_1) + \text{mean(MIF)}]$.
[42] This can be also confirmed by the parametric regressions.

too high to foster card acceptance in lieu of cash. Such a result is consistent with those obtained by Chakravorti *et al* (2009). Furthermore, from a first empirical investigation, such a relationship is not strictly monotonic, especially at the lower tails of the MIF level distribution.

Hence, we cannot also affirm that a zero MIF level would be optimal to increase electronic transactions, and we leave this topic for future research. We note that a relevant part of the economic literature perceives that a zero MIF rate may be counterproductive for the payment card diffusion and this may compromise the market incentive for innovation (Rochet 2007).

Finally, although the problem of the optimal interchange fee determination is beyond the scope of the present paper, our empirical model may inspire further research on the issue of interchange fee regulation. Indeed, our approach encourages investigation of the impact of different MIF rates on the substitutability between different means of payments, through relevant information collected on the issuing side. This payment data is usually more reliable and easy to obtain from the regulation authorities and our simple method of analysis may give some clues for how to enhance the efficiency of the payment system. For instance, policymakers may define some scenario analysis with the results of changes in the cash ratio as a result of changes in the average MIF.

REFERENCES

Alvarez, F., and Lippi. F. (2009). Financial innovation and transactions demand for cash. *Econometrica*, **77**(2), 363–402.

Ardizzi, G., and Coppola, G. (2002). The Italian case study: interchange fee, market structure and cost efficiency in the retail payment system. Conference on “The Economics of Payment Networks”. Working Paper, IDEI Université de Toulouse I.

Banca d’Italia (1998). Provision number 23 of 8 October 1998 – PAGOBANCOMAT.

Banca d’Italia (2012a). The social cost of payment instruments in Italy. *Institutional Issues* (November).

Banca d’Italia (2012b). *Annual Report 2011*.

Baxter, W. F. (1983). Bank interchange of transactional paper: legal and economic perspectives. *Journal of Law and Economics* **26**(3), 541–588.

Beck, N., and Katz, J. (2004). Time-series–cross-section issues. Working Paper.

Bolt, W., and Schmiedel, H. (2013). Pricing of payment cards, competition, and efficiency: a possible guide for SEPA. *Annals of Finance* **9**(1), 5–25.

Bolt, W., Humphrey, D., and Uittenbogaard, R. (2008). Transaction pricing and the adoption of electronic payments: a cross-country comparison. *International Journal of Central Banking* **4**(1), 89–123.

Bolt, W., Jonker, N., and van Renselaar, C. (2010). Incentives at the counter: an empirical analysis of surcharging card payments and payment behaviour in the Netherlands. *Journal of Banking and Finance* **34**(8), 1738–1744.

Börestam, A., and Schmiedel, H. (2011). Interchange fee in card payments. Occasional Paper, European Central Bank.

Breusch, T. S., and Pagan A. R. (1980). The Lagrange multiplier test and its application to model specifications in econometrics. *Review of Economic Studies* **47**(1), 239–253.

Cameron, C., and Trivedi, K. P. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

Chakravorti, S., Fernandez, F. R., and Valverde, S. C. (2009). Regulating two-sided markets: an empirical investigation. Working Paper, Federal Reserve Bank of Chicago.

Chang, H., Evans, D., and Garcia-Swartz, D. (2005). The effect of regulatory intervention in two-sided markets: an assessment of interchange-fee capping in Australia. *Review of Network Economics* **4**(4).

Chernozhukov, V., and Hansen, C. (2008). Instrumental variable quantile regression: a robust inference approach. *Journal of Econometrics* **142**(1), 379–398.

Coppola, G. (2011). Le surcharge sulle carte di pagamento. *Studi e Note di Economia* **16**(3), 323–340.

Doria, M. (2010). L'obiettivo di riduzione del contante: la facoltà di applicare "sconti" per l'acquisto di beni o servizi; il divieto di "surcharge". In *La Nuova Disciplina dei Servizi di Pagamento*. Giappichelli, Torino.

European Central Bank (2006). The Eurosystem's view of a SEPA for cards. Occasional Paper.

European Central Bank (2012). Payment statistics. Report incorporating 2011 data. ECB, September.

European Commission (2010). Decision relating to proceedings under Article 101 of the Treaty on the Functioning of the European Union and Article 53 of the EEA Agreement. Case COMP/39.398, Visan MIF.

European Commission (2012). Toward an integrated market for card, internet and mobile payments. Green Paper, COM(2011) 941.

Evans, D. S., and Schmalensee, R. L. (1999). *Paying with Plastic*. MIT Press, Cambridge, MA.

Gans, J. S. (2007). Evaluating the impact of the payment system reforms. Submission to the Reserve Bank of Australia's Payment System Board's 2007–08 Review of Payment System Reforms.

Gans, J. S., and King, S. P. (2003). The neutrality of interchange fees in payment systems. *Topics in Economic Analysis and Policy* **3**, Article 1.

Greene, W. (2000). *Econometric Analysis*. Prentice-Hall, Englewood Cliffs, NJ.

Hansen, B. (2012). *Econometrics*. University of Wisconsin.

Hayes, R. (2007). An econometric analysis of the impact of the RBAs credit card reforms. Mimeo, Melbourne Business School.

Hoechle, D. (2007). Robust standard errors for panel regressions with cross-sectional dependence. *Stata Journal* **7**(3), 281–312.

Humphrey, D. B., Pulley, L. B., and Vesala, J. M. (1996). Cash, paper, and electronic payments: a cross-country analysis. *Journal of Money, Credit and Banking* **8**(4, Part 2), 914–939.

ITM (2000). Study regarding the effects of the abolition of the non-discrimination rule in Sweden. European Commission Directorate General for Competition.

Jones, P., and Jones, C. (2006). Explaining differing European MCS levels. *Papeles de Economia Espanola*.

Jonker, N. (2011). Card acceptance and surcharging: the role of costs and competition. Working Paper 300, De Nederlandsche Bank.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis* **91**(1), 74–89.

Leinonen, H. (2007). On the efficiency of multilateral interchange fees (MIFs): how to price cash and cards in order to promote payment efficiency. Bank of Finland Online, no. 4. Financial Markets and Statistics.

Leinonen, H. (2011). Debit card interchange fees generally lead to cash-promoting cross-subsidisation. Bank of Finland Online, no. 3. Financial Markets and Statistics.

Lippi, F., and Secchi, A. (2009). Technological change and the households' demand for currency. *Journal of Monetary Economics* **56**, 222–230.

Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels. Working Paper, University of Cambridge.

Racine, J. S. (2008). Nonparametric econometrics: a primer. *Foundations and Trends in Econometrics* **3**(1), 1–88.

Rochet, J.-C. (2007). Understanding interchange fees. Mimeo, Toulouse School of Economics and IDEI.

Rochet, J.-C., and Tirole, J. (2002). Cooperation among competitors: some economics of payment card associations. *RAND Journal of Economics* **33**(4), 549–570.

Rochet, J.-C., and Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association* **1**(4), 990–1029.

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

Royston, P., and Cox, N. J. (2005). A multivariable scatterplot smoother. *Stata Journal* **5**(3), 405–412.

Schmiedel, H., Kostova, G., and Ruttenburg, W. (2012). The social and private cost of retail payments: a European perspective. Occasional Paper, European Central Bank.

Stata Technical Bulletin (1995). *Stata Technical Bulletin*, STB-25 (May).

Stix, H. (2004). How do debit cards affect cash demand? Survey data evidence. *Empirica* **31**(2–3), 93–115.

Visa International (2001). Visa International Service Association. Mimeo.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.

Zenger, H. (2011). Perfect surcharging and the tourist test interchange fee. *Journal of Banking and Finance* **35**(10), 2544–2546.