

Informational Synergies in Consumer Credit

Martin Hibbeln ^a, Lars Norden ^{b,*}, Piet Usselman ^c, Marc Gürtler ^c

^a *University of Duisburg-Essen, Lotharstraße 65, 47057 Duisburg, Germany*

^b *Brazilian School of Public and Business Administration, Getulio Vargas Foundation, Praia de Botafogo 190,
22250-900 Rio de Janeiro, Brazil*

^c *Braunschweig Institute of Technology, Abt-Jerusalem-Straße 7, 38106 Braunschweig, Germany*

Abstract

Lenders can tap into multiple sources of private information to assess consumer credit risk but little is known about the informational synergies between these sources. Using unique panel data on checking accounts and credit card accounts from the same customers during 2007-2014, we find that activity measures from both account types contain information beyond credit scores and other controls. Checking accounts display warning indications earlier and more accurately than credit card accounts. We also investigate the consistency of information, the reasons for defaults, and selection effects. The evidence highlights sizeable informational synergies that lenders can use to manage credit relationships.

Key words: Household finance, credit risk, lines of credit, credit cards, consumer bankruptcy

JEL classification: D10, G20, G21

* Corresponding author: Lars Norden; Brazilian School of Public and Business Administration (EBAPE), Getulio Vargas Foundation (FGV), Praia de Botafogo 190, 22250-900 Rio de Janeiro, Brazil. E-mail: lars.norden@fgv.br.

Parts of the paper were completed when Martin Hibbeln was visiting Rotterdam School of Management, Erasmus University.

The authors are grateful to Tobias Berg, Martin Götz, Oliver Entrop, Michael Haliassos, Steffen Meyer, Sascha Steffen, and participants at the Banking Workshop 2015 in Münster, the European Economic Association Conference 2015 in Mannheim, the Midwest Finance Association Conference 2015 in Chicago (Outstanding Paper Award in Financial Institutions), the German Finance Association Conference 2014 in Karlsruhe, and seminars at Aarhus University, Goethe University Frankfurt, the University of Münster, the University of Bremen, the University of Duisburg-Essen, the University of Giessen, the University of Innsbruck, the University of Leipzig, and the Braunschweig Institute of Technology.

Banks, credit card companies, and retailers apply specific lending technologies in appraising consumer credit that require information production. Lenders gather hard and soft information from public and private sources to assess the quality of their borrowers and the outcome of this assessment is aggregated to a credit score. There is consensus that a combination of different types of information is optimal for credit risk assessment. However, there is little evidence on the extent of synergies between different sources of private information in consumer finance, mainly due to the lack of data (Campbell 2006). Informational synergies affect the access to credit, terms of credit, credit allocation in the economy, and the risk taking of lenders.

We help fill this void by investigating whether activity measures from different credit products lead to informational synergies on the probability of default and credit line usage of consumers. Activity measures such as amplitude or overdrafts from credit products with time-varying exposure reflect private hard information that gives the lender a real time window into the borrower's cash in- and outflows. Our setting enables us to analyze why consumers default, how much credit they take (unconditional and conditional on default), and how early the lender learns about credit quality deterioration. We compare the account activity measures with traditional credit scores, borrower characteristics, and bank-borrower relationship characteristics, using combined information from checking accounts and credit card accounts from the same customers.

Researchers have increasingly paid attention to the link between account activity measures and information production. Mester, Nakamura, and Renault (2007) show that transaction information from the accounts of Canadian firms helps banks monitor small- and medium-sized enterprises (SMEs) in the context of asset-based lending. In a related study, Jimenez, Lopez, and Saurina (2009) investigate the determinants of the corporate credit line usage and exposure at default of Spanish firms. They find that the risk profile of the borrower, the bank, and the business cycle all have a significant impact on annual credit line usage. Norden and Weber

(2010) examine the credit line usage and checking account activity of different borrower types. They show that private hard information from checking accounts is particularly useful for monitoring small businesses and consumers. They also show how this information is used for loan pricing, credit limit management, and account closure decisions. Agarwal et al. (2009) document that credit card customers with close relationships to their bank exhibit lower probabilities of default and have higher credit line usage than non-relationship customers. Similarly, Puri, Rocholl, and Steffen (2013) find that consumers with prior bank relationships exhibit a significantly lower probability of default. Compared to Agarwal et al. (2009) and Puri, Rocholl, and Steffen (2013), in addition to considering the characteristics of prior bank relationships, we also investigate the informational synergies between the dynamic information from different credit products used by the same person. Stango and Zinman (2014) use consumer panel data for checking account activity and overdrafts from 2006 to 2008 to investigate how attention biases affect consumers' responses to financial shocks. Their analysis differs from ours in several dimensions. They take the consumer's perspective, consider overdrafts but not authorized lines of credit and/or credit card accounts, and examine attention biases, while we take the lender's perspective and examine how private hard information from consumers' checking accounts and credit card accounts can be used to manage credit relationships.

Vissing-Jorgensen's (2012) study is also related to ours but follows a different strategy. She considers consumer purchase data from a Mexican retail chain to investigate whether the type of products a consumer buys provides information about his/her default risk. In her setting, consumer purchases are financed with consumer loans provided by the retailer. All purchases are financed with the same credit product. She shows that products that lead to high losses on consumer loans tend to be luxury goods purchased by consumers who buy abnormally large fractions of luxury goods given their income. She documents a link between consumer choice

and the default risk of consumer loans, whereas we investigate potential synergies between private information from different credit products used by the same consumer.

Our study further relates to the cross-selling of financial products, which can be beneficial in various ways. For instance, the marketing costs are lower, switching costs for customers are higher, and banks can learn more about their customers' risk preferences and consumption behavior (e.g., Kamakura et al. 2003). Customers with more valuable information for cross-selling purposes may pay lower prices (e.g., Akçura and Srinivasan 2005). Studies on the cross-selling of financial products have mainly focused on the question of which product a customer is most likely to buy next in order to offer the right product to the right customer at the right time (e.g., Kamakura et al. 1991; Li, Sun and Wilcox 2005; Li, Sun and Montgomery 2011). In our study, the focus is on informational synergies that can be beneficial for credit risk management and customer relationship management.

We base our analysis on a large and unique panel dataset for the 2007-2014 period comprising more than 1.7 million customer-month observations with checking account and credit card observations for every customer. This dataset contains information from all checking accounts and credit card accounts of a large privately owned financial institution in Germany. The setting is ideal to study potential synergies between different sources of private information because universal banking is the dominant form of financial intermediation in Germany and it heavily relies on informational synergies over time and across products (e.g., Allen and Gale 2000; Krahnen and Schmidt 2004). The widespread use and simultaneous presence of consumer checking accounts with lines of credit attached to these accounts¹ and credit card accounts enables us to examine informational synergies. Credit cards are widely used in the United States but consumer checking accounts with lines of credit are not common.

¹ It is one account that may have a positive or negative balance. The negative balance has to stay below the authorized external limit. Overdrafts are bounced by default but can be authorized manually by loan officers. Such consumer checking accounts are widespread in Europe, Canada, and many Asian countries.

We obtain three principal results. First, we show that the activity measures from both credit products contain information beyond traditional credit scores, borrower characteristics, and bank-borrower relationship characteristics. Lenders can use this information for monitoring and screening. Second, we find that account activity measures from checking accounts are more useful regarding additional cross-product information than those from credit cards. The main explanation is that the former reflect cash in- *and* outflows, while the latter reflect *only* cash outflows. The benefits of cross-production informational synergies are economically sizeable: Lenders can lower the type I error by around eight percentage points when they consider checking account activity next to credit card account activity, corresponding to a 33% reduction of the type I error. Third, we differentiate between variables that affect the probability of default but not the credit line usage and vice versa to show why and how consumers default. We further document significant selection effects in the estimation of credit line usage at default. We conduct several additional empirical checks and show that our previous results are robust and not the product of particular choices of samples or model specifications.

Our study provides novel evidence on the time-varying nature of consumer credit risk and shows how lenders can exploit cross-product informational synergies. Our findings on informational synergies - next to the evidence on cost synergies through core deposit funding (Berlin and Mester 1999) - speak in favor of the universal bank model. Lending and risk assessment should be carried out by the same institution because lending itself generates proprietary real time information that a non-lender cannot produce (e.g., a credit bureau or rating agency). Moreover, our findings hint at sizeable economies of scope in the information production of banks. For instance, rating agencies cannot realize the same synergies as banks but they can take advantage of economies of scale (i.e., selling the same rating service for a larger bond or several bonds from the same issuer).

Our contribution to the literature on household finance and banking is threefold. First, we study the incremental value added of different sources of private information, while earlier studies have examined the value added of private information versus public information. Lenders can take advantage of private information only if the information production is not distorted by incentive problems arising from volume-based compensation schemes, career concerns, etc. (e.g., Berg, Puri and Rocholl 2014; Hertzberg, Liberti and Paravisini 2010). A risk assessment based on multiple sources of private information is less likely to be biased by strategic behavior of the borrower and/or loan officer. It is unlikely that a borrower or loan officer can manage to hide the true credit quality by manipulating the checking account and credit card activity. Understanding of how lenders gather, aggregate, and exploit private information to manage credit risk and their customer relationships is at the core of financial intermediation.

Second, we focus on private hard information, which has been neglected in earlier research due to the lack of comprehensive data. Consumer credit is characterized by a large number of transactions, small amounts, and standardized products and processes. The key products are credit cards, consumer loans, and mortgage lending. Lenders mainly rely on standardized hard information and this reliance is even stronger when consumer debt is securitized (Rajan, Seru, and Vig 2015). Hence, it is important to understand the potential synergies between different sources of private hard information.

Third, we focus on a subset of consumer credit products: credit cards and checking accounts bundled with lines of credit. Most of the recent credit-related research in household finance focuses on mortgage lending, while there has been little attention to credit cards and/or checking accounts (some exceptions are, for example, Stango and Zinman 2014; Norden and Weber 2010). Unlike mortgages, the latter credit products exhibit a time-varying risk exposure, no maturity, and no collateral. Monitoring their time-varying credit exposure is critical for the lender.

The remainder of this paper is organized as follows. In Section 1, we describe the institutional background, the data, and the methodology. In Section 2, we report the results of the analysis of informational synergies. In Section 3, we summarize the findings from further empirical checks and tests of robustness. We conclude in Section 4.

1. Institutional Background, Data, and Methodology

1.1 Institutional background

We consider proprietary data from a large privately owned German financial institution.² Checking accounts are used for receipts (e.g., salary, rental income, etc.), and expenses (e.g., rent, subscriptions, insurance, consumption expenses). The balance of a checking account can be positive or negative because by default a credit line is attached to it. If the balance is positive, consumers (may) receive a small amount of interest from the bank, and if the balance is negative, consumers have to pay interest to the bank for using the credit line. The credit line is pre-authorized up to a specific limit and can be redeemed at any point in time. Consumers use a checking account to withdraw cash at ATMs, to make payments with debit cards, or use electronic direct debits or paying their bills with electronic wire transfers. It is important that no brokerage is offered by the bank so that we can exclude that checking accounts are used as clearing accounts for the brokerage accounts.

We also consider credit card accounts (Visa credit cards). The balance of an account can be positive or negative. Our dataset comprises two types of credit card accounts, which differ in their redemption schemes. On the one hand, there are accounts with full repayment of the credit card bill every month. This monthly bill is charged on the individual's checking account on a fixed day per month and there is no interest to be paid. The payment day is one of the first days in

² We do not know the identity of the account holders. To ensure confidentiality the customer numbers and account numbers were anonymized by the financial institution before providing us with the data.

a month for almost all customers. On the other hand, there are accounts that allow the consumer to stretch out the redemption over time and make only minimum repayments per month. These monthly minimum repayments correspond to 5% of the outstanding credit card debt and at least 50 euro. Customers pay interest rates of more than 10% p.a. on the outstanding debt.

Checking accounts and credit card accounts are unsecured credit. Moreover, a consumer can open a credit card (checking) account without having a checking (credit card) account at the same bank. All accounts have an initial line of credit of 1,000 € and consumers may ask for an increase of the limit later on. A negative account balance has to stay within the limit but we also observe overdrafts. Accounts can be overdrawn by using the checking account or the credit card offline, manual approvals by loan officers, or interest debit charged by the bank.

The presence of these credit products is widespread in Europe, Canada, and many Asian countries. In 2012, more than 99.5% of the main payment instruments in Germany were credit transfers, direct debits, electronic cards, and credit cards. The same finding holds for most countries being part of the Committee on Payment and Settlement Systems (CPSS) (Bank for International Settlements, 2013). The setting differs from the U.S. where consumers typically use one account for receipts (a checking account with positive balance) and another account for expenses (a credit card account with negative balance).

We further use internal bank information on borrower defaults. According to the default definition, complying with Basel III and EU Regulation No. 575, a borrower default can occur if the borrower is “90 days past due” or if the bank expects that the person will not pay back all of his obligations. The latter occurs if the bank observes a limit violation, a bankruptcy, or receives negative credit information from a credit bureau (e.g., SCHUFA in Germany). The default status refers to the account and not to the account owner. This is common bank practice for retail products. A default of one account does not necessarily lead to a default of other accounts of the

same person. This practice is in line with the regulatory rules for retail exposures according to Basel III.

1.2 Data

We base our analysis on dynamic information from checking accounts and credit card accounts. Our proprietary raw data sample comprises 16,849,187 account-months observations. The sample period is from December 2007 to January 2014.

We apply five filter rules on the raw data. First, we choose only customers who have credit cards and checking accounts at the same bank. More than 90% of all customers with checking accounts also have a credit card, but only 22% of all credit card customers also have a checking account at the same bank. This reduces the dataset to 5,958,534 observations. Second, in our baseline analyses, we consider only one credit card and one checking account per customer. Most of the customers with both products have exactly one checking account and one credit card (88%), and about 10% use one credit card and two checking accounts or vice versa, so that we neglect 237,880 observations. For the 12% of customers where we observe more than one checking account or credit card, we focus on the most important account.³ Third, we drop observations with missing data (1,158,501 observations), which is mostly a consequence of a limit equal to zero in the denominator of account activity variables. Fourth, we only consider observations where the subsequent observation period regarding defaults or credit line usage is at least one year,⁴ which reduces the dataset by 967,795 observations. Fifth, we use only observations until the first default occurs, so that the observed account activity is not influenced

³ We assume that the account with the highest mean cash inflow per month is the most important one because it is likely that the customer's salary is paid to this account. For credit card accounts, inflows normally have to equal the outflows; hence, the most important credit card can also be identified using the highest mean inflow. The number of customers where we have to identify the most important account is rather low; hence, the choice of the procedure does not affect our analysis. In a robustness check, we repeat all subsequent analyses with customers who have only one checking account and one credit card. Our results show that the aggregation rule does not affect our results.

⁴ If the observation period is shorter than the forecasting horizon, the credit risk can be underestimated (Gürtler and Hibbeln 2013).

by a previous default of the same account, reducing the dataset by 33,594 observations. We winsorize all variables at the 1%-/99% quantile. The final sample consists of 3,560,764 account-month observations of 46,925 customers.

(Insert Table 1 about here)

Panel A of Table 1 reports the number of observations and defaults. There are 1,639 and 2,101 defaults at checking accounts and credit cards, respectively. Customers are more frequently in default on credit cards than on checking accounts. This result is not due to differences between customers because the checking accounts and credit cards come from the *same* customers. Moreover, people can default on multiple accounts or on individual accounts. The reasons for a joint default are bankruptcy or negative credit information from a credit bureau. A default on an individual account can occur if the monthly expenses sufficiently exceed monthly receipts. Similarly, a customer can default on a credit card account. A high credit card balance might lead to an overdraft of the checking account, so that the debit is bounced. A bounced debit does not cause a default on the checking account, but it might do on the credit card account.

Panel B of Table 1 shows that about 50% of the monthly default observations (13,887 observations) are joint defaults of both accounts of the same person, whereas the remaining 50% occur either on the checking account (5,857 observations) or on the credit card account (7,565). These numbers confirm that both cases occur frequently and are therefore relevant for our study.

Panel C of Table 1 reports summary statistics of the main variables. We define the variable *Rating* as the logarithm of the probability of default according to the internal rating system of the bank. Because this rating is account specific, different rating systems are used for checking accounts and credit cards (the correlation between the corresponding credit scores per customer is 0.337). The mean *Rating* for checking accounts is significantly better than for credit card accounts, which is consistent with a lower number of defaults for checking accounts.

Net Inflow/Lim is the difference between monthly cash inflows and outflows as a percentage

of the external limit. To measure the account variation, we use *Amplitude* as the difference between the maximum and minimum exposure in each month as a percentage of *Limit*. We observe higher amplitude for checking accounts than for credit card accounts as checking accounts are used for income and expenses. The mean and median of the external limit in euros indicates that most customers choose a higher limit than the initial limit; moreover, most customers choose a similar limit for both products (the correlation between the limit of both accounts is 0.767). *Bounced* is the average number of bounced debits in the preceding 12 months (per month). In most cases, this number is zero for checking accounts, whereas for credit cards bounced debits can be observed more frequently. *Days Usage* is the percentage of days with a negative balance in the preceding 12 months. This number is significantly higher for credit cards because for this product we usually observe no credit balance. We see a relatively active use of credit cards since 55% of the days in the preceding 12 months have a negative balance. Similar to *Days Usage*, we define *Days Overdrafts* as the percentage of days with overdrafts in the preceding 12 months. Only a very small percentage of checking accounts exhibit overdrafts.

Moreover, we have information on bank-customer relationship characteristics. *Duration* indicates the length of the relationship in months measured separately for each account. For checking accounts, *Duration* is on average 46 months (about four years), which is slightly longer than for credit card accounts. Seven percent of the credit card observations belong to accounts without full monthly repayment (*Full Payment* = 0). Around 38% of observations with defaults belong to this product type, which indicates that higher risk customers self-select into this redemption mode. We therefore use the redemption mode (*Full Payment*) as control variable. We also use information about the customer's age, gender, job, marital status, number of children, nationality, online versus offline banking, and academic degrees (not reported in Table 1).

1.3 Methodology

We analyze informational synergies on the probability of default (PD), the credit line usage (CLU), and the credit line usage at default ($CLUAD$) of checking accounts and credit card accounts. We use these variables as dependent variables and the variables described beforehand as explanatory variables.

To estimate the PD of one of these accounts, we first consider the account activity information from the same account. Second, we study the informational content of cross-product information. For this purpose, we estimate the PD of checking accounts based on activity variables from credit card accounts and vice versa. Third, we take account activity from both accounts at the same time to estimate the PD . We follow the same three steps to estimate the CLU and $CLUAD$ of checking accounts and credit card accounts.⁵

We define PD as the probability of default of an account within a time horizon of one year. Thus, for a given month t , we define a default variable $Def_{i,t+12}$ that equals one if a jump to default is observed at $t+\tau$ with $\tau \in \{1, 2, \dots, 12 \text{ months}\}$ for account i , and zero otherwise. Given the explanatory variables $Z_{i,t}$ that can be observed in month t , the estimated PD is:

$$PD_{i,t} := P(Def_{i,t+12} = 1 | Z_{i,t}) = f(Z_{i,t}). \quad (1)$$

The explanatory variables contain account activity variables, which can include information from the particular account only, or we consider cross-product information as well. Furthermore, $Z_{i,t}$ includes bank-borrower relationship and borrower-specific variables. In addition, we use the internal credit scores to account for variables that can be observed by the bank but are not available to us (e.g., information from a credit bureau). We integrate the credit score ($=PD_{i,t}$) of

⁵ We also estimate the PD and the CLU at default on the full dataset of all 12.4 million account-month observations instead of considering 3.5 million observations where we have information about checking accounts and credit cards of the same customer. We obtain similar estimation results regarding the sign of coefficients, statistical significance, and adjusted R^2 , which confirms that restricting the analysis to the matched sample of checking account and credit cards from the same customer does not introduce a selection bias.

the internal rating as $Rating_{i,t} = \log(PD_{i,t})$. We perform pooled probit regressions and cluster the standard errors at the customer level.

To estimate the credit line usage (CLU), we choose a time horizon of one year. The CLU of account i in month $t+12$ can be defined as $CLU_{i,t+12} = Exposure_{i,t+12}/Limit_{i,t+12}$. However, we would have to know the limit in $t+12$ to calculate the expected exposure in $t+12$ based on the estimated CLU . Thus, for predicting the credit line usage, we calculate the target variable $CLU_{i,t+12}$ as the ratio of the (at time t unknown) exposure in $t+12$ months and the (known) limit in t : $CLU_{i,t+12} = Exposure_{i,t+12}/Limit_{i,t}$. To estimate CLU , we apply:

$$E(CLU_{i,t+12} | X_{i,t}) = X_{i,t} \beta, \quad (1)$$

where the explanatory variables $X_{i,t}$ includes account activity with or without cross-product information, as well as bank-borrower relationship and borrower characteristics. Following Jimenez, Lopez, and Saurina (2009), we add the credit rating ($Rating$) to equation (1) because it is likely that the CLU depends on the customer's default risk. To estimate the CLU , we perform ordinary least squares (OLS) regressions with standard errors clustered at the customer level. We refer to this method as CLU_{all} .

We further examine informational synergies regarding the credit line usage at default ($CLUAD$). We implement different methods to estimate the $CLUAD$ because of possible biases. First, we use CLU_{all} to estimate the $CLUAD$, but the estimates can be biased because the rating does not fully capture differences in the credit line usage between defaulters and non-defaulters: $(CLU_{i,t+12} | Def_{i,t+12} = 1, X_{it}) \neq E(CLU_{i,t+12} | Rating_{i,t}, X_{it})$. Second, we calculate equation (1) on the subsample of defaulted loans. We call this method CLU_{Def} .⁶

$$E(CLU_{i,t+12} | Def_{i,t+12} = 1, X_{i,t}) = X_{i,t} \beta. \quad (2)$$

⁶ Discussions with bankers confirm that this approach is a common banking practice.

Third, we use a Heckman selection model (see Heckman, 1976, 1979) to account for a possible selection bias. We refer to this method as $CLU_{Heckman}$:

$$E(CLU_{i,t+12} | Def_{i,t+12} = 1, X_{i,t}, Z_{i,t}) = X_{i,t}\beta + \delta_{\lambda}\lambda(Z_{i,t}\gamma), \quad (3)$$

where the inverse Mills ratio λ is determined by a probit regression. Fourth, we consider an endogenous treatment-regression model (see Heckman, 1978, 1979) of the following form (unrestricted control function approach):

$$E(CLU_{i,t+12} | Def_{i,t+12}, X_{i,t}, Z_{i,t}) = X_{i,t}\beta + \nu Def_{i,t+12} + (Def_{i,t+12} X_{i,t})\eta + \delta_{h1}(1 - Def_{i,t+12})h_1(Z_{i,t}\gamma_1) + \delta_{h2}Def_{i,t+12}h_2(Z_{i,t}\gamma_2). \quad (4)$$

The hazards h_1 and h_2 are determined by probit regressions of the first stage participation model. We include interaction terms between $Def_{i,t+12}$ and all explanatory variables $X_{i,t}$ to allow for different intercepts and slopes for defaulters and non-defaulters. We refer to this estimator as $CLU_{End.Treat}$. The results would be identical to an IV regression if we implemented the restrictions $\delta_{h1} = \delta_{h2}$ and $\beta = \eta$, and if we used a linear probability model instead of a probit model for the probability of being selected (e.g., Vella and Verbeek 1999).

2. Results

2.1 Defaults

We conduct a univariate event study of the balance and limit of checking accounts for defaulted and non-defaulted customers. We show the explanatory variables from 24 months before default to 12 months after default; the default time is $\tau = 0$. We calculate the median of the explanatory variables for defaulted customers at a monthly frequency for a time window of 37 months ($\tau - 24, \tau - 23, \dots, \tau, \dots, \tau + 12$). For non-defaulted customers, we calculate default weighted medians to allow a direct comparison between the explanatory variables for defaulted and non-defaulted customers. We first calculate the median account variable for each month ($t = 12/2007$

to 1/2014) for all non-defaulted customers that will not default in the next 12 months ($Def_{i,t+12} = 0$). Then, we determine the number of defaults in each month and calculate on this basis how often each month must be considered at each event time ($\tau-24, \tau-23, \dots, \tau, \dots, \tau+12$) and compute the median of the monthly medians for each event time.

Figure 1 shows both variables for defaulted and non-defaulted checking accounts. On the x-axis, the time relative to the default event at τ is displayed in months, and on the y-axis the median *Limit* and the median of $(-1) \times \text{Balance}$ are shown in euros. For defaulted accounts, we see that the negative *Balance* is already increasing around 14 months before default (analogous to the credit line usage discussed below). The negative *Balance* is slightly decreasing after default because the *Limit* at the time of default is often set to zero and no further increase in exposure is possible. *Limit* is typically reduced to zero in the month of default. Default is often due to overdrafts, which is supported by the observation that *Limit* approximately equals the negative *Balance* one month before default. All findings hold for the 10% and 90% quantiles, as well as for credit card accounts (not reported). For non-defaulted customers, *Limit* (positive) and *Balance* are usually higher. This difference between non-defaulted and defaulted accounts can already be observed 24 months prior to default and increases as the time to default approaches. From Figure 1, we cannot see whether the credit line usage is increasing due to decreasing inflows (e.g., unemployment) or increasing outflows (e.g., higher consumption or unexpected expenses).

(Insert Figure 1 about here)

In Figure 2, we show cash inflows and outflows separately for defaulted and non-defaulted checking accounts. The gap between cash inflows and outflows, the *Net Inflow*, directly leads to the change of *Balance* shown in Figure 1. These inflows and outflows, as well as the *Net Inflow* are more informative for checking accounts than for credit card accounts. For credit cards only,

the outflows are influenced by the customer, while the inflows correspond to the sum of the outflows because of monthly clearing.

(Insert Figure 2 about here)

Figure 2 shows that the inflows and outflows of non-defaulted checking accounts are similar and rather stable, implying that *Net Inflow* is close to zero. For customers who default, inflows decrease earlier and faster than outflows. This inflow decrease starts about six months before default and continues until default. Outflows, on the other hand, only slightly decrease, but around three months before default, they decrease rapidly. The latter is mainly a consequence of the customers approaching the limit, as shown in Figure 1. The findings indicate that for most customers, a default on a checking account is due to decreasing cash inflows, not increasing cash outflows. Hence, the *Net Inflow* is negative for defaulted accounts and increases in the months before default; however, at the time of default, the inflows and outflows are almost identical; after default, inflows are even greater than outflows, which is a consequence of customers being constrained by the limit reduction at the time of default.

Building on the studies of Mester, Nakamura, and Renault (2007) and Norden and Weber (2010), we consider limit violations as a predictor for customer defaults. However, unlike earlier studies, we use dynamic private information from different sources at a daily frequency. We observe the number of days with a positive usage of the credit line and with overdrafts in each month. These two numbers are very informative for predicting defaults and hint at substantial informational synergies across different credit products.

In Figure 3, we present the median number of days with positive credit line usage (Panels A1 and B1), as well as the median number of days with overdrafts (Panels A2 and B2) for defaulted and non-defaulted accounts. In Panel A, we present these variables for defaults of checking accounts, whereas Panel B refers to defaults of credit cards.

(Insert Figure 3 about here)

Panel A1 of Figure 3 shows for checking accounts that the number of days with positive credit line usage is higher for accounts with subsequent default. This difference is visible 24 months prior to default, but the difference increase as default approaches. We also show the number of days with positive credit line usage of credit cards prior to a default of the checking account. This number slightly increases before default too, meaning that cross-product information is informative. Panel A2 indicates that checking account overdrafts strongly increase in the three months prior to default, peaking at the default event. The number of days with overdrafts is much higher for defaulted checking accounts than for non-defaulted accounts, which confirms that overdrafts provide useful early warning indications. Interestingly, the number of days with overdrafts on credit card accounts is also substantially increasing in the period prior to default of checking accounts. Thus, account activity from other credit products is indeed informative for default prediction.

Panel B1 and B2 of Figure 3 show the number of days with positive credit line usage and overdrafts regarding the default events of credit card accounts. The findings for Panel B1 are similar to the defaults of checking accounts (Panel A1). The median number of days with positive usage is lower for defaulted credit card accounts than for non-defaulted accounts. Panel B2 indicates that overdrafts on credit card accounts strongly increase four months prior to default with a peak at default, while increasing overdrafts on checking accounts of the same customers can be observed substantially earlier (Panel A2).

We continue with the analysis of the probability of default within a time horizon of one year. We examine at time t whether a default event will occur in the next year, which means in $\{t+1, t+2, \dots, t+12\}$. We analyze additional variables that may influence the probability of default (next to *Net Inflow*, *Days Usage*, and *Days Overdrafts*)⁷ and report the average differences between

⁷ *Days Usage* and *Days Overdrafts* are the percentage of days in the preceding 12 months, whereas we present in Figure 3 the median number of days per month with a negative balance or days with overdrafts.

defaults and non-defaults in Table 2. Columns (1) and (2) refer to checking account defaults, whereas columns (3) and (4) refer to credit card account defaults. The results without cross-product information are shown in columns (1) and (3); the results for cross-product information are shown in columns (2) and (4).

(Insert Table 2 about here)

Table 2 shows that defaulted checking and credit card accounts have significantly worse ratings, meaning that the internal ratings of banks are informative for the default risk of customers. We find different values for *Rating* depending on the information source due to different weights of account and customer characteristics in the scoring function. Moreover, defaulted checking accounts have lower *Net Inflows*, *Amplitude*, and *Limit*, along with a higher number of bounced credits (*Bounced*), a more frequent usage of credit line (*Days Usage*), more frequent *Days Overdrafts*, and a shorter relationship (*Duration*). As an additional variable to evaluate the bank-customer relationship, we use the dummy variable *Prev. Def.* to measure whether a previous default occurred on a different account of the same customer. For defaulted checking accounts, the percentage of customers with a previous default on a credit card account is significantly higher, which shows that previous defaults on other accounts can be used as early warning signals. The overall findings for credit card accounts are similar except for *Amplitude*, where *Amplitude* is higher for defaulted accounts. There is a significantly higher percentage of customers without full monthly payments for defaulted accounts than for non-defaulted accounts. The cross-product analysis results in columns (2) and (4) show that checking or credit card account defaults are due to information from both credit products.

We use multivariate probit regression models to estimate the probability of default.⁸ The results are reported in Table 3. Columns (1)-(4) show the results for checking account defaults and columns (5)-(8) for credit card defaults. The upper half of the table refers to checking account variables and the lower half to credit card variables. We first discuss checking account defaults. In the probit regression (1), we use *Rating* as independent variable to predict defaults in the next 12 months, showing that the rating system of the bank is effective. Column (2) confirms that account activity variables are useful early warning indicators for default risk, with the coefficients of all variables being identical to the univariate findings. When the estimation is performed on account activity variables only, we obtain an adjusted R^2 of 0.232 for checking accounts (not reported), which shows that these activity variables are even more informative than the bank's current internal rating system. The main findings remain stable if we additionally control for *Rating* in column (3). The adjusted R^2 slightly increases when the internal rating of the bank is added, which indicates that there is additional information contained in the rating system beyond our measures of default risk. Column (4) reports the results for cross-product information. The influence of rating, account activity, and bank-customer relationship variables of credit cards on checking account defaults mostly points in the same direction as the impact of the corresponding checking account variables, which means that the information is widely consistent. We find that most of these variables are highly significant. Moreover, the explanatory power is higher if cross-product information is considered.

(Insert Table 3 about here)

Columns (5)-(8) of Table 3 report the results of credit card defaults. The results are similar to checking accounts except for *Amplitude*, as a higher amplitude on checking accounts is a positive signal but on credit cards it is rather a negative signal, confirming the above univariate results. As

⁸ We also use logistic and rare events logistic regression models (King and Zeng 2001). The choice of the model does not affect our results. Moreover, we perform all subsequent analyses with absolute account activity measures instead of using account activity ratios. We find that all the main results are robust.

discussed above, a likely reason is that high activity on a credit card is mainly due to consumption, whereas for checking account this could also be due to a high income. Furthermore, regarding cross-product information on credit card defaults, we confirm that the information is consistent (column (8)). The increase in the adjusted R^2 from column (7) to column (8) shows that the prediction of credit card defaults can be improved if account activity from checking accounts is considered as well.

We find evidence that banks can achieve substantial informational synergies when they monitor customer account activity for different products. The additional cross-product information helps improve the prediction of defaults beyond traditional credit scores, customer characteristics, and bank-customer relationship characteristics.

2.2 Credit line usage

We now estimate the OLS regression CLU_{all} to examine which variables at time t influence the credit line usage in $t+12$ months. Table 4 reports the OLS regression results. As for the PD regression, columns (1)-(4) show the findings for checking accounts and columns (5)-(8) the findings for credit cards. Column (1) shows that *Rating* is not only predictive of future defaults but also of the CLU in 12 months. Similarly, column (2) shows that the account activity variables are informative about the future CLU . We find that CLU is high if *Net Inflow* is low or more overdrafts are observed. If rating information is included (column (3)), the coefficients of *Days Usage* and *Days Overdrafts* are significantly decreased, which is mainly due to a high positive correlation between *Rating* and *Days Usage* ($\rho = 0.31$), as well as *Rating* and *Overdrafts* ($\rho = 0.54$). Customers with a higher credit line usage and more frequent overdrafts have worse ratings, so that this effect is contained in the coefficient of the rating. Adding information from credit cards to the regression in column (4) only slightly improves the estimation of CLU .

(Insert Table 4 about here)

Columns (5)-(8) in Table 4 report the results of analyses of credit card defaults. The results are mostly similar to those for checking accounts shown in columns (1)-(4). Cross-product information is not significantly predictive of credit line usage, which is in contrast to the findings on the prediction of default from Table 3.

2.3 Credit line usage at default

Banks have to estimate the credit line usage at default (or the absolute value of exposure at default) for credit risk management and due to regulatory capital requirements. There are several possible approaches to estimate the credit line usage at default, some leading to biased estimates. We consider this problem and propose an approach that leads to unbiased estimates.

The rating and the default status are key inputs for modeling CLU at default. Figure 4 presents the median CLU for checking accounts between $\tau-24$ and $\tau+12$ by rating. The rating at $\tau-12$ discriminates between high and low CLU prior to default. The better the rating, the lower the CLU in the period between $\tau-24$ and τ . This indicates that *Rating* is informative about the CLU at default. Unreported results indicate that the results for credit card accounts are similar.

(Insert Figure 4 about here)

In Table 5 we compare the different methods we described in Section 1.3 to estimate the CLU at default for checking accounts.

(Insert Table 5 about here)

Columns (1) and (2) of Table 5 repeat the results for CLU_{All} of Table 4 to allow a comparison to the other methods. Columns (3) and (4) report the results for CLU_{Def} , which is the estimate based on the subsample of defaulted accounts. A comparison of columns (3) and (4) with the results in columns (1) and (2) indicates that the estimation on the full sample is influenced by the variation of CLU of non-defaulted accounts even if controlling for rating. For instance, CLU_{Def} is

almost not affected by *Amplitude* or *Limit*, whereas both variables are highly significant in the results for CLU_{All} . Furthermore, the *Rating* effect in the CLU_{Def} model is much smaller than for CLU_{All} , which means that conditional on default the impact of *Rating* is smaller. This finding is consistent with Figure 4, where we see the interdependence between rating categories and CLU . However, comparing the values at τ and $\tau-12$, this interdependence becomes weaker at default.

Importantly, the models shown in columns (3) and (4) of Table 5 might suffer from a sample selection bias, as discussed in Section 1.3. The estimator can be biased if there are unobserved effects in the model CLU_{Def} that also influence the likelihood of being in the subsample of defaults (selection). We therefore show the results from the Heckman selection model $CLU_{Heckman}$ in column 5 of Table 5. We find evidence that the results of CLU_{Def} (column (3)) are affected by a sample selection bias as the coefficient of the inverse Mills ratio (λ) is statistically significant. Interestingly, we observe almost the same coefficients in columns (4) and (5). The reason for this finding is that *Rating* and the inverse Mills ratios of the Heckman selection model are highly correlated. Thus, unbiased estimates of CLU at default can also be obtained by using a standard OLS regression based on the subsample of defaulted accounts, provided that *Rating* is considered as a covariate. In contrast, the results of CLU_{All} and $CLU_{Heckman}$ are different, which confirms that considering *Rating* as a covariate in a full sample regression is insufficient to explain the difference between the CLU of defaulted and non-defaulted accounts. Even if CLU_{All} does not suffer from the sample selection bias (column (3)), the results confirm that the model is not appropriate to estimate CLU conditional on default (also see equation 3 in Section 1.3).

Furthermore, we examine the differences of CLU for defaulted and non-defaulted accounts with otherwise identical characteristics. We thus report estimation results from the endogenous treatment-regression model $CLU_{End.Treat}$ in column (6) in Table 5. Due to the specification of the $CLU_{End.Treat}$ model presented in Section 2.3, the results for the default subsample are identical to the standard Heckman selection model $CLU_{Heckman}$. We report only the estimates for CLU

conditional on *non*-default to conserve space. These coefficients are similar to those in column (1), which implies that the regression model CLU_{All} provides estimates of CLU of non-defaulting customers. The difference between the scenarios “ CLU at default” versus “ CLU at non-default” is represented by the non-reported interaction terms $Def. * Acc.-controls$ and $Def. * Rel.-controls$; these coefficients can be calculated by subtracting the values of columns (5) and (6). As expected, the high coefficient of *Default* indicates that CLU is significantly higher for defaulted than for non-defaulted accounts. We find that a higher number of *Bounced* credits and *Overdrafts* reduces CLU for non-defaulted accounts. One interpretation is that for some customers, *Bounced* credits or *Overdrafts* have a disciplining effect, so that they react with a reduction in their credit line usage. In contrast, for the subset of defaults, a high number of *Bounced* and *Overdrafts* indicates that their credit line usage is likely to be high at default. There can be a substantial bias if a bank estimates a model of (unconditional) CLU to obtain the (conditional) CLU at default. Instead, a Heckman selection model or endogenous treatment-regression model should be used to achieve unbiased results. Alternatively, a model based on the subsample of defaulted customers could give unbiased estimates if the rating is considered as a control variable.

In addition, we examine the consistency of information from different sources for estimating the CLU at default and whether cross-product information improves the estimate. In Table 6, we report the results for the Heckman selection models with different sources of information. Columns (1) and (2) provide the results for checking accounts, while columns (3) and (4) provide the results for credit cards.

(Insert Table 6 about here)

The results in column (1) of Table 6 allow for a comparison with the previous results for checking accounts without cross-product information. The results show that most coefficients for account activity and bank-customer relationship variables have the same sign for checking accounts and credit card accounts. However, a high *Amplitude* value is related to lower CLU at

default for checking accounts, but to higher *CLU* at default for credit card accounts. A substantially larger fraction of variance can be explained for credit card accounts than for checking accounts (the adjusted R^2 is 0.562 vs. 0.173). This is contrary to our results for default prediction, where we find early warning indicators for checking accounts to be more informative. Regarding informational synergies, we find that using information from different accounts can slightly improve the estimate of *CLU* at default. However, the informational benefit for credit line usage prediction is lower compared to default prediction.

2.4 Screening with cross-product information

Significant informational synergies can be realized by monitoring the activity from different accounts of a customer. We now investigate whether informational synergies can be used to screen potential customers. When someone opens a new account, the bank can gather information about their characteristics but it does not have information about their past account activity. However, in case the person already has a different account at the bank, the bank can observe the past activity on this account and use this source of information for screening. In our dataset, 20% of all customers of the bank have both a checking account *and* a credit card account. Hence, there are many potential customers with one existing account, for whom this additional source of information could be used for screening. In Table 7, we present the results when we use the account activity variables from credit card accounts and not from checking accounts (and vice versa). This analysis is based on clean cross-product information that can be used for screening.

(Insert Table 7 about here)

Checking account activity is more predictive of defaults on credit card accounts than credit cards variables (adj. R^2 of 0.25 vs. 0.20; see Table 3 (7)). For *CLU* at default, we observe that the informational content of cross-product information is weaker but still statistically significant.

2.5 Consistency of information

Figure 5 provides information about the consistency of *PD* and *CLU* at default estimates derived from different information sources. In Panel A1, we present the *PD* estimates for checking accounts, where the estimates are based on the account activity of checking accounts on the one hand and credit card accounts on the other hand. Similarly, in Panel B1 we show *PD* estimates for credit card accounts. In both cases the correlation between the estimates derived from checking account versus credit card account variables is around 50% (0.558 for checking accounts and 0.540 for credit card accounts). If we analyze the pairwise correlations between individual account activity measures, we find that the correlations are at maximum 0.35 for all variables except for *Days Overdraft* (0.483) and *Limit* (0.777). Thus, the account activity of both products potentially provides different informational content.

(Insert Figure 5 about here)

In Panel A2 of Figure 5, we show the estimates of *CLU* at default for checking accounts derived from account activity on checking accounts versus credit card accounts. The predictions are similar no matter on which source of information the estimates are based, and the correlation between the predicted values equals 0.950. In contrast, for credit card accounts, the correlation between both predictions of *CLU* at default (see Panel B2) is only 0.514, so that the informational content of both accounts is more likely to differ for this variable.

2.6 Economic and statistical significance of the informational synergies

In the analysis above we document synergies between information from checking and credit card accounts. We show that account activity variables have a significant impact on default predictions and credit line usage (at default) predictions.

We now examine how large these synergies are. How much can financial institutions gain if they exploit cross-product account activity information? What are the marginal benefits of using

additional sources of information? Quantifying the synergies in form of a monetary equivalent is challenging because we would have to make various assumptions about the size and composition of the portfolio, the types of loans and loan terms, the account activity and control variables used in the model, the empirical default rate, the empirical loss given default, and others. We therefore refrain from following this route. Instead, we assess the economic and statistical significance by computing standard measures of prediction accuracy and goodness of fit to assess the economic and statistical significance of the cross-product informational synergies (e.g., Grunert, Norden and Weber, 2005). These measures indicate the economic significance in the sense that we can measure how many false credit decisions a lender can avoid when exploiting cross-product informational synergies. Table 8 reports the accuracy of *PD* and *CLU* at default prediction results, which are based on the account activity of checking accounts (column 1), credit card accounts (column 2), and of both accounts (column 3). For the accuracy of default prediction, we report the adjusted McFadden R^2 , the value of the area under the receiver operating characteristic (ROC), type I errors (predicting no default for consumers that do default) and type II errors (predicting default for consumers that do not default). We calculate the type I and II errors based on binary default/no default-predictions using the empirical default rate in the respective sample as cut-off point. A value of one for the R^2 or ROC indicates a perfect prediction. For a random prediction these statistics are $R^2 = 0$ and $ROC = 0.5$.

(Insert Table 8 about here)

We find evidence for sizeable informational synergies in both accounts. First, lenders can improve their decision making significantly when they consider information from both accounts rather than information from one account. Second, the improvement in the adjusted McFadden R^2 , ROC and type I and type II errors is particularly strong when we compare information from credit card accounts (column 2) with information from both accounts (column 3). For defaults of credit card accounts, the adjusted McFadden R^2 increases from 0.204 to 0.278, the ROC from

0.845 to 0.892, the type I error decreases from 23.76% to 15.89%, and the type II error decreases from 22.24% to 18.48%. Strikingly, when lenders additionally consider checking account information they can reduce the type I error by around eight percentage points.⁹ The absolute number of the type I error decreases from 5,096 to 3,408 (minus 1,688), corresponding to a reduction by 33%.¹⁰ We explain this strong finding with the fact that changes in consumer income (e.g., due to unemployment) are identifiable in checking accounts, but not in credit card accounts. As shown earlier, warning indications about decreases of cash inflows are critical for the prediction of defaults and credit line usage at default. Third, there is an improvement of all measures when we compare information from checking accounts (column 1) with information from both accounts (column 3) but the benefits are moderate. This finding shows again that checking accounts are more informative about changes in borrower quality than credit card accounts. Fourth, regarding the *CLU* at default, we find that information from checking accounts and credit cards accounts is useful but there is only a small improvement when cross-product information is considered.

By and large, there are sizeable benefits due to informational synergies. The benefits are highest when lenders consider checking account information next to credit card information.

3. Further empirical checks and robustness tests

3.1 Alternative definition of default

The previous analyses refer to the account level, as stipulated in the Basel III accord and EU Regulation No 575. We now repeat the analysis with default at the customer level. Under this

⁹ We note that the type I error is significantly more costly than the type II error. The loss of the credit granted to a consumer who will subsequently default is higher than the loss of the interest income of a rejected loan to a consumer who will subsequently not default.

¹⁰ The effect is even stronger for predicting the defaults of checking accounts: lenders can lower the type I error by around ten percentage points - corresponding to a 42% reduction of the type I error - when they consider checking account activity in addition to credit card account activity.

definition, a default occurs if at least one account of the customer is in default. To estimate the *PD*, we use the variables of both accounts individually or aggregate the variables.¹¹ Table 9 reports the results. Analyses with aggregated variables are shown in columns (1) and (3) and analyses with variables of both individual accounts in columns (2) and (4).

(Insert Table 9 about here)

For the model of a customer default with aggregated customer variables (column (1)), the adjusted R^2 is lower than for account-specific default models that were based on cross-product information (adj. R^2 of 0.191 versus 0.296 for checking accounts and 0.278 for credit card accounts). The sign of coefficients and statistical significances are as expected. The estimates using variables of both accounts individually (column (2)) are similar compared to the estimates for aggregated customer variables except for *Rating* and *Days Usage*. The coefficients for *Rating* in column (2) must be summed to be comparable with the *Rating* in column (1) because this is defined as the mean of both ratings. *Days Usage* in column (1) is slightly significant because of opposite effects for individual variables in column (2).

Estimating the credit line usage at customer default using aggregate customer variables is superior to estimation with individual account variables (adj. R^2 of 0.506 vs. 0.383). There are some differences regarding *Amplitude* and *Days Overdraft*. While high *Amplitude* on checking accounts is a positive signal, it is a negative signal on credit card accounts, which is consistent with our previous findings. We also find that the aggregated overdraft information is a better early warning signal of default probability than the separate information. For estimating credit line usage at default with aggregated customer variables, the adjusted R^2 of 0.506 is higher/lower

¹¹ For the aggregated variables, we define the *Rating* as the mean of both ratings, *Net Inflow/Lim* as the difference between monthly total inflows and outflows as a percentage of the total external limit, *Amplitude* as the difference between the maximum and minimum exposure in each month as a percentage of *Limit* at both accounts, *Limit* and *Bounced* as the sum of limits and average number of bounced debits, and *Days Usage*, *Days Overdraft*, and *Duration* as the maximum of the corresponding variables. The *CLU* at default is limit weighted.

than for the models with cross-product information for checking accounts (0.182) and credit card accounts (0.579).

3.2 Alternative definition of credit line usage

To measure credit line usage, it is possible to consider only positive values or to consider positive and negative values. From the lender's perspective, it is important to estimate potential losses, which only occur if the credit line usage is positive. We therefore consider restricted values of *CLU* and compare them with the unrestricted estimation. Table 10 reports the results.

(Insert Table 10 about here)

The results in Table 10 are similar to those from Table 6 in terms of sign, significance, magnitude, and model fit because the correlation between restricted and unrestricted *CLUAD* is rather high, especially for credit card accounts (0.62 for checking accounts and 0.97 for credit card accounts). However, the sign for *Amplitude* in checking accounts becomes positive, meaning that accounts with higher *Amplitude* have a higher restricted *CLUAD*. This is because accounts with higher amplitude have higher inflow, along with a higher variance in their in- and outflows. Customers with high amplitude take less credit, but in the event of default, the high amplitude often leads to a high credit line usage and, thus, to higher values of the (restricted) *CLUAD*. This interpretation is supported by the finding that the coefficient of *Amplitude* is significantly negative in regressions with a restriction of *CLUAD* to *negative* values. Overall, our main results hold both for restricted and unrestricted *CLUAD*.

3.3 Interaction effects of cross-product account activity measures

We further investigate whether similar information from different sources leads to an amplification of early warning indications, which could be seen as further evidence for cross-product informational synergies. To address this issue we add interaction terms of the activity

measures from both accounts (*Net Inflow/Lim*, *Amplitude*, *Limit*, *Bounced*, *Days Usage*, and *Days Overdraft*) to the default and *CLU* at default prediction models and repeat the analyses corresponding to Table 3, 4 and 6. Our key results are unchanged and the adjusted McFadden R^2 increases only marginally. The findings suggest that adding the interaction terms to the individual terms from both accounts results only in a marginal improvement.

3.4 Alternative prediction horizons

We examine whether our results are sensitive to the prediction horizon. For this purpose, we repeat the analyses with time horizons of one month and three months. At a one month time horizon, we achieve a higher adjusted R^2 when predicting credit card defaults compared to a time horizon of one year (adj. R^2 of 0.308 versus 0.278; Table 3, column (8)). However, the effects in terms of sign, significance, and magnitude are largely unchanged. Thus, our results also hold for shorter prediction horizons.

3.5 Fixed effects models

Our estimation of the probability of default is based on pooled probit regression models. Although we included various control variables, our estimates could be biased due to unobserved customer characteristics. We therefore add customer fixed effects to our regression models. In unreported analyses we find that the coefficients of *Rating*, *Net Inflow/Lim*, *Amplitude*, *Bounced*, *Days Usage*, and *Days Overdrafts* are similar to the pooled probit estimates for both types of account. The coefficients for bounced debits and for an increased percentage of days with negative balance or overdrafts are even higher in the fixed effects model. Moreover, we find that cross-product information results in substantially increased estimates of the probability of default. Overall, the fixed effects regression results confirm our previous findings.

3.6 Impact of the recent financial crisis

The recent financial crisis could affect our results because customers are more likely to default and they have a higher demand for credit. Indeed, when we compare 2009-2010 (crisis) with 2011-2014 (post-crisis), we find higher default rates during the crisis for checking accounts (1.05% versus 0.83%, $p < 0.001$) and credit card accounts (1.36% versus 1.00%, $p < 0.001$). In addition, we find that the credit line usage at default is significantly higher in 2009-2010 ($p < 0.001$); this shows that in a financial crisis, customers have a higher demand for credit. We repeat all analyses using observations from 2011-2014. The findings for the *PD*, *CLU*, and *CLUAD* are similar, indicating that they are robust in good and bad times.

4. Conclusion

We investigate potential synergies between different sources of private information in consumer credit. Informational synergies are important because they affect the access to credit, terms of credit, credit allocation in the economy, and the risk taking of lenders. Our setting enables us to analyze why consumers default, how much credit they take (unconditional and conditional on default), and how lenders can obtain early warning indications that capture the time-varying nature of credit risk in household finance. The analysis is based on a large and unique dataset of more than 1.7 million customer-month observations for 2007-2014, comprising dynamic information from checking accounts and credit card accounts for each customer.

We provide comprehensive evidence for substantial informational synergies between different credit products used by the same person. We find that activity measures from both account types contain information beyond credit scores, borrower characteristics, bank-borrower relationship characteristics and many other controls. Interestingly, information from checking accounts is more useful than that from credit cards. Activity measures from checking accounts indicate credit quality deterioration earlier and more accurately than those from credit cards. The

benefits of cross-production informational synergies are sizeable: Lenders can lower the type I error by around eight percentage points - corresponding to a 33% reduction of the type I error - when they consider checking account activity in addition to credit card account activity. We show that these different sources of information are beneficial for ex ante screening and ex interim monitoring of borrower quality. Lenders can exploit these synergies to improve their credit and customer relationship management, for example in the context of limit changes, account closures, or restructuring. We also investigate the consistency of information, the reasons for consumer defaults, and selection effects. The latter should be considered when estimating the credit line usage or exposure at default with a Heckman selection model or an endogenous treatment-regression model. Unbiased estimates of the exposure at default can also be achieved with econometric models on the subsample of defaulted accounts if ex ante measures of default risk, such as credit ratings, are taken into account.

Our results on cross-product informational synergies suggest that financial institutions can realize significant economies of scope in credit risk and customer relationship management when they simultaneously offer different services to the same customer. Individuals should consider informational synergies, and their potential to serve as a basis for relationship banking, when purchasing credit products from one or multiple financial institutions. Our findings suggest that benefits from informational synergies can be realized when consumers purchase different credit products from one financial institution. Investigating the potential interaction of informational synergies with product-specific competition in the credit market is a fruitful avenue for future research. A better understanding of the costs and benefits of an information monopoly versus those of perfect competition may improve the availability and allocation of consumer credit in the economy.

References

- Agarwal, S., S. Chomsisengphet, C. Liu, and N. Souleles. 2009. Benefits of relationship banking: evidence from consumer credit markets. Working Paper, Federal Reserve Bank of Chicago, No. 2010-05.
- Akçura, M.T., and K. Srinivasan. 2005. Research note: customer intimacy and cross-selling strategy. *Management Science* 51, 1007-1012.
- Allen, F., and D. Gale. 2000. Comparing Financial Systems. MIT Press, Cambridge MA.
- Bank for International Settlements. 2013. Statistics on payment, clearing and settlement systems in the CPSS countries.
- Berg, T., M. Puri, and J. Rocholl. 2014. Loan officer incentives, internal ratings and default rates. Working Paper, September 2014.
- Berlin, M., and L. Mester. 1999. Deposits and Relationship Lending. *Review of Financial Studies* 12, 579-607.
- Campbell, J. 2006. Household Finance. *Journal of Finance* 61, 1553-1604.
- Grunert, J., L. Norden, and M. Weber, 2005. The role of non-financial factors in internal credit ratings. *Journal of Banking and Finance* 29, 509-531.
- Gürtler, M., and M. Hibbeln. 2013. Improvements in loss given default forecasts for bank loans. *Journal of Banking and Finance* 37, 2354-2366.
- Heckman, J. 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* 5, 475-492.
- Heckman, J. 1978. Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46, 931-959.
- Heckman, J. 1979. Sample selection bias as a specification error. *Econometrica* 47, 153-161.

- Hertzberg, A., J. Liberti, and D. Paravasini. 2010. Information and incentives inside the firm: Evidence from loan officer rotation. *Journal of Finance* 65, 795-828.
- Jiménez, G., J. Lopez, and J. Saurina. 2009. Empirical Analysis of Corporate Credit Lines. *Review of Financial Studies* 22, 5069-5098.
- Kamakura, W., M. Wedel, F. De Rosa, and J. Mazzon. 2003. Cross-selling through database marketing: a mixed data factor analyzer for data augmentation and prediction. *International Journal of Research in Marketing* 20, 45-65.
- Kamakura, W., S. Ramaswami, and R. Srivastava. 1991. Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services. *International Journal of Research in Marketing* 8, 329-349.
- King, G., and L. Zeng. 2001. Logistic Regression in Rare Events Data. *Political Analysis* 9, 137-163.
- Krahnert, J., and R. Schmidt. 2004. The German Financial System. Oxford University Press.
- Li, S., B. Sun, and A.L. Montgomery. 2011. Cross-selling the right product to the right customer at the right time. *Journal of Marketing Research* 48, 683-700.
- Li, S., B. Sun, and R.T. Wilcox. 2005. Cross-selling sequentially ordered products: An application to consumer banking services. *Journal of Marketing Research* 42, 233-239.
- Mester, L., L. Nakamura, and M. Renault. 2007. Transaction accounts and loan monitoring. *Review of Financial Studies* 20, 529-556.
- Norden, L., and M. Weber. 2010. Credit line usage, checking account activity, and default risk of bank borrowers. *Review of Financial Studies* 23, 3665-3699.
- Puri, M., J. Rocholl, and S. Steffen. 2013. What kinds of bank-client relationships matter in reducing loan defaults and why? Working Paper, October 2013.
- Rajan, U., A. Seru, and V. Vig. 2015. The failure of models that predict failure: Distance, incentives, and defaults. *Journal of Financial Economics* 115, 237-260.

- Stango, V., and J. Zinman. 2014. Limited and Varying Consumer Attention Evidence from Shocks to the Salience of Bank Overdraft Fees. *Review of Financial Studies* 27, 990-1030.
- Vella, F., and M. Verbeek. 1999. Estimating and Interpreting Models with Endogenous Treatment Effects. *Journal of Business and Economic Statistics* 17, 473-478.
- Vissing-Jorgensen, A. 2012. Consumer Credit: Learning Your Customer's Default Risk from What (S)he Buys. Working Paper, April 2012.

Figure 1: Limit and negative balance of checking accounts

The long-dashed line presents negative balances (median, in euros) for defaulted checking accounts from 24 months prior to default to 12 months after default. The dashed line presents negative balances (default-weighted median, in euros) for non-defaulted checking accounts. The short-dashed and solid lines present limits (median, in euro/default-weighted median, in euros) for defaulted/non-defaulted checking accounts from 24 months prior to default.

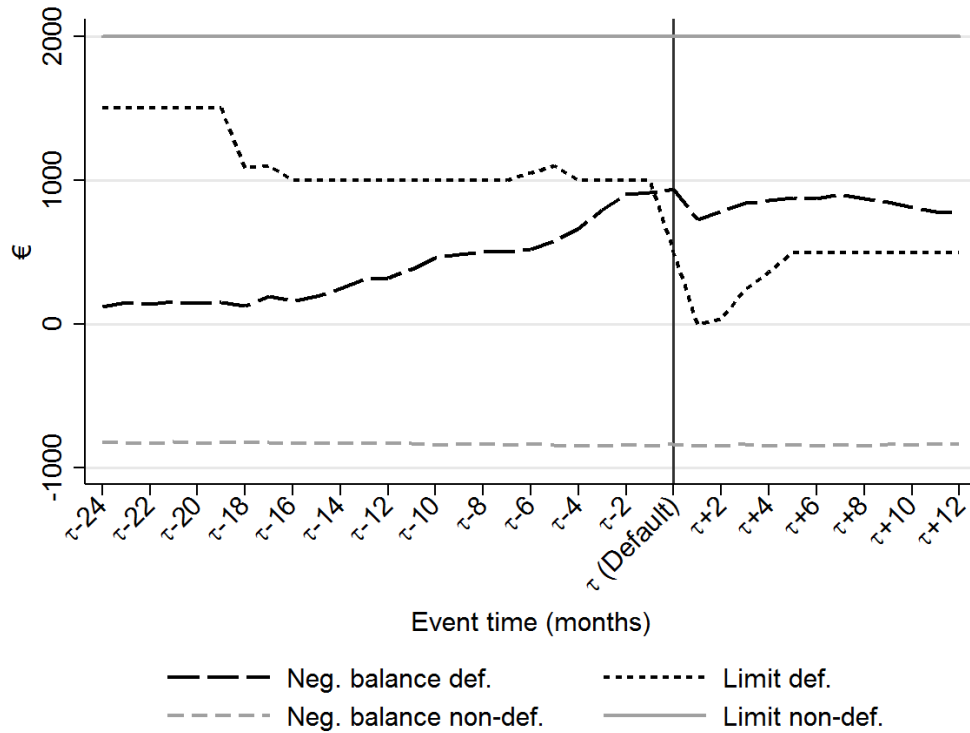


Figure 2: Cash inflows and outflows of checking accounts

The long-dashed line presents inflows (median, in euros) for defaulted checking accounts from 24 months prior to default to 12 months after default. The dashed line presents inflows (default weighted median, in euros) for non-defaulted checking accounts. The short-dashed and solid lines present outflows (median, in euros/default weighted median, in euros) for defaulted/non-defaulted checking accounts up to 24 months prior to default.

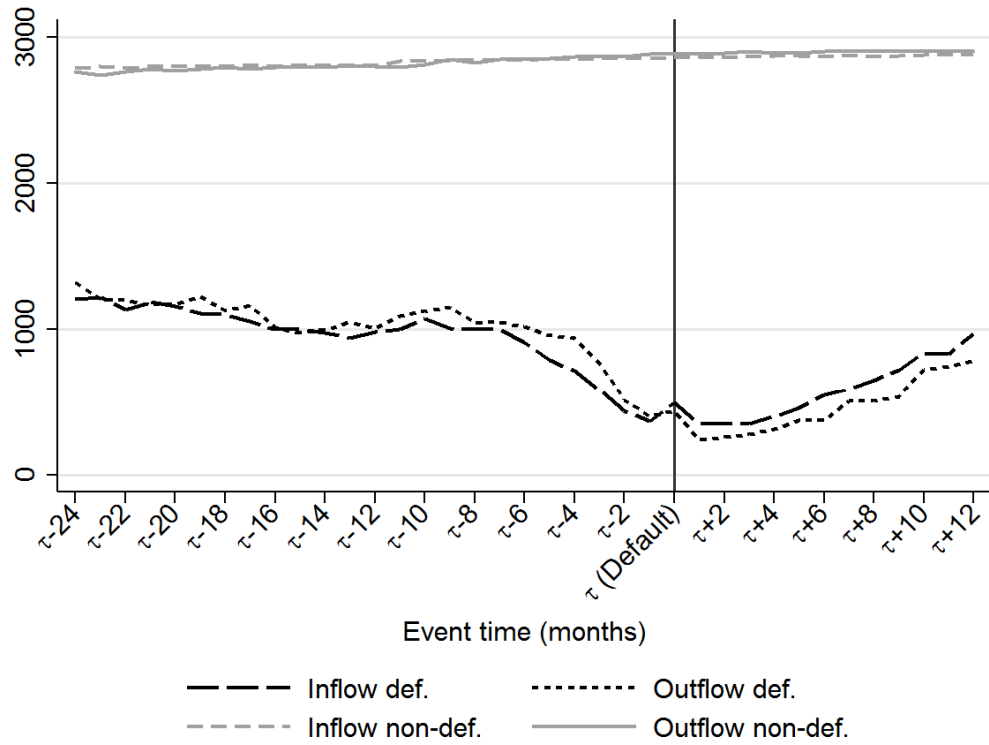
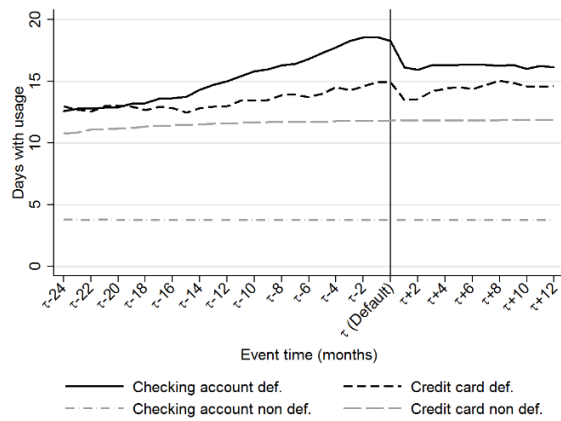


Figure 3: Days with credit line usage and overdrafts

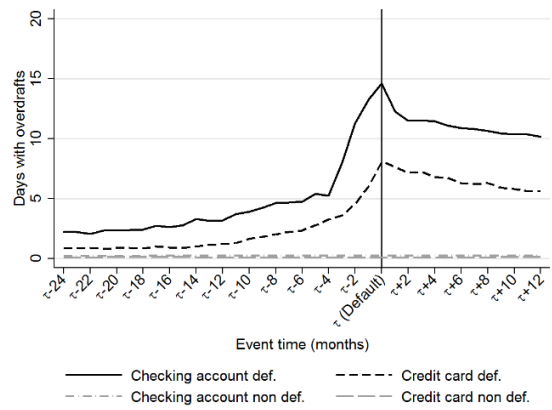
Panel A displays defaults of checking accounts, whereas Panel B displays defaults of credit cards. Days with usage are presented in Panel A1/B1 and days with overdrafts in Panel A2/B2. The solid and dash-dotted lines present days for defaulted (mean) and non-defaulted (default weighted mean) checking accounts, whereas the dashed and long-dashed lines present days for defaulted (mean) and non-defaulted (default-weighted mean) credit card accounts.

Panel A: Default of checking accounts

Panel A1: Days with usage

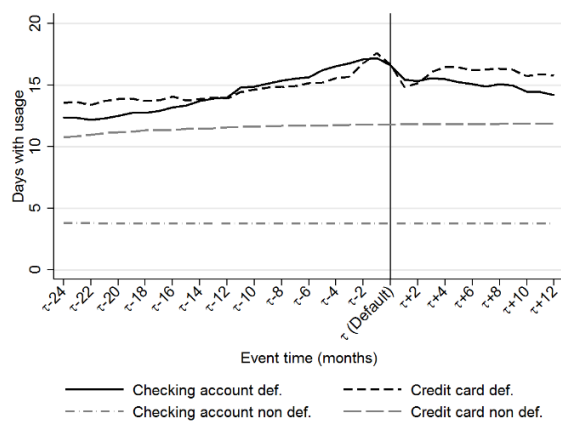


Panel A2: Days with overdrafts



Panel B: Default of credit cards

Panel B1: Days with usage



Panel B2: Days with overdrafts

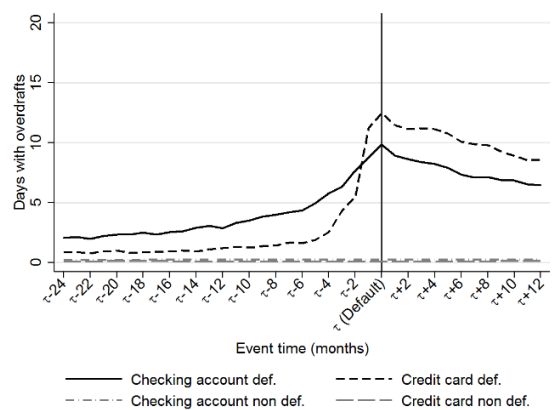


Figure 4: Credit line usage by credit rating

This figure displays the median of different rating categories for defaulted checking accounts from 24 months prior to default to 12 months after default. The five groups are formed at about 20%, 40%, 60%, and 80% quantiles with best ratings in category 1 and worst ratings in category 5.

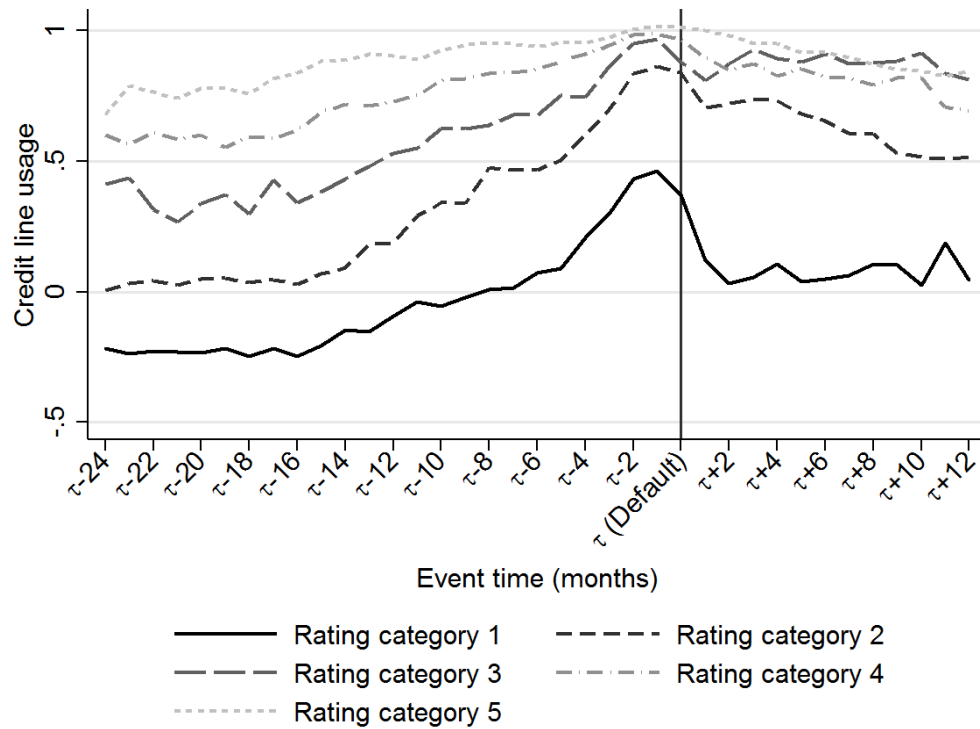
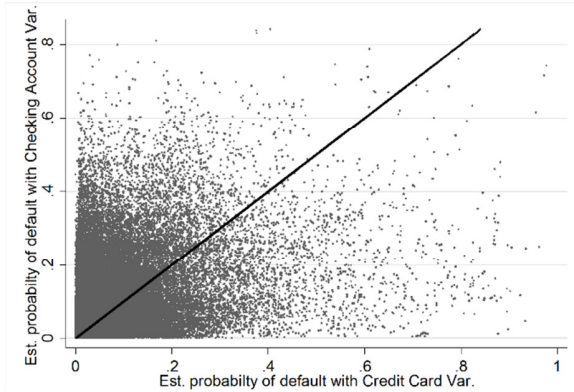


Figure 5: Consistency of information across credit products

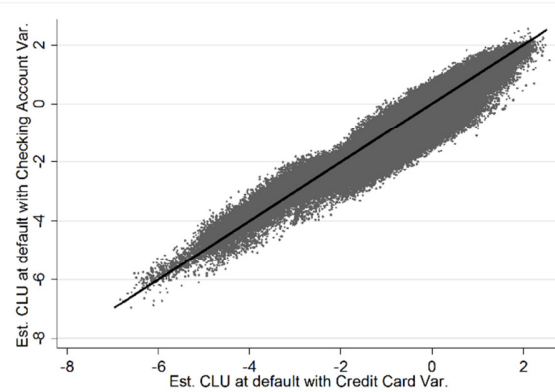
Panel A1 shows estimates of the probability of default (*PD*) for checking accounts where the estimates are based on account activity of checking accounts versus credit cards. Panel A2 shows the estimates of credit line usage at default (*CLUAD*). Panel B shows the corresponding *PD* estimates (Panel B1) and *CLUAD* estimates (Panel B2) for defaults of credit card accounts.

Panel A: Default of checking accounts

Panel A1: *PD* estimates

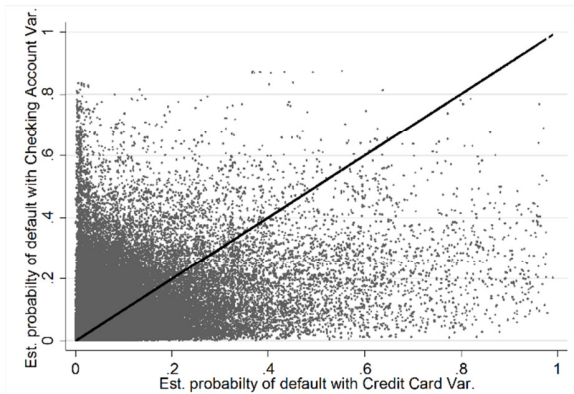


Panel A2: *CLUAD* estimates



Panel B: Default of credit cards

Panel B1: *PD* estimates



Panel B2: *CLUAD* estimates

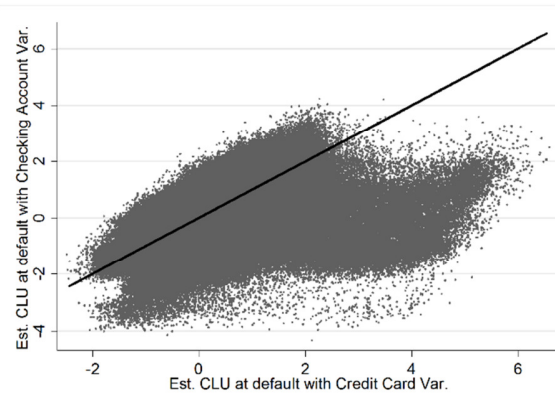


Table 1: Summary statistics

This table reports summary statistics. The sample covers the period from December 2007 to January 2014 (74 months). Panel A reports the number of account-months observations and the frequency of default events. The default events are at account level. Panel B reports the number of monthly observations without default, with default of only checking accounts or credit cards, or with joint defaults of both accounts. Panel C provides summary statistics of default risk, account activity, relationship, and account type variables. *Rating* is the probability of default estimate of the bank's internal rating system. *Net Inflow/Limit* is the ratio of monthly inflows minus outflows to the external limit and can be positive or negative. *Amplitude* is the difference of the monthly maximum and minimum balance per Limit and hence positive. *Limit* is reported in Euro values. *Bounced* is the average number of bounced credits in the previous year. *Days Usage* and *Days Overdrafts* are the average percentage of days in the previous year with positive credit line usage and overdrafts, respectively. *Duration* is defined as the time period since account opening in months. *Full Payment* refers to credit cards with monthly full payments instead of delayed payments. The last column reports the pairwise comparison of checking accounts and credit cards on the customer level. *** Statistically significant at the 0.1% level.

Panel A: Number of account-month observations and default events

Statistic	Checking Account	Credit Card
Number of account-month observations	1,779,356	1,781,408
Number of account-months with default in the subsequent 12 months	19,149	22,964
Number of defaults	1,639	2,101

Panel B: Number of account-month observations with individual vs. joint defaults

	Checking Account: No Default	Checking Account: Default
Credit Card: No default	1,754,099	5,857
Credit Card: Default	7,565	13,887

Panel C: Summary statistics and comparisons of checking accounts and credit cards

Variable	Checking Account		Credit Card		Checking Account Minus Credit Card	
	Mean	Median	Mean	Median	Pairwise Difference	t-stat.
Default Risk						
<i>Rating_t</i>	1.79	0.88	3.79	1.94	-1.939***	-428.25
Account Activity						
<i>Net Inflow/Lim_t</i>	0.02	0.00	0.00	0.00	0.0195***	19.18
<i>Amplitude_t</i>	1.80	1.03	0.39	0.24	1.407***	727.72
<i>Limit_t</i>	2437	2000	2581	2000	-143.7***	-163.29
<i>Bounced_t</i>	0.00	0.00	0.05	0.00	-0.0500***	-182.49
<i>Days Usage_t</i>	0.18	0.03	0.55	0.67	-0.367***	-1269.96
<i>Days Overdraft_t</i>	0.01	0.00	0.00	0.00	0.0070***	201.34
Relationship						
<i>Duration_t</i>	46.38	41.00	34.99	34.00	11.46***	616.02
Account Type						
<i>Full Payment_t</i>			0.93	1.00		

Table 2: Characteristics of accounts in default vs. non-default

This table reports the average differences between the accounts in default and non-default for several explanatory variables. In addition to the variables presented in Table 1, models (2) and (4) contain the variable *Prev. Def*, which is one if a previous default is observed on a different account of the same customer, and zero otherwise. The explanatory variables are observed in month t , and the default variable is equal to 1 if a default occurs in the period $[t, t+12 \text{ months}]$. Columns (1) and (2) refer to defaults versus non-defaults of checking accounts, and columns (3) and (4) refer to defaults versus non-defaults of credit card accounts. The explanatory variables in models (1) and (4) are checking account variables, whereas the explanatory variables in models (2) and (3) are from credit card accounts. Thus, models (2) and (4) refer to cross-product information. We report t -statistics in parentheses. *, **, and *** Statistically significant at the 5%, 1%, and 0.1% levels, respectively. (Note: Limit in 1,000 euros.)

	Default _{t+12} vs. Non-Default _{t+12} : Checking account		Default _{t+12} vs. Non-Default _{t+12} : Credit card	
	Inform. Source: Checking Account	Inform. Source: Credit Card	Inform. Source: Credit Card	Inform. Source: Checking Account
	(1)	(2)	(3)	(4)
Default Risk				
<i>Rating_t</i>	1.825*** (191.16)	1.079*** (129.85)	1.110*** (146.77)	1.729*** (197.96)
Account Activity				
<i>Net Inflow/Lim_t</i>	-0.0940*** (-9.97)	-0.0474*** (-16.07)	-0.0651*** (-24.21)	-0.0880*** (-10.23)
<i>Amplitude_t</i>	-0.814*** (-41.11)	0.128*** (29.70)	0.280*** (71.81)	-0.587*** (-32.51)
<i>Limit_t</i>	-0.716*** (-54.84)	-1.072*** (-83.33)	-1.063*** (-87.05)	-0.693*** (-58.19)
<i>Bounced_t</i>	0.00677*** (16.19)	0.336*** (120.29)	0.345*** (139.36)	0.00761*** (19.41)
<i>Days Usage_t</i>	0.456*** (218.64)	0.0355*** (14.03)	0.0644*** (27.93)	0.420*** (219.72)
<i>Days Overdraft_t</i>	0.109*** (309.61)	0.0464*** (216.29)	0.0482*** (304.21)	0.115*** (325.25)
Relationship				
<i>Duration_t</i>	-17.42*** (-71.83)	-12.22*** (-77.06)	-12.27*** (-84.87)	-16.51*** (-73.23)
<i>Prev. Def_t</i>		0.0210*** (51.18)		0.0419*** (93.13)
Account Type				
<i>Full Payment_t</i>		-0.263*** (-135.71)	-0.231*** (-130.24)	
Number of observations	1,779,356	1,779,356	1,781,408	1,781,408

Table 3: PD from checking accounts, credit cards, and cross-product information

This table reports probit estimates of the probability of default (*PD*) for several explanatory variables, using different sources of information. The explanatory variables are observed in month t , and the default variable is equal to 1 if a default occurs in the period $[t, t+12 \text{ months}]$. Columns (1)-(4) refer to defaults of checking accounts, and columns (5)-(8) refer to defaults of credit card accounts. The explanatory variables in the upper half of the table are checking account variables, whereas the explanatory variables in the bottom half are from credit cards. Thus, columns (4) and (8) include cross-product information. We report t -statistics clustered at the customer level in parentheses. † , * , ** , and *** Statistically significant at the 10%, 5%, 1%, and 0.1% levels, respectively. (Note: Limit in 1,000 euros.)

	Default _{t+12} : Checking Account				Default _{t+12} : Credit Card			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CHECKING ACCOUNT VARIABLES								
Default Risk								
<i>Rating_t</i>	0.4626 ^{***} (52.184)		0.2215 ^{***} (19.633)	0.1703 ^{***} (14.915)				0.1314 ^{***} (13.178)
Account Activity								
<i>Net Inflow/Lim_t</i>		-0.0479 ^{***} (-14.394)	-0.0561 ^{***} (-15.847)	-0.0616 ^{***} (-17.407)				-0.0475 ^{***} (-18.548)
<i>Amplitude_t</i>		-0.0687 ^{***} (-7.902)	-0.0645 ^{***} (-7.470)	-0.0614 ^{***} (-7.348)				-0.0360 ^{***} (-6.242)
<i>Limit_t</i>		-0.1289 ^{***} (-12.371)	-0.1322 ^{***} (-12.231)	-0.0842 ^{***} (-6.678)				-0.0750 ^{***} (-6.681)
<i>Bounced_t</i>		0.2471 [*] (2.489)	0.2288 [*] (2.249)	0.1853 [†] (1.779)				0.1649 (1.571)
<i>Days Usage_t</i>		1.2349 ^{***} (38.682)	0.8366 ^{***} (22.237)	0.9115 ^{***} (22.089)				0.7657 ^{***} (20.266)
<i>Days Overdraft_t</i>		2.1567 ^{***} (21.982)	1.8995 ^{***} (19.126)	1.4664 ^{***} (13.412)				1.1692 ^{***} (11.174)
Relationship								
<i>Duration_t</i>		-0.0084 ^{***} (-14.432)	-0.0055 ^{***} (-9.420)	-0.0027 ^{***} (-4.113)				-0.0018 ^{**} (-3.006)
<i>Prev. Def_t</i>								-0.0191 (-0.236)
CREDIT CARD VARIABLES								
Default Risk								
<i>Rating_t</i>				0.1284 ^{***} (10.279)	0.3947 ^{***} (40.648)		0.2814 ^{***} (23.018)	0.1626 ^{***} (12.498)
Account Activity								
<i>Net Inflow/Lim_t</i>				-0.1078 ^{***} (-17.330)		-0.1226 ^{***} (-28.027)	-0.1493 ^{***} (-26.392)	-0.1418 ^{***} (-22.740)
<i>Amplitude_t</i>				0.0811 ^{***} (8.698)		0.1323 ^{***} (19.145)	0.1518 ^{***} (19.972)	0.1765 ^{***} (23.282)
<i>Limit_t</i>				-0.0241 [†] (-1.802)		-0.1057 ^{***} (-11.646)	-0.0667 ^{***} (-7.303)	0.0053 (0.468)
<i>Bounced_t</i>				0.0343 [*] (2.323)		0.1628 ^{***} (7.113)	0.1402 ^{***} (6.692)	0.0554 ^{***} (3.594)
<i>Days Usage_t</i>				-0.5843 ^{***} (-13.448)		0.1516 ^{***} (5.276)	-0.3906 ^{***} (-10.949)	-0.4539 ^{***} (-11.408)
<i>Days Overdraft_t</i>				0.8898 ^{***} (5.346)		4.7102 ^{***} (22.811)	4.1120 ^{***} (20.775)	2.2709 ^{***} (12.012)
Relationship								
<i>Duration_t</i>				-0.0068 ^{***} (-6.929)		-0.0120 ^{***} (-17.658)	-0.0104 ^{***} (-15.669)	-0.0080 ^{***} (-8.839)
<i>Prev. Def_t</i>				-0.1909 [†] (-1.651)				
Account Type								
<i>Full Payment_t</i>				-0.1120 ^{***} (-3.627)		-0.5604 ^{***} (-22.549)	-0.2670 ^{***} (-9.105)	-0.0398 (-1.224)
<i>Constant</i>	-2.2031 ^{***} (-47.816)	-1.9650 ^{***} (-24.078)	-1.9854 ^{***} (-23.254)	-1.9658 ^{***} (-23.286)	-2.5644 ^{***} (-60.191)	-1.2397 ^{***} (-15.919)	-1.9346 ^{***} (-22.661)	-2.1022 ^{***} (-23.216)
<i>Customer controls</i>	No	Yes	Yes	Yes	No	Yes	Yes	Yes
<i>Year-month FE</i>	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	1,779,356	1,779,356	1,779,356	1,779,356	1,781,408	1,781,408	1,781,408	1,781,408
McFadden adj. R ²	0.185	0.259	0.277	0.296	0.101	0.185	0.204	0.278

Table 4: CLU from checking accounts, credit cards, and cross-product information

This table reports OLS estimates of the credit line usage (*CLU*) for several explanatory variables, using different sources of information. The explanatory variables are observed in month t , and the credit line usage variable is defined as described in Section 2.3. Columns (1)-(4) refer to *CLU* of checking accounts, and columns (5)-(8) refer to *CLU* of credit card accounts. The explanatory variables in the upper half of the table are checking account variables, whereas the explanatory variables in the bottom half are from credit cards. Thus, columns (4) and (8) include cross-product information. We report t -statistics clustered at the customer level in parentheses. The symbols † , * , ** , and *** Statistically significant at the 10%, 5%, 1%, and 0.1% levels, respectively. (Note: Limit in 1,000 euros.)

	Credit Line Usage _{t+12} : Checking Account				Credit Line Usage _{t+12} : Credit Card			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CHECKING ACCOUNT VARIABLES								
Default Risk								
<i>Rating_t</i>	0.9318 ^{***} (58.151)		0.8904 ^{***} (47.511)	0.8988 ^{***} (47.058)				-0.0539 ^{***} (-5.895)
Account Activity								
<i>Net Inflow/Lim_t</i>		-0.0830 ^{***} (-12.691)	-0.0981 ^{***} (-15.356)	-0.0917 ^{***} (-14.307)				-0.0311 ^{***} (-10.838)
<i>Amplitude_t</i>		-0.3330 ^{***} (-33.391)	-0.2838 ^{***} (-28.685)	-0.2926 ^{***} (-29.630)				-0.0258 ^{***} (-4.797)
<i>Limit_t</i>		0.2932 ^{***} (31.228)	0.3313 ^{***} (36.989)	0.3354 ^{***} (25.976)				-0.0474 ^{***} (-3.554)
<i>Bounced_t</i>		-0.2499 (-0.475)	-0.3205 (-0.664)	-0.2873 (-0.605)				0.0025 (0.025)
<i>Days Usage_t</i>		2.4052 ^{***} (76.779)	0.4131 ^{***} (12.564)	0.2693 ^{***} (7.582)				-0.0862 ^{**} (-3.027)
<i>Days Overdraft_t</i>		0.5874 ^{***} (5.470)	-0.4634 ^{***} (-3.700)	-0.3974 ^{**} (-2.861)				-0.4173 ^{***} (-4.236)
Relationship								
<i>Duration_t</i>		-0.0093 ^{***} (-18.320)	0.0004 (0.890)	-0.0033 ^{***} (-6.055)				0.0020 ^{***} (3.947)
<i>Prev. Def_t</i>								0.0920 (1.483)
CREDIT CARD VARIABLES								
Default Risk								
<i>Rating_t</i>				-0.0137 (-0.935)	1.3460 ^{***} (45.662)		1.5714 ^{***} (44.522)	1.5955 ^{***} (44.413)
Account Activity								
<i>Net Inflow/Lim_t</i>				-0.0012 (-0.124)		-0.5696 ^{***} (-16.870)	-0.6229 ^{***} (-20.547)	-0.6357 ^{***} (-20.574)
<i>Amplitude_t</i>				0.1469 ^{***} (7.034)		-1.4780 ^{***} (-29.263)	-1.1732 ^{***} (-27.496)	-1.1470 ^{***} (-28.117)
<i>Limit_t</i>				-0.0552 ^{**} (-3.983)		-0.1365 ^{***} (-12.802)	0.0733 ^{***} (7.274)	0.0998 ^{***} (6.833)
<i>Bounced_t</i>				-0.0850 ^{***} (-3.978)		0.1309 ^{***} (6.827)	-0.0755 ^{***} (-4.788)	-0.0409 [*] (-2.544)
<i>Days Usage_t</i>				0.5244 ^{***} (12.351)		2.9418 ^{***} (47.752)	-0.3151 ^{***} (-9.688)	-0.3086 ^{***} (-9.559)
<i>Days Overdraft_t</i>				-0.5038 [*] (-2.276)		4.3640 ^{***} (17.963)	-2.4887 ^{***} (-9.865)	-1.8416 ^{***} (-8.262)
Relationship								
<i>Duration_t</i>				0.0112 ^{***} (12.476)		-0.0128 ^{***} (-15.324)	-0.0070 ^{***} (-10.425)	-0.0092 ^{***} (-11.112)
<i>Prev. Def_t</i>				-0.0316 (-0.250)				
Account Type								
<i>Full Payment_t</i>				-0.0014 (-0.044)		-0.6265 ^{***} (-9.799)	-1.5919 ^{***} (-24.303)	-1.5312 ^{***} (-22.895)
Constant	-0.5063 ^{***} (-13.864)	-1.4147 ^{***} (-10.955)	-1.3495 ^{***} (-11.111)	-1.5040 ^{***} (-11.984)	-1.5633 ^{***} (-34.066)	0.0759 (0.785)	-1.8783 ^{***} (-19.076)	-1.8372 ^{***} (-18.667)
Customer controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Year-month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	1,745,534	1,745,534	1,745,534	1,745,534	1,742,049	1,742,049	1,742,049	1,742,049
Adj. R ²	0.119	0.179	0.234	0.239	0.229	0.186	0.319	0.320

Table 5: Alternative estimation methods for CLU at default

This table contrasts different methods for estimating the credit line usage at default. The explanatory variables are observed in month t . Columns (1) and (2) refer to the CLU_{all} method described in Section 1.3. Columns (3) and (4) contain estimates of the model CLU_{Def} , which is the same method as CLU_{all} , but estimated on the subsample of defaulted accounts. Column (5) refers to the Heckman selection model $CLU_{Heckman}$ and column (6) provides estimates for the endogenous treatment-regression $CLU_{End.Treat}$. We report t -statistics clustered at the customer level in parentheses. † , * , ** , and *** Statistically significant at the 10%, 5%, 1%, and 0.1% levels, respectively. (Note: Limit in 1,000 euros)

	CLU_{all}		CLU_{Def}		$CLU_{Heckman}$	$CLU_{End.Treat}$
	(1)	(2)	(3)	(4)	(5)	(6)
Default Risk						
<i>Rating_t</i>		0.8904 ^{***} (47.511)		0.2806 ^{***} (4.678)		
<i>Default_t</i>						3.8564 ^{***} (10.132)
Account Activity						
<i>Net Inflow/Lim_t</i>	-0.0830 ^{***} (-12.691)	-0.0981 ^{***} (-15.356)	-0.0718 [†] (-1.822)	-0.0793 [*] (-2.102)	-0.0797 ^{***} (-6.939)	-0.0838 ^{***} (-43.451)
<i>Amplitude_t</i>	-0.3330 ^{***} (-33.391)	-0.2838 ^{***} (-28.685)	-0.0365 (-0.844)	-0.0275 (-0.634)	-0.0269 ^{***} (-4.673)	-0.3326 ^{***} (-342.296)
<i>Limit_t</i>	0.2932 ^{***} (31.228)	0.3313 ^{***} (36.989)	-0.0296 (-1.335)	-0.0453 [*] (-2.038)	-0.0448 ^{***} (-5.600)	0.2985 ^{***} (171.053)
<i>Bounced_t</i>	-0.2499 (-0.475)	-0.3205 (-0.664)	0.0472 (0.558)	0.0696 (1.063)	0.0726 (1.102)	-0.3023 ^{***} (-6.803)
<i>Days Usage_t</i>	2.4052 ^{***} (76.779)	0.4131 ^{***} (12.564)	1.1493 ^{***} (8.267)	0.7672 ^{***} (6.958)	0.7438 ^{***} (19.015)	2.1981 ^{***} (195.769)
<i>Days Overdraft_t</i>	0.5874 ^{***} (5.470)	-0.4634 ^{***} (-3.700)	0.7602 ^{***} (6.992)	0.4798 ^{***} (4.098)	0.4762 ^{***} (6.842)	-0.1039 (-1.638)
Relationship						
<i>Duration_t</i>	-0.0093 ^{***} (-18.320)	0.0004 (0.890)	-0.0074 ^{***} (-4.218)	-0.0039 [*] (-2.521)	-0.0037 ^{***} (-8.520)	-0.0083 ^{***} (-90.576)
Constant	-1.4147 ^{***} (-10.955)	-1.3495 ^{***} (-11.111)	0.2805 (0.933)	0.1997 (0.670)	2.3201 ^{***} (13.925)	-1.5363 ^{***} (-32.471)
<i>Lambda: Def.</i>					-0.7217 ^{***} (-26.960)	-0.7217 ^{***} (-26.755)
<i>Lambda: No Def.</i>						3.3774 ^{***} (43.298)
Customer Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year-month FE	Yes	Yes	Yes	Yes	Yes	Yes
Def. x Acc. controls	No	No	No	No	No	Yes
Def. x Rel. controls	No	No	No	No	No	Yes
Number of observations	1,745,534	1,745,534	19,149	19,149	19,149	1,745,534
Adj. R ²	0.179	0.234	0.142	0.171	0.173	0.181

Table 6: CLU at default and cross-product information

This table reports estimates of the credit line usage at default ($CLUAD$) for checking accounts, credit card accounts, and with cross-product information. Estimates are based on the Heckman selection model $CLU_{Heckman}$. The explanatory variables are observed in month t . Columns (1) and (2) refer to defaulted checking accounts and columns (3) and (4) refer to defaults of credit card accounts. The explanatory variables in the upper half of the table are checking account variables, whereas the explanatory variables in the bottom half are observed on credit card accounts. Thus, columns (2) and (4) include cross-product information. We report t -statistics clustered at the customer level in parentheses. † , * , ** , and *** Statistically significant at the 10%, 5%, 1%, and 0.1% levels, respectively. (Note: Limit in 1,000 euros)

	CLU at Default $_{t+12}$: Checking Account		CLU at Default $_{t+12}$: Credit Card	
	(1)	(2)	(3)	(4)
CHECKING ACCOUNT VARIABLES				
Account Activity				
<i>Net Inflow/Lim_t</i>	-0.0797 ^{***} (-6.939)	-0.0772 ^{***} (-6.752)		0.0391 ^{***} (5.409)
<i>Amplitude_t</i>	-0.0269 ^{***} (-4.673)	-0.0431 ^{***} (-7.311)		-0.0670 ^{***} (-17.159)
<i>Limit_t</i>	-0.0448 ^{***} (-5.600)	-0.0565 ^{***} (-5.621)		0.0989 ^{***} (14.023)
<i>Bounced_t</i>	0.0726 (1.102)	0.0842 (1.282)		0.0093 (0.205)
<i>Days Usage_t</i>	0.7438 ^{***} (19.015)	0.7127 ^{***} (17.905)		0.0127 (0.480)
<i>Days Overdraft_t</i>	0.4762 ^{***} (6.842)	0.2343 ^{**} (3.049)		0.3970 ^{***} (7.242)
Relationship				
<i>Duration_t</i>	-0.0037 ^{***} (-8.520)	-0.0022 ^{***} (-4.265)		0.0004 (1.149)
<i>Prev. Def_t</i>				0.2033 ^{***} (5.685)
CREDIT CARD VARIABLES				
Account Activity				
<i>Net Inflow/Lim_t</i>		-0.0384 (-1.582)	-0.2376 ^{***} (-15.200)	-0.2122 ^{***} (-13.784)
<i>Amplitude_t</i>		0.1228 ^{***} (9.919)	0.9247 ^{***} (122.519)	0.9308 ^{***} (123.307)
<i>Limit_t</i>		0.0074 (0.655)	-0.1519 ^{***} (-23.634)	-0.2062 ^{***} (-26.502)
<i>Bounced_t</i>		0.0037 (0.384)	-0.0077 (-1.077)	-0.0219 ^{**} (-3.101)
<i>Days Usage_t</i>		0.2206 ^{***} (5.636)	0.0656 [*] (2.087)	0.0349 (1.109)
<i>Days Overdraft_t</i>		0.0502 (0.406)	0.9162 ^{***} (10.823)	0.5432 ^{***} (6.080)
Relationship				
<i>Duration_t</i>		-0.0026 ^{**} (-3.270)	-0.0027 ^{***} (-5.785)	-0.0039 ^{***} (-6.930)
<i>Prev. Def_t</i>		0.1705 [*] (2.271)		
Account Type				
<i>Full Payment_t</i>		-0.0766 ^{**} (-2.802)	-0.5989 ^{***} (-25.640)	-0.5521 ^{***} (-24.383)
<i>Constant</i>	2.3201 ^{***} (13.925)	2.1924 ^{***} (13.107)	2.2606 ^{***} (18.071)	2.2203 ^{***} (17.959)
<i>Lambda: Default</i>	-0.7217 ^{***} (-26.960)	-0.6953 ^{***} (-25.688)	-0.5551 ^{***} (-18.083)	-0.5242 ^{***} (-17.316)
<i>Customer Controls</i>	Yes	Yes	Yes	Yes
<i>Year-month FE</i>	Yes	Yes	Yes	Yes
Number of observations	19,149	19,149	22,964	22,964
Adj. R ²	0.173	0.182	0.562	0.579

Table 7: Screening with cross-product information

This table reports estimates of the probability of default and credit line usage at default for several explanatory variables. The explanatory variables are observed in month t , and the default variable is equal to 1 if a default occurs in the period $[t, t+12 \text{ months}]$. Estimates for credit line usage at default are based on the Heckman selection model $CLU_{Heckman}$. Columns (1) and (2) refer to the estimation of probability of default and credit line usage at default of checking accounts, and columns (3) and (4) refer to the corresponding estimations of credit card accounts. The explanatory variables for checking accounts defaults are credit card variables, whereas the explanatory variables for credit card defaults are from checking accounts. Thus, this table includes pure cross-product information. We report t -statistics clustered at the customer level in parentheses. †, *, **, and *** Statistically significant at the 10%, 5%, 1%, and 0.1% levels, respectively. (Note: Limit in 1,000 euros.)

Dependent Variable	Default _{t+12} Of Checking Account	CLU At Default _{t+12} Of Checking Account	Default _{t+12} of Credit Card Account	CLU at Default _{t+12} of Credit Card Account
Independent Variables	Credit Card Account	Credit Card Account	Checking Account	Checking Account
	(1)	(2)	(3)	(4)
Default Risk				
<i>Rating_t</i>	0.2715*** (24.155)		0.1718*** (16.769)	
Account Activity				
<i>Net Inflow/Lim_t</i>	-0.1144*** (-21.253)	-0.0517* (-2.003)	-0.0426*** (-15.034)	0.0424*** (4.259)
<i>Amplitude_t</i>	0.0387*** (3.950)	0.0714*** (5.365)	-0.0153** (-2.873)	0.0209*** (4.111)
<i>Limit_t</i>	-0.1012*** (-9.525)	-0.0521*** (-5.508)	-0.0981*** (-9.795)	-0.0163* (-2.095)
<i>Bounced_t</i>	0.1225*** (7.837)	0.0568*** (5.827)	0.1986* (1.994)	-0.0366 (-0.563)
<i>Days Usage_t</i>	-0.4614*** (-12.775)	0.2497*** (5.244)	0.7627*** (21.797)	0.3011*** (7.689)
<i>Days Overdraft_t</i>	2.1146*** (13.249)	0.6657*** (5.878)	1.7604*** (18.710)	1.5044*** (22.035)
Relationship				
<i>Duration_t</i>	-0.0094*** (-13.951)	-0.0061*** (-8.876)	-0.0052*** (-9.849)	-0.0042*** (-10.285)
<i>Prev. Def_t</i>	0.1383 (1.426)	-0.4741*** (-9.116)	0.1251† (1.799)	-0.2392*** (-5.974)
Account Type				
<i>Full Payment_t</i>	-0.3769*** (-14.280)	-0.0400 (-1.064)	-0.1618*** (-5.984)	-0.2159*** (-9.284)
<i>Constant</i>	-1.8416*** (-21.021)	1.8315*** (9.145)	-1.6782*** (-20.378)	2.3731*** (14.179)
<i>Lambda: Default</i>		-0.5279*** (-10.187)		-0.3307*** (-12.352)
<i>Customer controls</i>	Yes	Yes	Yes	Yes
<i>Year-month FE</i>	Yes	Yes	Yes	Yes
Number of observations	1,779,356	19,149	1,781,408	22,964
McFadden adj. R ² /adj. R ²	0.185	0.093	0.250	0.124

Table 8: Accuracy of PD and CLU at default estimates by information source

This table reports the accuracy for predicting defaults of checking accounts and credit card accounts, as well as the accuracy of corresponding *CLU* at default predictions. The estimates are based on (1) checking account information, (2) credit card information, and (3) information from both accounts. The accuracy of *PDs* is measured by the adjusted McFadden R^2 , the area under the receiver operating characteristic (ROC) and the type I and type II errors. We use the empirical default rate in the checking account (credit card account) sample of 1.00% (1.20%) as cut-off point to calculate the type I and type II errors. We measure the accuracy of the *CLU* at default estimates with the adjusted R^2 . The checking account sample comprises 1,779,356 observations and the credit card account sample 1,781,408 observations.

Dependent Variable	Measure	Information Source		
		(1) Checking Account	(2) Credit Card Account	(3) Both Accounts
<i>Default_{t+12} of checking account</i>	Adj. McFadden R^2	0.277	0.185	0.296
	ROC	0.906	0.843	0.913
	Type I error	13.99	23.81	13.69
	Type II error	18.03	23.63	17.38
<i>CLU at default_{t+12} of checking account</i>	Adj. R^2	0.173	0.093	0.182
<i>Default_{t+12} of credit card account</i>	Adj. McFadden R^2	0.250	0.204	0.278
	ROC	0.882	0.845	0.892
	Type I error	16.36	23.76	15.89
	Type II error	19.55	22.24	18.48
<i>CLU at default_{t+12} of credit card account</i>	Adj. R^2	0.124	0.562	0.579

Table 9: Default at the customer level

This table reports estimates of the probability of default and credit line usage at default at the customer level (instead of account level) for several explanatory variables. The explanatory variables are observed in month t , and the default variable is equal to 1 if a customer default occurs in the period $[t, t+12]$ months]. Estimates for credit line usage at default are based on the Heckman selection model $CLU_{Heckman}$. Columns (1)-(3) refer to the estimation of probability of default, and columns (4)-(6) refer to credit line usage at default. Models (1) and (3) refer to aggregated customer variables as independent variables, whereas models (2) and (4) contain estimates with non-aggregated variables. Part 1 refers to checking account variables and part 2 refers to credit card variables. (Note: For brevity, we do not present these models in one column.) We report t -statistics clustered at the customer level in parentheses. † , * , ** , and *** Statistically significant at the 10%, 5%, 1%, and 0.1% levels, respectively. (Note: Limit in 1,000 euros.)

Dependent Variable Independent Variables	Default _{t+12}			CLU at Default _{t+12}		
	Customer	Checking Acc. (part 1)	Credit Card Acc. (part 2)	Customer	Checking Acc. (part 1)	Credit Card Acc. (part 2)
	(1)	(2)	(3)	(4)		
Default Risk						
<i>Rating_t</i>	0.2870 ^{***} (16.681)	0.1266 ^{***} (9.963)	0.1129 ^{***} (6.920)			
Account Activity						
<i>Net Inflow/Lim_t</i>	-0.0650 ^{***} (-12.227)	-0.0304 ^{***} (-11.037)	-0.0604 ^{***} (-6.333)	-0.0235 (-1.397)	-0.0351 ^{***} (-4.012)	-0.0713 ^{**} (-3.010)
<i>Amplitude_t</i>	-0.0125 (-1.264)	-0.0068 (-1.496)	-0.0104 (-0.720)	-0.3231 ^{***} (-64.016)	-0.0346 ^{***} (-7.496)	0.0721 ^{***} (4.776)
<i>Limit_t</i>	-0.0332 ^{***} (-5.395)	-0.0581 ^{***} (-4.457)	-0.0024 (-0.184)	-0.0000 ^{***} (-6.540)	-0.0000 (-0.553)	-0.0279 ^{**} (-2.914)
<i>Bounced_t</i>	0.0435 (1.119)	0.1571 (1.481)	0.0272 (1.355)	0.0230 (1.450)	0.0735 (1.526)	-0.0049 (-0.518)
<i>Days Usage_t</i>	-0.1101 [*] (-2.311)	0.4570 ^{***} (9.817)	-0.2879 ^{***} (-5.866)	0.3409 ^{***} (9.420)	0.5208 ^{***} (14.287)	0.1273 ^{***} (3.756)
<i>Days Overdraft_t</i>	2.5958 ^{***} (22.454)	1.8337 ^{***} (14.296)	1.0824 ^{***} (4.924)	0.5638 ^{***} (9.949)	-0.0974 (-1.376)	0.2214 [†] (1.665)
Relationship						
<i>Duration_t</i>	-0.0020 ^{***} (-3.449)	-0.0004 (-0.529)	-0.0050 ^{***} (-4.761)	-0.0029 ^{***} (-8.563)	-0.0007 (-1.607)	-0.0021 ^{**} (-3.031)
Account Type						
<i>No Full_t</i>	-0.1517 ^{***} (-3.887)		-0.1750 ^{***} (-4.561)	0.0122 (0.541)		-0.0360 (-1.444)
Constant	-2.4611 ^{***} (-22.179)	-2.3155 ^{***} (-20.966)		2.5455 ^{***} (16.645)	3.1527 ^{***} (18.367)	
<i>Lambda: Default</i>				-0.6631 ^{***} (-22.396)	-1.0522 ^{***} (-31.304)	
Customer controls	Yes	Yes		Yes	Yes	
Year-month FE	Yes	Yes		Yes	Yes	
Number of observations	1,767,269	1,767,269		11,621	11,621	
McFadden adj. R ² /adj. R ²	0.191	0.207		0.506	0.383	

Table 10: CLU at default restricted to positive values

This table reports estimates of the restricted credit line usage at default ($CLUAD_{res}$), which is defined as $CLUAD_{res,t+12} = \max(CLUAD_{t+12}, 0)$. The estimates refer to the potential loss from a bank's perspective as the loss equals zero if the account balance is positive at time of default. Analogous to the results presented in Table 6, the estimates are based on the Heckman selection model $CLU_{Heckman}$. We report t -statistics clustered at the customer level in parentheses. † , * , ** , and *** Statistically significant at the 10%, 5%, 1%, and 0.1% levels, respectively. (Note: Limit in 1,000 euros.)

	CLUAD _{res,t+12} : Checking Account			CLUAD _{res,t+12} : Credit Card	
	(1)	(2)		(3)	(4)
CHECKING ACCOUNT VARIABLES					
Account Activity					
<i>Net Inflow/Lim_t</i>	-0.0631 ^{***} (-11.364)	-0.0608 ^{***} (-11.036)	□		0.0362 ^{***} (5.321)
<i>Amplitude_t</i>	0.0783 ^{***} (28.156)	0.0697 ^{***} (24.533)			-0.0559 ^{***} (-15.239)
<i>Limit_t</i>	-0.0528 ^{***} (-13.622)	-0.0642 ^{***} (-13.238)			0.0819 ^{***} (12.353)
<i>Bounced_t</i>	-0.0169 (-0.531)	-0.0114 (-0.359)			0.0181 (0.427)
<i>Days Usage_t</i>	0.1716 ^{***} (9.062)	0.1461 ^{***} (7.614)			-0.0345 (-1.385)
<i>Days Overdraft_t</i>	0.7013 ^{***} (20.786)	0.5308 ^{***} (14.307)			0.4998 ^{***} (9.703)
Relationship					
<i>Duration_t</i>	-0.0004 [†] (-1.684)	0.0008 ^{**} (3.003)			-0.0007 [*] (-2.227)
<i>Prev. Def_t</i>					0.1682 ^{***} (5.006)
Account Activity					
<i>Net Inflow/Lim_t</i>		-0.0527 ^{***} (-4.490)		-0.1865 ^{***} (-12.735)	-0.1632 ^{***} (-11.289)
<i>Amplitude_t</i>		0.0537 ^{***} (8.991)		0.9860 ^{***} (139.468)	0.9899 ^{***} (139.586)
<i>Limit_t</i>		0.0104 [†] (1.916)		-0.1189 ^{***} (-19.749)	-0.1628 ^{***} (-22.268)
<i>Bounced_t</i>		0.0049 (1.047)		-0.0092 (-1.373)	-0.0229 ^{***} (-3.456)
<i>Days Usage_t</i>		0.1765 ^{***} (9.358)		0.0161 (0.547)	-0.0043 (-0.145)
<i>Days Overdraft_t</i>		0.0161 (0.270)		1.1643 ^{***} (14.681)	0.7640 ^{***} (9.101)
Relationship					
<i>Duration_t</i>		-0.0023 ^{***} (-5.870)		-0.0013 ^{**} (-2.923)	-0.0013 [*] (-2.474)
<i>Prev. Def_t</i>		0.1504 ^{***} (4.147)			
Account Type					
<i>Full Payment_t</i>		-0.0656 ^{***} (-4.975)		-0.8607 ^{***} (-39.343)	-0.8189 ^{***} (-37.633)
<i>Constant</i>	1.6152 ^{***} (20.010)	1.5424 ^{***} (19.110)		1.4379 ^{***} (12.268)	1.4024 ^{***} (12.072)
<i>Lambda: Default</i>	-0.2762 ^{***} (-21.167)	-0.2573 ^{***} (-19.594)		-0.1531 ^{***} (-5.323)	-0.1236 ^{***} (-4.346)
<i>Customer Controls</i>	Yes	Yes		Yes	Yes
<i>Year-month FE</i>	Yes	Yes		Yes	Yes
Number of observations	19,149	19,149		22,964	22,964
Adj. R ²	0.162	0.177		0.586	0.599