Forecasting Recoveries in Debt Collection - Debt Collectors and Information Production*

Johannes Kriebel
University of Muenster†

Kevin Yam
Seghorn AG ‡

May 21, 2019

Abstract

Recent theoretical work suggests that debt collection agencies are important in gathering and processing debtor information. We study a comprehensive dataset with information provided by original creditors and information gathered in third-party debt collection. In line with the theoretical results, the initial information is sparse, and the gathered information is essential for better-informed predictions.

Keywords: Loss given default, recovery rate, debt collection.

JEL Classification: G21, G22, G29, G3.

*We are deeply indebted to Seghorn AG for providing the dataset and kindly giving a wide range of advice on the debt collection industry that enabled this study. We are further grateful to the participants and discussants of the HypoVereinsbank-Seminar 2017 in Bochum, the GOR Financial Management and Financial Institutions Workshop 2017 in Magdeburg, the European Conference on Data Analysis 2017 in Wroclaw, the 35th Annual Conference of the French Finance Association in Paris, the 2018 Annual Meetings of the European Financial Management Association in Milan, Christopher Jung, Henning Cordes, Jörn Debener, Judith Schneider, and other members of the Finance Center Muenster for their valuable comments, which helped to improve this study.

†Corresponding author, Universitaetsstr. 14-16, 48143 Muenster, Germany, johannes.kriebel@wiwi.uni-muenster.de, phone +49-251-83-22692, fax +49-251-83-22882

‡Seghorn AG, Legienstrasse 1, 28188 Bremen, Germany, yam@seghorn.de, phone +49-421-4391-404, fax +49-421-4391-316

1 Introduction

The management of accounts receivable and accounts payable plays a vital role in the balance sheets of many producers of goods and services. It is common practice in many industries (e.g., insurance, telecommunication, and mail-order services) to commission specialized collection agencies to collect distressed receivables. Likewise, banks tend to resort to collection agencies in difficult cases (Thomas et al., 2012). According to industry studies, collection firms managed a total of €60bn of receivables in Germany at the end of 2015 (Bülow, 2016) and over $792bn in the United States at the end of 2016 (Ernst & Young, 2017).

Surprisingly, little is known on how collection agencies manage accounts successfully. There is particularly little knowledge on what factors drive collection recoveries. Han and Jang (2013) and Thomas et al. (2012) study these factors for bank loans, Beck et al. (2017) study these factors for other goods and services, and Hoechstoetter et al. (2012) study the choice of prediction models. To the best of our knowledge, these are the only results in this field.

A collection agency can use information from different sources for predicting collection recoveries. One important differentiation is between the information original creditors provide and the information debt collection agencies gather themselves.[1] Earlier empirical work, such as Hoechstoetter et al. (2012), Thomas et al. (2012), and Beck et al. (2017), mainly focuses on initially available information. Thomas et al. (2012) and Beck et al. (2017) find this information to be sparse.

Interestingly, recent theoretical work, such as Drozd and Serrano-Padial (2017), suggests that debt collection has an important role in gathering and processing debtor information. Fedaseyeu and Hunt (2018) put

[1] We refer to the collection process at the original creditor as ”in-house collection” and to the later process at the debt collector as ”third-party collection”. We use the term ”debtor” when referring to the individual or company that is overdue on one or more claims. The initial holder of the claim is referred to as the ”original creditor” or ”business partner”.

a stronger emphasis on reputational issues but mention that debt collection agencies particularly gain information on a customer's willingness to pay. There is no empirical evidence on the importance of information production in debt collection agencies to date.

Given these considerations, the main purpose of this study is to examine how valuable information gathered in third-party debt collection is for collection recovery predictions. We make three important contributions to the literature:

First, we study predictive characteristics on a large proprietary dataset of more than 300,000 distressed claims in a field where there is no publicly available data and almost no academic literature to date. This allows complementing the few earlier works on the drivers of collection recoveries.

Second, we explicitly apply variable importance methods to the predictive characteristics. Using these methods, we show that the initially provided information is not only sparse but also insufficient for making precise predictions.

Third, and this is our most central contribution, debt collection agencies make use of a much more valuable set of information that is gathered from external sources, by repeated contact with the debtor, investigations into a debtor's financial situation, or by experience with a debtor in general.

The rest of the paper is structured as follows. In the second section, we present a short review of the available literature on debt collection agencies. The third section introduces our dataset and descriptive statistics. The fourth section presents the research design and the regression results. The fifth section describes the various checks of robustness that we use. The sixth section concludes this paper.

2 Related Literature

There is generally little theoretical and empirical work on the debt collection industry. Beck et al. (2017) study predictive characteristics of collections on a German debt collection dataset. They further find the average level of collections to be at around 65% and having a strong bimodal shape. Thomas et al. (2012) study differences in the characteristics of in-house and third-party collection claims on loans issued in the UK. They identify three main differences: There is less debtor and contract information in third-party debt collection, the claims are older, and the quality of claims is lower. They further identify several characteristics to be predictive in the two cases. Our own results make use of and extend these findings. Hoechstoetter et al. (2012) study the choice of prediction methods on an enormous German dataset with several million claims from nine different industries. Their results further show that mean recoveries vary tremendously across industries. We extend this literature by providing more evidence on the drivers of debt collection recoveries.

Beck et al. (2017) and Thomas et al. (2012) highlight that debtor and contract information provided by the initial creditor is sparse. These studies further consider sparsity of information to be a typical aspect of the debt collection business. This is further in line with the data in Hoechstoetter et al. (2012). Building on these important insights, we study how valuable additional gathered information is for recovery predictions.

The works of Makuch et al. (1992), Chin and Kotak (2006), and Miller et al. (2012) further document considerable improvements in the profitability of debt collection firms and departments by introducing operational innovations. These methods rely, in some part, on debtor and contract characteristics for making optimal workout process decisions. Identifying the key drivers of recoveries is essential in implementing these operational improvements. Han and Jang (2013) further document that choices on workout actions in a collection process for Korean loans explain the recovery success better than individual characteristics.

One very central paper underlying our argument is Drozd and Serrano-Padial (2017) which argues that a crucial factor in the expansion of consumer credit over the past decades is an improved information procession in the debt collection industry by the use of information technology. This improved information procession allows to focus more work on the debtors, which are more likely to pay. The enhanced efficiency of debt collection allowed an expansion of consumer credit, while the risk of consumer credit in terms of charge-off rates increased over this period. This argument is important, as it implies that debt collection agencies are critical in information production and procession. Fedaseyeu and Hunt (2018) argue that debt collection agencies are rather important, as they can use harsher debt collection actions compared to original creditors; however, the authors note that debt collection agencies, particularly, produce information on a debtor's willingness to pay. Our results shed more light on the question of, whether debt collection agencies, in fact, play a role in information production and procession. We evaluate how important this role is in terms of increased prediction accuracy.

Drozd and Serrano-Padial (2017) further show that debt collection has a positive influence on the credit supply. Several studies point in a similar direction. Fedaseyeu (2015) reports that more restrictive debt collection state laws in the US have a negative impact on the availability of collection services and via this channel on the credit supply. Fonseca et al. (2017) support these findings and, in addition, discover that this affects borrowers with a bad credit assessment, in particular. A better-informed regulation of the debt collection industry is, therefore, beneficial to users of goods and services.

3 Data Description

3.1 Debt Collection Dataset

Our dataset is provided by Seghorn AG, a large German debt collection agency, which mainly offers collection services in the insurance, banking, and mail order industries. The data consists of insurance receivables; original creditors transferred these claims on a commission basis. This means that the collection agency receives a compensation for the collection service, while the debtor payments and the claim itself remain in the possession of the initial holder.[2] The collection agency further guarantees a minimum average collection rate to the holder of the claim. This minimum average collection rate is contractually agreed on with the original creditor.[3]

There are two different types of receivables in our data. The first and main type is missed premium payments. Over 98% of our data consists of this category. The second category is recourse receivables. Recourse receivables result from insurance events where the insurance company makes payments but considers the insurance customer or a third party to be reliable for these payments. Regarding the premium payments, insurance companies could choose to withdraw from a contract (new contracts) or cancel a contract within certain periods (ongoing contracts). However, debt collection is regarded as a way to maintain the customer basis in a trade-off between retaining existing customers and investing marketing expenses to gain new customers. In cases, in which there is a payment, the insurance contract can just continue. In the contrary

2 The collection agency can charge the debtor with the compensation. Third-party debt collection is a legal service in the German Legal Services Act (Rechtsdienstleistungsgesetzt). The size of the compensation for this debt collection service must accord with the Lawyers' Compensation Act (Rechtsanwaltsverguetungsgesetz).

3 Only looking at the minimum average collection rate agreement, this could provide incentive to focus more work on large accounts. The individual compensation sets incentive to work out all claims where the expected payments cover at least the claim itself and possible variable workout costs.

case, the insurance company could terminate the contract later on. The minimum collection rate offered by the debt collection agency is further used by the insurance companies in pricing and accounting insurance contracts.

The dataset contains three subsamples from different insurers: A, B, and C. Samples A and B result from the same insurance product, while Sample C results from a different insurance product. As the samples were each transferred by different insurance companies, they were independent before the transfer to the collection agency. The collection agency received the claims over the years 2012 to 2014 on a regular basis, and all of these claims are towards debtors located in Germany. Debt collection processes usually comprise of written reminders and telephone calls, then possibly legal actions. The collection efforts of the original creditors are limited to written reminders and possibly telephone calls.

Regarding data cleansing, 250 claims with exposures missing or smaller than €5 and ten accounts with missing genders are removed. Twenty-two customer ages lower than fifteen are set to missing. The dataset used in the analysis contains 182,880 claims in Sample A, 16,623 claims in Sample B, and 126,615 claims in Sample C.

Our dataset contains a comprehensive body of contract and debtor characteristics that were provided by the collection agency. We further collect spatial and macroeconomic information from the following sources:[4] (1) unemployment rates: Federal Employment Agency (Bundesagentur fuer Arbeit); (2) postal codes (Postleitzahl) matched to counties (Landkreis) and independent cities (kreisfreie Staedte): OpenGeoDB; and (3) geographic coordinates of postal code areas: OpenGeoDB.

[4] The regional characteristics were collected by the authors. This, therefore, does not necessarily reflect the collection agency's practice.

3.2 Construction of the Dependent Variable

The data contains monthly collection payments for each individual account, starting from the date of transfer to the collection agency. Following Dermine and de Carvalho (2006) and Gürtler and Hibbeln (2013), we analyze payments over a standardized payment period. Accordingly, only accounts with this minimum number of months in debt collection are considered. This aims at ensuring consistency across the claims.[5] We use a standardized fixed period of around four years. The exact number of months and the level of collections are not explicitly stated here on request of the collection agency.

The collection rate is then calculated by dividing the sum of monthly payments over t time periods for an account i by its exposure at the time of transfer $t = 0$ (see Eq. 1). As a common step, workout expenses and collection fees paid by the debtor to the collection agency do not enter the calculation. Debtor payments first service the initial claim. Fees and possible late payment interest imposed by the collection agency are serviced last, and they are accounted separately.

$$CR_{i,t} = \frac{\sum_{j=1}^{t} Payment_{i,j}}{Initial\ exposure_{i,t=0}} \tag{1}$$

Table 1 presents the distribution of cumulative collections for each account over the repayment horizons for all three samples. The distribution has strong concentrations at the boundaries of full or no payment. This is similar to the strong bimodal distribution of collection recoveries found in Beck et al. (2017). Only around 5% and below of all accounts lie strictly between zero and one.

[Insert Table 1 here]

---

[5] Accounts that are more recent will tend to have smaller collection rates, on average, because there are less payment months. Moreover, using only accounts that are "closed" would introduce a bias because successful accounts are "closed" by definition, whereas unsuccessful accounts could remain "open" over several years.

3.3 Initially Provided Information

The sample contains a comprehensive set of independent variables. Table 2 presents descriptive statistics and definitions for the explanatory characteristics. We highlight additional important information in this section and Section 3.4. Table 2 indicates whether a variable belongs to the initial information or the subsequently obtained information.

[Insert Table 2 here]

The exposure size (EXP) is the amount in Euro that is due when the debt collection agency receives the claim. The age of the debtor (AGE) is the debtor age at that time. The dummy for insolvent accounts ($INS.acqu$) indicates whether the account was insolvent before the transfer to the debt collection agency. As in Beck et al. (2017), the corporate dummy ($FIRM$) indicates whether a debtor has the legal form of a corporation. An individual debtor can, therefore, be a private person, a private company, or a partnership. The telephone contact dummy (TEL) indicates whether there are known telephone contact details available. $PREMIUM$ indicates whether a claim results from a missed premium payment. The remaining cases are recourse receivables. The age of the account ($AGE.ACC$) is calculated as the difference between the beginning of the contract period and the time of transfer to the collection agency. In the case of Samples A and B, the insurance product generally requires an insurance fee payment before the beginning of the contract period; therefore, the claims are in arrears when the payment was not received until the beginning of the insurance period. In Sample C, fee payments could generally be acceptable shortly after the beginning of the insurance period. The payments are, however, expected to arrive at least closely after the beginning of the insurance period. A claim from Sample C is, therefore, considered to be in arrears when the payment is not received shortly after the beginning of the insurance term.

3.4 Accumulated Information

We collect four groups of accumulated information: Spatial information, external credit assessments, customer relationship information, and information on financial and non-financial assets.

Considering the spatial information, there are multiple reasons how regions could differ in the collection rate (refer to Fernandes and Artes, 2016 for a discussion of regional scoring). These include a higher proportion of steady jobs, stronger sources of income, differences in debt enforcement institutions, or also cultural factors that influence the willingness to service debt. We include three proxies for the quality of an area. The first two proxies are calculated as the mean collection rate over the accounts in a county/city ($CR.c$) and in a postal code area ($CR.p$). The mean is calculated over all other debtors without the respective one. For postal code areas with less than ten accounts of other debtors, the value is replaced by the mean in the closest postal code area with at least ten accounts to ensure reliable numbers. These proxies allow taking several sources of differences between regions into account. The mean collection rate by postal code area allows for a granular consideration of spatial differences. We further include the unemployment rate ($UNEMPL.c$) published by the Federal Employment Agency on a county level. It is particularly addressed at measuring spatial differences in the available sources of income. Beck et al. (2017) consider the unemployment rate on the federal state level and find a negative relation to the collection rate. Likewise, Bellotti and Crook (2012) distinguish between council/poor housing, suburban/wealthy, rural, and other areas and find these categories to be predictive of loss given default in consumer finance. We, therefore, expect the mean collection rates on the county and postal code level to be positively related to the collection rate and the unemployment rate to be negatively related to collection rates.

A very important source of external credit assessments are credit scores (SCORE.num). Drozd and Serrano-Padial (2017) list the easier provision of credit scores as one factor as to how information technology made debt collection processes more efficient over the past decades. Dierkes et al. (2013) study the importance of credit scores for assessing the default probability of small firms and find it to be essential.

The scores (*SCORE.num*) that we use are obtained from SCHUFA, which is a large supplier of individual and corporate credit assessments in Germany. According to self-reported numbers, it maintains data on more than 65 million individuals and more than 5 million firms. The score mainly relies on information regarding ongoing and past loans, bank accounts, mobile phone contracts, etc. Both the existence of these contracts and whether there were missed or delayed payments enter the score. The providers of the mentioned services transfer this information to SCHUFA. This is similar to the score outlined for Creditreform in Dierkes et al. (2013). Where applicable, relevant public credit information published by authorities, such as insolvencies, etc., further affect the score; however, soft public information, such as social media, is not taken into account. Spatial information is only used in cases, in which no other information is available (0.03% of scored entities and individuals). Information such as profession, income, property, marital status, or nationality are not reported to SCHUFA and do not enter the scores.[6]

As a comparison, FICO reports that the characteristics used in scoring consist of the following information[7]: 35% payment history and missed or delayed payments including insolvencies, etc.; 30% amount already owed; 15% length of credit history; 10% mix of credit products; and 10 % extent of recently obtained new credit. As the SCHUFA score is, as well, mainly calculated from information on existing contracts and delays on these contracts, these scores are comparable to FICO scores in the United States.

The credit bureau score is obtained by the debt collection agency in cases where there was no sufficient payment after written reminders and telephone calls. The score is obtained to make a decision whether

[6] Details on the information that enters the score are available on https://www.schufa.de.

[7] Refer to https://www.myfico.com for the composition of the score and more detailed descriptions of the individual components.

legal actions are considered promising. The score is not obtained earlier, as obtaining the score is costly. Obtaining the score is, therefore, indicative of a collection process in which initial collection actions did not result in sufficient payments and obtaining it will, as such, be related to lower collection rates. We discuss this and implications for our estimated coefficients in much detail in Section 4.2.3. For Sample B, the credit score was not obtained due to an agreement with the original creditor; therefore, there are no credit bureau scores in this sample. The reported credit score is the one that was last obtained. In cases with multiple accounts, it is, hence, possible that the score was obtained in the most recent collection process. We discuss this in the checks of robustness in Section 5. Furthermore, there may be some information in the score that is already captured in the insolvency dummies, the dummy for enforceable claims, and, arguably, in a dummy for single accounts, as (multiple) delays in payments will likely be reflected in the score. We discuss this in the robustness checks in Section 5, as well.

The levels of the score range from ’A’, which is the best score value, to ’M’, which is the worst. We use this score as a numerical variable by setting scores of ’A’ to 1, scores of ’B’ to 2, and so forth. We generally expect better scores to coincide with higher collection rates and the scores to have a high impact on the prediction quality.

Debt collection agencies further obtain information on debtors and their willingness to pay over the course of a customer relationship within one debt collection process or over processes with multiple claims. This could lead to a better understanding of how to effectively communicate with a customer or what extent of payments to expect. We add two variables that proxy for this. The first is a dummy whether the debtor only has one account in collection or whether there are multiple accounts (*SINGL.ACC*); the second is the average collection rate over all accounts of the same debtor except the respective account (*CR.OTHER*). The variables are calculated in-sample in the sense that they are calculated over observations within our dataset.[8]

There are reasonable arguments that could lead to the single account dummy having a positive relation to the collection rate, as in Thomas et al. (2012), or a negative relation, as in Beck et al. (2017); therefore, we do not expect a specific relation. We do expect, however, a strong positive coefficient for the collection rate on other accounts. One more piece of information that becomes available over time, as well, is a dummy for insolvencies during the debt collection process ($INS.proc$). We include this variable and expect a negative relation to the debt collection rate. Overall, when customer relationship information is influential, it should largely increase the prediction quality.

An important characteristic in deciding how to proceed with difficult cases is financial status information. Over the course of a debt collection process, a debt collection agency learns about whether a debtor possesses any financial or non-financial assets. This is important information in order to find agreements to settle the claim. In cases where a court declared a debt legally enforceable (we add $TITLE$ as a dummy for these cases), there is a formal process for obtaining financial status information.[9] We add a dummy for cases where this process is applied ($FIN.STATUS$). The financial status information is obtained from a bailiff which is costly.[10] The bailiff sets an appointment with the debtor to make a list of all financial and non-

8 The value of this information might, therefore, even be higher when calculating these measures out-of-sample for a larger extent of data.

9 There are two ways that the debt can become legally enforceable. In the first case, the creditor can request a judicial reminder to pay from a court (legal dunning proceedings). If the creditor does not dispute this reminder, the claim becomes enforceable. The second alternative is to sue the debtor. A creditor could sue the debtor directly without legal dunning proceedings. These steps are similar in the United States. Atradius (2018) gives a concise overview of legal debt collection actions in different countries.

10 There is a similar process for obtaining financial status information in the United Sates. Instead of a bailiff, an attorney produces the financial status. The debtor is also obliged to disclose all information on the financial status. Refer to Atradius (2018) for more details.

financial assets the debtor owns. The debtor is obliged to disclose this information correctly. In the case of payment, the debtor bears the costs of the bailiff; in cases without payments, the debt collection agency will not be compensated for the costs. The financial status information is, therefore, rather obtained for larger exposures.[11]

The overall effect of obtaining financial status information might be ambiguous, as there will be several unsuccessful collection actions before obtaining the financial status; therefore, obtaining financial status information indicates a difficult debt collection process. This could result in claims with financial status information being weaker than the overall sample. However, as the financial status information removes incentives for the debtor to withhold payment, the effect should be at least strongly positive in comparison to accounts that are in collection for an equally long time.

In summary, we look at four types of additionally gathered information. Among the spatial characteristics, we expect regions that display higher collection rates on other debtors and have lower unemployment rates to be predictive of higher collection rates. Credit bureau scores will contain important information regarding what payments to still expect. Considering customer relationship information, when debt collection agencies have multiple accounts, the success on other accounts will be predictive of the collection rate. Financial status information is valuable information in cases where there are no sufficient payments, even after several collection actions. Especially compared to other similarly difficult cases, financial status information should provide a better picture on how to effectively recover a claim. This should result in higher collection rates.

[11] We discuss implications for this in Section 4.2.5.

4 Analysis

4.1 Empirical Model and Research Design

We aim to study the value of information gathered in third-party debt collection in comparison with the information provided by original creditors. We examine first, whether the information extent of the initially disclosed information is in fact low. In doing so we also contrast our findings with findings of previous studies to extend the small knowledge on drivers of collection recoveries. Only the initial information enters this step.

In the second step, we analyze information that the debt collection agency gathers. Namely, we use the spatial information, the credit score, the customer relationship information, and the financial status information.

In empirical modeling, we use the fractional regression model of Papke and Wooldridge (1996) in our analysis, which is a common approach in modeling the recovery rate of bank loans (Ingermann et al., 2016, Dermine and de Carvalho, 2006). Furthermore, Beck et al. (2017) use this approach to model collection rates. All of the tables state standard errors that are clustered on the level of the year times the postal code. The coefficients are calculated as the conditional marginal effects, which are given by the change in the outcome of the collection rate for a change of one standard deviation for continuous and a change of one unit for dichotomous variables. The marginal effects are calculated for the continuous variables at their median and for the dichotomous variables at zero. In order to keep as much of the data as possible, we replace the missing variable values with the variable mean and add a dummy for missing variable values. This is done for the AGE, $AGE.ACC$, $SCORE.num$, and $CR.OTHER$ (dummy for missing $CR.OTHER$: $SINGL.ACC$).[12]

[12] As an alternative approach, we conduct our analysis excluding accounts with missing variable values (see Section 5). The

When talking of characteristics being ”main” or ”important” drivers of recoveries, we measure this in three different ways. Our first assessment is whether a characteristic has a statistically significant effect in the regression model, as outlined above. If this is the case, we assess whether the marginal effects indicate a large effect on the collection rate. As a third criterion, we assess the change in the adjusted R^2 resulting from including or excluding certain variables. In defining what relevant changes are in this setting, we refer to results from the recovery rate literature that is methodologically similar to the prediction of collection rates. Bellotti and Crook (2012) state adjusted values between 10.5% and 11.1% in a linear regression on a comprehensive set of debtor, contract, spatial, and macroeconomic characteristics; Gürtler and Hibbeln (2013) state values between 4.4% and 18.9%; Loterman et al. (2012) study six different loan datasets and find R^2 values between 1.2% and 44.12%. The R^2 is considerably higher in studies including information on collateral (see Ingermann et al., 2016: up to 76.9% or Qi and Yang, 2009: up to 61%). Given that these values not including collateral are often low, predicting the recovery rate is considered a difficult problem. An improvement of several percentage points represents an important contribution.

One strong aspect of our study is that we can compare results on different datasets that were independently managed before being handed over to third-party debt collection. This allows us to examine, in detail, whether results can be generalized over different portfolios. We, therefore, conduct our analysis on the three samples, individually, to exploit this important characteristic.

results remain qualitatively unchanged.

4.2 Regression Results

4.2.1 Initially Disclosed Information

Table 3 presents the regression results for the baseline model with the initially disclosed information. It is interesting to notice that most of the variables have a consistent relationship to the collection rate over all three samples, which applies to the firm dummy, the dummy for insolvencies before third-party debt collection, the telephone contact details, and the dummy for missed premium payments. The exposure size is not significant in each of the three samples, but the sign is consistent. The coefficients are less consistent for the customer age, the age of the account, and the male dummy.

[Insert Table 3 here]

The exposure size (*EXP*) has negative coefficients for all three samples. The coefficient is significant in Sample A and Sample C; for Sample A, an increase of the exposure of one standard deviation results in a 2.7 percentage points lower expected collection rate, and, in Sample C, the collection rate decreases stronger by 12.6 percentage points. This appears large, but, given that the standard deviation of the exposure size is relatively high compared to the mean values, this effect seems more relevant for particularly high exposures. Table 4 states the change in the adjusted R^2 when removing the exposure size from the three full models in Table 5, which is at less than one percentage point in the Samples A and B. The influence on the prediction quality is higher in Sample C, with a change of 1.46 percentage points, which is quite considerable at least in Sample C. The negative coefficients are in line with the results of Hoechstoetter et al. (2012) on collection recoveries and Thomas et al. (2012). Interestingly, Beck et al. (2017) find a positive relationship, arguing that this is due to incentives leading to a greater focus on large exposures. Our rationale is that larger exposures are more difficult to repay.

[Insert Table 4 here]

The firm dummy ($FIRM$) has a negative significant coefficient in all three samples. The effect in terms of the coefficients is considerable, as firms have collection rates that are lower by 5.2, 11.3, and 10.8 percentage points, respectively. This is in contrast to Beck et al. (2017) finding a positive relation to recoveries. Hoechstoetter et al. (2012) find inconsistent results.[13] The influence on the adjusted R^2 is small for all three samples (refer to Table 4).

The availability of telephone contact information (TEL) has a large positive relation to the collection rate. The effect is at around 28.7, 23.8, and 21.9 percentage points, respectively, in the three samples. A strong effect is further supported by the change in the adjusted R^2 (refer to Table 4). This is in line with Thomas et al. (2012) and Hoechstoetter et al. (2012) finding lower recoveries for worse contact information. This could result from an easier contact to debtors or better-maintained contact information being a signal of debtor quality.

Debtors who are insolvent before the transfer ($INS.acqu$) consistently have a far lower collection rate, which is intuitive and expected. The age of the account ($AGE.ACC$) is significantly negative in Samples A and B but positive in some models of Sample C (refer to Table 5). The relation in this sample is, hence, unclear, which is supported by the change in the adjusted R^2 (refer to Table 4). Beck et al. (2017) argues that a high age of the account value is informative of worse debtor quality, when more time in in-house collection leaves less recoveries to the third-party. This is in line with the empirical results in Beck et al. (2017). When controlling for more characteristics, our numbers support this argument. The coefficient for missed premium payment claims ($PREMIUM$) is significantly positive for Samples A and C, where there are two types of

[13] We analyze individuals and firms collectively in line with these studies. Robustness checks find similar driving characteristics for both groups (refer to Section 5).

claims. The variable importance in Table 4 is low in both samples.

The age of the customer and the male dummy have less consistent or insignificant coefficients over the three samples; however, the coefficient of the age (*AGE*) becomes significantly negative in all three samples when controlling for more information (refer to Table 5). It appears that older debtors tend to be less likely to pay. The dummy for male debtors (*MALE*) is mostly significant; however, the signs are inconsistent. There does not seem to be a clear and consistent relationship to the collection rate. The influence on the adjusted R^2 is small for both variables (refer to Table 4).[14]

The overall adjusted R^2 of the models built on the initially disclosed set of information is between 9.4% and 14.1%. Taking the results of Loterman et al. (2012), which state values up to 44.12%, or the results of Qi and Yang (2009), which state values up to 61%, as a benchmark, these values are rather low. In summarizing the individual coefficients, we can confirm a positive relation to better contact information and a negative relation to the firm dummy. Some predictive characteristics seem to be unreliable, such as the gender dummy. When taking more characteristics into account, older debtors and claims that remained longer in in-house collection have lower collection rates. Larger exposures tend to be related to lower collection rates, as well. Most of the initially provided variables individually tend to contribute little to the prediction quality. The overall adjusted R^2 only including the initially available information is also moderate.

[14] The results on age, gender dummy, and firm dummy need to be considered with some caution because firms do not have age and gender values, and they are much less likely to have credit scores, which might induce multicollinearities. However, the results remain qualitatively unchanged when estimating results for individuals and firms separately (Section 5). Furthermore, the dummy for firms remains significantly negative when excluding age, gender, and credit score (available on request).

4.2.2 Spatial Characteristics

The results, including the spatial characteristics, are stated in the first column of the three sections of Table 5. All three characteristics are significant in the Samples A and C, and, in Sample B, the unemployment rate (*UNEMPL.c*) is significant.[15] In line with the idea of a regional scoring, the mean collection rates in the county and the postal code area have a positive coefficient. The unemployment rate, as a presumed measure negatively related to the income of the local population, has negative coefficients. The coefficients indicate a change of the dependent variable between 0.3 and 3.3 percentage points for a standard deviation change of the spatial characteristic. Assessing the individual influence on the adjusted R^2 in Table 4, all three variables are rather weak, and the overall improvement in the prediction quality is relevant but rather moderate. The increase in the adjusted R^2 is at 0.9, 0.1, and 0.3 percentage points in Samples A, B, and C, respectively.

Given that we analyze characteristics down to the level of the postal code area, the prediction quality might improve beyond the values outlined here when building even more fine-grained spatial scores. More data further seems to improve the value of spatial scores, as reflected by the increase in the prediction quality in line with the sample size from Samples B and C to A.

[Insert Table 5 here]

4.2.3 External Credit Assessment

The credit bureau score (*SCORE.num*) is added in the second column. In Sample A, a credit score that is worse by one standard deviation decreases the collection rate by 10.2 percentage points; in Sample C,

---

[15] The mean collection rate in the county (*CR.c*) and the mean collection rate in a postal code area (*CR.p*) individually become significant when removing the two other variables but are insignificant when controlling for them (available on request).

the collection rate decreases by 11.7 percentage points for a credit score that is weaker by one standard deviation. Considering that even the worst levels of the score appear regularly and that the score has a wide range of values, this is a very considerable variation in the collection rate.[16] The dummy for missing scores has a significant positive coefficient in Samples A and C. This is consistent with a collection process that was difficult before obtaining the score.

The strong relation of the score to the collection rate is further emphasized by the overall increase in the adjusted R^2 in Table 5 from Column 1 to Column 2. The values increase by 11.9 percentage points in Sample A and 10.4 percentage points in Sample C, which is very sizable, especially considering the general level of the adjusted R^2.

Obtaining the credit bureau score does, in a sense, already contain some information of the pace of the collection process that is ex-post knowledge and is not, in fact, available in a prediction context. Therefore, we additionally calculate the change in the adjusted R^2 for removing the credit bureau score from a model estimated only on accounts with an obtained credit bureau score. These results are included in Table 4. The adjusted R^2 is, overall, lower on this subset in Sample A (39.9%) and is higher on this subset in Sample C (37.3%). In Sample A, the change in the adjusted R^2 for excluding the score is about half as high compared to the values including accounts with no score, indicating that the lack of a score contains some information of the pace of the collection process. The difference is, however, still considerable. In Sample C, the change in the adjusted R^2 is lower but remains in a similar magnitude compared to the earlier results, which is in line with more unconditionally obtained credit scores.

[16] When including dummies for the credit score levels to replace the numerical score variable, it is noteworthy that the dummy coefficients are strictly monotonic over all of the levels of the score both in Sample A and Sample C. (These additional regression tables are available on request.) This further points towards a strong discriminatory power of the credit score.

4.2.4 Experience with the Customer

The third column includes the outcome of other collection processes with the debtor ($CR.OTHER$) and further includes information on insolvencies during the respective collection process ($INS.proc$). The mean collection rate on other accounts ($CR.OTHER$) has a significantly positive coefficient over all three samples. Given an increase of one standard deviation, the collection rate increases by 25.3, 1.9, and 6.2 percentage points, respectively; the effect is particularly large in Sample A. Including the dummy for debtors becoming insolvent during the third-party collection process decreases the collection rate significantly by 56.3, 15.4, and 13.8 percentage points, respectively.

From the second to the third columns, the adjusted R^2 increases by 15.9, 6.1, and 6.9 percentage points in Samples A, B, and C, respectively. In the larger samples, the majority of this increase results from payments on other accounts. The influence on the adjusted R^2 of the dummy for insolvencies during the collection process in Table 4 is at 1.8, 4.1, and 0.8 percentage points, respectively.

4.2.5 Financial Status Information

Obtaining a financial status is very important information in cases that are particularly difficult. The dummy for accounts where this information ($FIN.SATUS$) is obtained is added in the fourth columns of Table 5. As the financial status could only be obtained when there is already a court order to pay, there is a dummy for this ($TITLE$) as well as a control.

The dummy for the court order to pay has a strong negative and significant coefficient in all three samples, particularly in Sample A. The importance is also rather high when considering Table 4. When assessing the coefficient of the financial status dummy, it is significantly positive and considerable in Sample C, slightly negative in Sample A, and insignificant in Sample B. The overall influence on the adjusted R^2 in Table 4 is also relatively low in all three samples. However, the results do not capture the entire impact of the financial status information when comparing the cases with this information to the overall sample.

Therefore, we apply a matching procedure pairing all financial status information accounts with a similar account without financial status information that was in the third-party debt collection process for a similar amount of time without being fully paid. We perform a matching on the exposure size and the credit score, as the exposure size plays some role in the decision to obtain the financial status information and we want to control for differences in solvability. The exposure size is matched using a Mahalanobis metric, and the credit score is matched exactly.

Table 6 presents the matching results.[17] The lower part of the table lists the means of the exposure size and the credit score. The results show that the financial status information sample and the matched sample are statistically similar in the exposure size using t-tests. The credit score is matched exactly. The number of pairs indicates that there is a match for almost all the cases with financial status information (8.500, 2.574, and 8.053, respectively). The top line of the table lists the group means of the collection rate. The differences in the mean are quite considerable in all three samples but are particularly large in Sample B.

In order to have a representation of the variable importance in terms of the R^2, we calculate a regression model, including the full list of variables from Table 5 after combining the set of the accounts with financial status information and the related matches. When excluding the financial status dummy, the R^2 decreases by 5, 11, and 2.7 percentage points, respectively.[18]

[17] We produced several alternative specifications, including the full list of variables from Table 5 and several subsets of it. The results are in a similar magnitude. These specifications are available on request.

[18] These results are available on request.

[Insert Table 6 here]

In summing up the results, the change from the initial adjusted R^2 values to the values including the characteristics gathered after the time of transfer is quite noteworthy. The overall increases sum up to 31.2, 8.9, and 17.9 percentage points for Samples A, B, and C, respectively. In Samples A and C, the increase exceeds the initial adjusted R^2 values, and, in Sample B, the increase is at about the same magnitude as the initial level. We infer that a major part of the quality of predictions relies on information that needs to be gathered later rather than the information provided initially. The value of the initially provided information is moderate compared to the R^2 values typically stated in recovery rate studies. The levels of the full model, including all gathered information, is considerable vice-versa typical levels stated in these studies. Assessing the characteristics provided initially, we find a negative effect of the exposure size and, possibly, the age of the account. The availability of telephone contact details has a positive relation to the collection rate and older customers and firms tend to have lower collection rates. Among the gathered information, better spatial areas and accounts with successful other collection processes have higher collection rates. The collection rate is much lower for worse credit scores. Financial status information is very important in convincing difficult debtors to pay.

5 Robustness Checks

In this section, we present several checks of robustness. We estimate results for shorter standardized fixed payment periods. We then estimate results excluding missing variable values, cases with hard negative credit information, and multiple accounts. In a last step, we examine whether there are differences between individual and corporate debtors.

In Section 3.2, we calculate the collection rate over a uniform payment period of about four years. This period is chosen to estimate results on a uniform and in particular long payment period to assess results on ultimate collection recoveries (Hoechstoetter et al. (2012) discuss the choice of payment periods but use shorter periods compared to our data). However, to show that our results are not dependent on this specific choice of the payment period, we replicate the results from Section 4.2 using two different payment periods. These periods, respectively, are half as long and a quarter as long as the main payment period (refer to Fig. 1 and Table 7 in the Appendix; further results for other payment horizons are available on request). These additional analyses also make results more comparable to possible studies using shorter debt collection processes.

Looking at the results, a small number of coefficients become insignificant. The premium payment dummy further becomes significantly negative. The financial status information has a consistent negative coefficient in the shortest payment period, which is reasonable, as this action is rather applied in late stages of the debt collection process. This is reflected in the fact that the coefficients become less negative or positive in the longer of the two short collection periods. The other results remain qualitatively unchanged.[19]

The dataset contains many accounts with missing characteristics. This is a typical feature in debt collection due to the information provided by the in-house collection department. We treat these cases by including dummies for missing values and setting the missing values to the variable mean. To make doubly sure this does not affect the results, we fit the regression models by only including complete accounts (Table 8 in the Appendix). The results remain qualitatively unchanged.

[19] When assessing Fig. 1, it is further interesting to notice that the mean collection rate depends on the length of the collection process. According to statements of practitioners, knowledge on what recoveries to ultimately expect is an important asset for collection agencies gained over experience with many cases and portfolios. This knowledge is usually not available to original creditors.

In the three samples, some debtors are linked to multiple accounts. This could, for example, be problematic in the case of the credit score, as we only observe the latest score and the score may have changed from requesting it in an earlier case to a later case. Therefore, we further estimate the results from Section 4.2 by only including accounts of one-time debtors. This is done excluding all accounts in which the debt collection agency has at least one other case of the same debtor in the database. Table 8 in the Appendix presents these results. The results remain qualitatively unchanged. The coefficient for the credit score is more negative than in the case including multiple accounts in Sample C and remains similar in Sample A. When excluding the credit score from this sample, the adjusted R^2 decreases by 5.9 and 6.1 percentage points, respectively.

Credit scores could further contain publicly available information that is also included in other independent variables; this applies to insolvencies and enforceable claims. Therefore, we estimate one full model excluding the cases with publicly available negative credit information in addition to cases with multiple accounts. The results are presented in Table 9. The coefficients of the credit score become smaller but remain in a considerable magnitude. When excluding the credit score from this sample, the adjusted R^2 decreases by 3.1 and 4.2 percentage points, respectively. The variable importance of the credit score is, therefore, still strong.

Our samples contain both individual and corporate debtors. Given that it could be argued that these groups might be driven by different factors, we estimate the regression models from Section 4.2 for both groups separately. The results are presented in Table 10 in the Appendix. For the individual debtors that account for the majority of cases, all of the results remain qualitatively unchanged. For the corporate debtors, the exposure size becomes insignificant in Sample A and becomes positive in Sample B. One could hypothesize that the specific exposure amount is less informative for corporations that could largely differ in size. The spatial characteristics also lose some significance and the coefficient for the financial status is also less consistent. However, doing a similar matching as in Section 4.2.5, the results are qualitatively unchanged.[20]

6 Conclusion

Delegating the collection of distressed receivables to debt collection agencies is common practice in many industries, but, surprisingly, very little is known about the successful management of receivables in third-party debt collection. Earlier work has, furthermore, mainly studied information provided at the beginning of debt collection processes. Recent theory suggests that producing additional information is an important function of the debt collection industry. This work, therefore, studies how valuable information gathered in third-party debt collection is for predictions of collection rates.

We add three important contributions to the literature:

(1) We collect and study a large proprietary debt collection dataset. This allows us to complement the few earlier works on the drivers of collection recoveries.

(2) Using variable importance methods, we empirically support the notion that the information usually provided by the initial creditor has a limited share in explaining the variation in collection rates.

(3) Using our extensive set of variables, we show that the information that is gathered in third-party debt collection plays a very crucial role in making predictions more reliable. This finding supports the theoretical argument that debt collection agencies perform a function of producing information.

Our results, therefore, empirically complement the theoretical picture of debt collectors that are active in

[20] The difference in means are significant. The exact values are 0.106, 0.1, and 0.148, respectively. These results are available on request.

information production. Our results are important for a better understanding as well as, in this way, a better regulation and management of the debt collection industry.

Appendix

[Insert Figure 1 here]

[Insert Table 7 here]

[Insert Table 8 here]

[Insert Table 9 here]

[Insert Table 10 here]

[Insert Table 11 here]

References

Atradius (2018). International debt collections handbook. Technical report.

Beck, T., J. Grunert, W. Neus, and A. Walter (2017). What determines collection rates of debt collection agencies? Financial Review 52(2), 259–279.

Bellotti, T. and J. Crook (2012). Loss given default models incorporating macroeconomic variables for credit cards. International Journal of Forecasting 28(1), 171–182.

Bülow, S. (2016). Update Branchenstudie Inkasso: Neue Zahlen, Daten und Fakten über das Forderungsmanagement. Zeitschrift für das Forderungsmanagement (6/2016), 223–226.

Chin, A. G. and H. Kotak (2006). Improving debt collection processes using rule-based decision engines: A case study of Capital One. International Journal of Information Management 26(1), 81–88.

Dermine, J. and C. N. de Carvalho (2006). Bank loan losses-given-default: A case study. Journal of Banking & Finance 30(4), 1219–1243.

Dierkes, M., C. Erner, T. Langer, and L. Norden (2013). Business credit information sharing and default risk of private firms. Journal of Banking & Finance 37(8), 2867–2878.

Drozd, L. A. and R. Serrano-Padial (2017). Modeling the revolving revolution: The debt collection channel. American Economic Review 107(3), 897–930.

Ernst & Young (2017). The impact of third-party debt collection on the US national and state economies in 2016. Technical report.

Fedaseyeu, V. (2015). Debt collection agencies and the supply of consumer credit. Working paper, Research Department, Federal Reserve Bank of Philadelphia (15-23).

Fedaseyeu, V. and R. M. Hunt (2018). The economics of debt collection: Enforcement of consumer credit contracts. Working paper, Research Department, Federal Reserve Bank of Philadelphia (18-4).

Fernandes, G. B. and R. Artes (2016). Spatial dependence in credit risk and its improvement in credit scoring. European Journal of Operational Research 249(2), 517–524.

Fonseca, J., K. Strair, and B. Zafar (2017). Access to credit and financial health: Evaluating the impact of debt collection. Staff Report, Federal Reserve Bank of New York (814).

Gürtler, M. and M. Hibbeln (2013). Improvements in loss given default forecasts for bank loans. Journal of Banking & Finance 37(7), 2354–2366.

Han, C. and Y. Jang (2013). Effects of debt collection practices on loss given default. Journal of Banking & Finance 37(1), 21–31.

Hoechstoetter, M., A. Nazemi, S. T. Rachev, and C. Bozic (2012). Recovery rate modelling of non-performing consumer credit using data mining algorithms. RMI Working Paper - National University of Singapore (12/09).

Ingermann, P.-H., F. Hesse, C. Bélorgey, and A. Pfingsten (2016). The recovery rate for retail and commercial customers in Germany: A look at collateral and its adjusted market values. Business Research 9(2), 179–228.

Loterman, G., I. Brown, D. Martens, C. Mues, and B. Baesens (2012). Benchmarking regression algorithms for loss given default modeling. International Journal of Forecasting 28(1), 161–170.

Makuch, W. M., J. L. Dodge, J. G. Ecker, D. C. Granfors, and G. J. Hahn (1992). Managing consumer credit delinquency in the US economy: A multi-billion dollar management science application. Interfaces 22(1), 90–109.

Miller, G., M. Weatherwax, T. Gardinier, N. Abe, P. Melville, C. Pendus, D. Jensen, C. K. Reddy, V. Thomas, J. Bennett, G. Anderson, and B. Cooley (2012). Tax collections optimization for New York State. Interfaces 42(1), 74–84.

Papke, L. E. and J. M. Wooldridge (1996). Econometric methods for fractional response variables with an application to 401 (K) plan participation rates. Journal of Applied Econometrics 11(6), 619–632.

Qi, M. and X. Yang (2009). Loss given default of high loan-to-value residential mortgages. Journal of Banking & Finance 33(5), 788–799.

Thomas, L. C., A. Matuszyk, and A. Moore (2012). Comparing debt characteristics and LGD models for different collections policies. International Journal of Forecasting 28(1), 196–203.

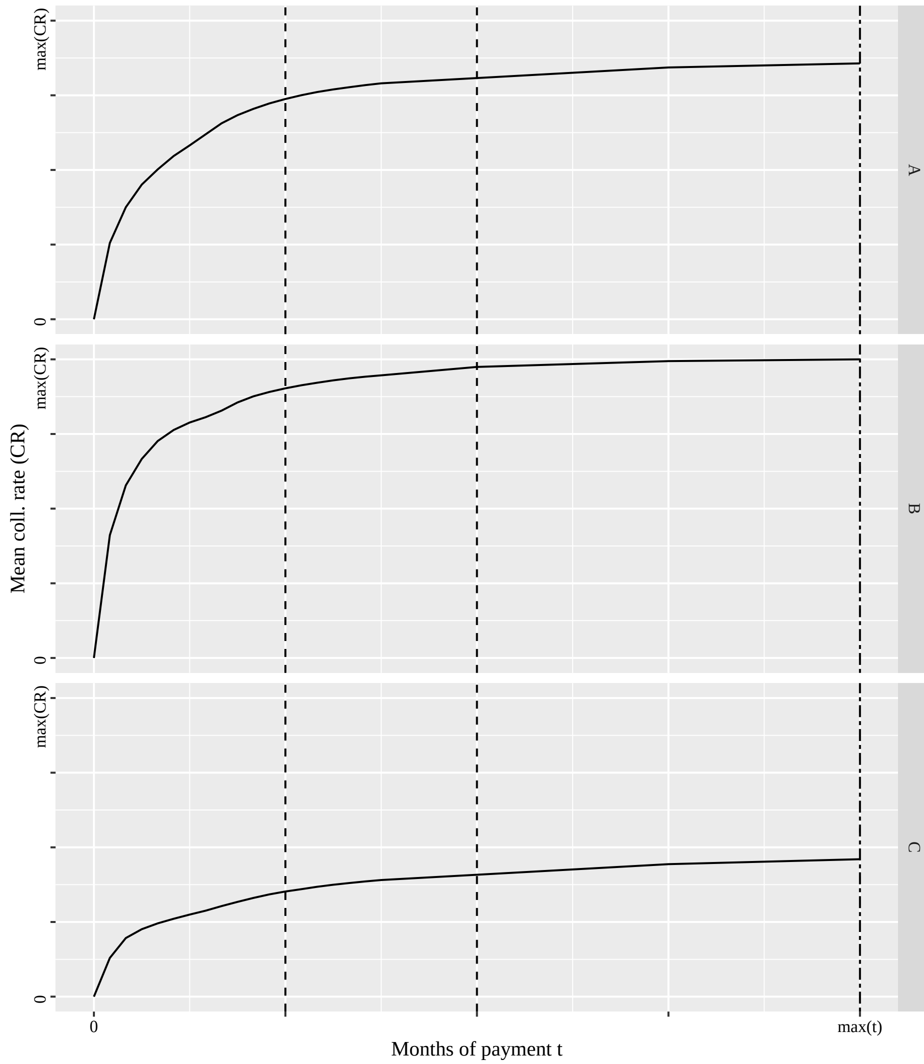

Figure 1

Overview robustness checks: collection period

Figure 1 displays the robustness checks for the choice of the payment period. For each sample, there is one vertical line on the right side of the figure indicating the position of the baseline payment period. There are two more vertical lines for shorter payment periods on the left which we use in the checks of robustness. The results of the checks of robustness are displayed in Table 7. The plot is true to scale.

Table 1

Distribution of the collection rate to boundary and non-boundary cases

		Sample		
Interval		**A**	**B**	**C**
Full or no payment	($CR = 0$ & $CR = 1$)	0.944	0.977	0.959
Partial payment	($0 < CR < 1$)	0.056	0.023	0.041

Table 1 displays the fraction of the accounts that yield a cumulative partial, full, or no payment. Full and non-payments are combined to conceal the specific level of recoveries.

Table 2

Summary statistics

Numerical Characteristics

	Initial	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Sample A									
EXP: Exposure size handed over to debt collection	Yes	182,880	200.996	762.768	6.41	61.25	99.31	185.46	>100,000
AGE: Age of the debtor	Yes	172,118	44.927	12.798	15	35	45	53	110
AGE.ACC: Time in in-house collection	Yes	182,698	145.112	93.378	−21	87	120	170	4,456
UNEMPL.c: County unemployment rate	No	182,880	7.226	2.970	1.1	5	6.6	9.2	18.5
SCORE.num: Numerical credit score	No	49,472	8.599	3.618	1	6	9	12	13
Sample B									
EXP: Exposure size handed over to debt collection	Yes	16,623	173.578	298.542	5	57.81	98.58	197.55	>15,000
AGE: Age of the debtor	Yes	6,979	45.309	13.279	18	35	45	54	96
AGE.ACC: Time in in-house collection	Yes	10,501	131.809	62.911	35	105	108	144	1,904
UNEMPL.c: County unemployment rate	No	16,623	6.872	2.912	1.1	4.5	6.5	8.8	16.5
SCORE.num: Numerical credit score	No	0	—	—	—	—	—	—	—
Sample C									
EXP: Exposure size handed over to debt collection	Yes	126,015	352.285	828.646	5	39.765	114.1	330.17	>50,000
AGE: Age of the debtor	Yes	116,174	42.376	12.441	15	32	41	51	101
AGE.ACC: Time in in-house collection	Yes	102,449	170.367	212.956	−146	66	106	234	7,879
UNEMPL.c: County unemployment rate	No	126,015	7.333	3.177	1.1	4.7	7	9.8	18.5
SCORE.num: Numerical credit score	No	76,498	8.504	4.030	1	5	9	12	13

		Sample A	Sample B	Sample C
	Initial	N	N	N
CR.OTHER: Mean coll. rate on other accounts	No	113,588	1,809	34,267
CR.c: Mean coll. rate in the same county	No	182,880	16,623	126,015
CR.p: Mean coll. rate in the same postal code area	No	182,880	16,623	126,015

Dichotomous Characteristics

		Sample A	Sample B	Sample C
	Initial	True	True	True
INS.acqu: Insolvency before third-party collection	Yes	0.019	0.020	0.049
FIRM: Legal form of a corporation	Yes	0.046	0.069	0.060
MALE: Male debtor	Yes	0.647	0.658	0.642
TEL: Telephone contact details available	Yes	0.624	0.419	0.396
PREMIUM: Premium payment	Yes	0.992	1.000	0.981
INS.proc: Insolvency in third-party collection	No	0.023	0.017	0.035
TITLE: Legally enforceable claim	No	0.386	0.259	0.388
FIN.STATUS: Financial status information	No	0.046	0.154	0.064

The three upper sections of Table 2 report summary statistics for the numerical variables in the collection samples. CR.OTHER, CR.c and CR.p are separately listed in the fourth section to conceal the specific level of recoveries. The lower section reports the relative frequencies of true values for the dichotomous variables. "N" is the number of non-missing values, "Mean" the mean, "St. Dev." the standard deviation, "Min" the minimum, "Pctl(25)", "Median", and "Pctl(75)" the first, second and third quartile, and "Max" the maximum. "Initial" indicates, whether a characteristic is provided by the initial creditor.

Table 3

Regression results - initially disclosed information

	DV: Collection rate		
	Sample A	Sample B	Sample C
EXP	-0.027^{***}	-0.007	-0.126^{***}
	(0.003)	(0.005)	(0.003)
AGE	0.016^{***}	-0.014^{***}	-0.006^{***}
	(0.002)	(0.004)	(0.001)
AGE.NA	-0.008	0.058^{***}	0.045^{***}
	(0.013)	(0.011)	(0.008)
MALE	-0.003	-0.020^{**}	0.006^{**}
	(0.004)	(0.009)	(0.002)
FIRM	-0.052^{***}	-0.113^{***}	-0.108^{***}
	(0.016)	(0.017)	(0.011)
INS.acqu	-0.370^{***}	-0.386^{***}	-0.096^{***}
	(0.009)	(0.021)	(0.005)
TEL	0.287^{***}	0.238^{***}	0.219^{***}
	(0.003)	(0.011)	(0.002)
AGE.ACC	-0.091^{***}	-0.045^{***}	0.014^{***}
	(0.002)	(0.007)	(0.002)
AGE.ACC.NA	-0.267^{***}	-0.063^{***}	-0.008^{***}
	(0.046)	(0.007)	(0.003)
PREMIUM	0.149^{***}		0.083^{***}
	(0.020)		(0.013)
R^2 adj.	0.114	0.094	0.141
Chi2	10,242.51	999.67	10,624.10
Chi2 Prob.	0.000	0.000	0.000
Observations	182,880	16,623	126,015

*p<0.1; **p<0.05; ***p<0.01; Table 3 displays the fractional regression results for the Samples A, B, and C. For each sample, the results are presented for including only the information provided to the collection agency initially. The table reports standard errors in brackets. The standard errors are clustered by year and postal code. The coefficients are calculated as the conditional marginal effects, which are given by the change in the outcome of the collection rate for a change of one standard deviation for continuous and a change of one unit for dichotomous variables. The marginal effects are calculated for the continuous variables at their median and for the dichotomous variables at zero.

Table 4

Change in the adjusted R^2 for removing individual characteristics

	Sample		
	A	**B**	**C**
Full model			
Baseline R^2	0.4262	0.1834	0.3201
Excluded variable (ΔR^2)			
EXP	-0.0003	-0.0000	-0.0146
AGE	-0.0008	-0.0011	-0.0018
MALE	-0.0000	-0.0002	-0.0010
FIRM	-0.0001	-0.0037	-0.0001
INS.acqu	-0.0069	-0.0356	-0.0017
TEL	-0.0132	-0.0263	-0.0479
AGE.ACC	-0.0033	-0.0049	0.0000
PREMIUM	0.0000		0.0000
CR.c	-0.0002	0.0000	-0.0002
UNEMPL.c	-0.0001	-0.0011	-0.0004
CR.p	-0.0006	-0.0001	-0.0001
SCORE	-0.0279		-0.0512
INS.proc	-0.0177	-0.0414	-0.0076
CR.OTHER	-0.1205	-0.0145	-0.0591
TITLE	-0.0247	-0.0154	-0.0032
FIN.STATUS	-0.0000	-0.0001	-0.0010
Full model - non-missing SCORE			
Baseline R^2	0.3984		0.3733
Excluded variable (ΔR^2)			
SCORE	-0.0444		-0.0884
Full model - TITLE			
Baseline R^2	0.3921	0.1077	0.3358
Excluded variable (ΔR^2)			
FIN.STATUS	-0.0000	-0.0004	-0.0013

Table 4 displays the change in the adjusted R^2 for excluding individual characteristics from the full models corresponding to Table 5. The first section of the table lists the baseline adjusted R^2 in the first row and the change in the adjusted R^2 when excluding the respective variable in the lines below. The second section of the table first states the baseline adjusted R^2 for a full model on accounts with non-missing credit scores. The second value in this section indicates by how much the adjusted R^2 of this model drops for excluding the score. The third section reports numbers for enforceable claims and for excluding the financial status information from this model.

Table 5

Regression results - stepwise addition of gathered characteristics

	DV: Collection rate										
	Sample A				Sample B			Sample C			
	(1)	(2)	(3)	(4)	(1&2)	(3)	(4)	(1)	(2)	(3)	(4)
EXP	-0.029^{***}	-0.032^{***}	-0.030^{***}	-0.016^{***}	-0.007	-0.001	0.0004	-0.126^{***}	-0.100^{***}	-0.063^{***}	-0.080^{***}
	(0.003)	(0.003)	(0.003)	(0.002)	(0.005)	(0.001)	(0.001)	(0.003)	(0.003)	(0.002)	(0.002)
AGE	0.013^{***}	0.002	-0.006^{***}	-0.010^{***}	-0.014^{***}	-0.005^{***}	-0.005^{***}	-0.007^{***}	-0.014^{***}	-0.010^{***}	-0.014^{***}
	(0.002)	(0.002)	(0.002)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
AGE.NA	-0.007	-0.117^{***}	-0.077^{***}	-0.063^{***}	0.058^{***}	0.009^{***}	0.004	0.046^{***}	-0.016^{**}	-0.014^{***}	-0.021^{***}
	(0.013)	(0.012)	(0.011)	(0.007)	(0.011)	(0.003)	(0.002)	(0.008)	(0.008)	(0.005)	(0.007)
MALE	-0.004	0.022^{***}	0.014^{***}	0.006^{**}	-0.021^{**}	-0.006^{**}	-0.004^{*}	0.005^{**}	0.031^{***}	0.020^{***}	0.026^{***}
	(0.004)	(0.003)	(0.003)	(0.002)	(0.009)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
FIRM	-0.048^{***}	-0.026^{*}	-0.029^{**}	-0.035^{***}	-0.112^{***}	-0.025^{***}	-0.027^{***}	-0.108^{***}	-0.083^{***}	-0.022^{***}	-0.033^{***}
	(0.016)	(0.015)	(0.013)	(0.009)	(0.017)	(0.005)	(0.004)	(0.011)	(0.010)	(0.006)	(0.008)
INS.acqu	-0.368^{***}	-0.352^{***}	-0.314^{***}	-0.230^{***}	-0.383^{***}	-0.112^{***}	-0.103^{***}	-0.097^{***}	-0.035^{***}	-0.036^{***}	-0.059^{***}
	(0.009)	(0.009)	(0.010)	(0.007)	(0.021)	(0.006)	(0.005)	(0.005)	(0.005)	(0.003)	(0.005)
TEL	0.284^{***}	0.209^{***}	0.174^{***}	0.103^{***}	0.235^{***}	0.063^{***}	0.045^{***}	0.220^{***}	0.183^{***}	0.115^{***}	0.160^{***}
	(0.003)	(0.003)	(0.003)	(0.002)	(0.011)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
AGE.ACC	-0.089^{***}	-0.069^{***}	-0.060^{***}	-0.027^{***}	-0.045^{***}	-0.012^{***}	-0.008^{***}	0.013^{***}	0.003^{*}	0.001	0.0004
	(0.002)	(0.002)	(0.002)	(0.001)	(0.007)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
AGE.ACC.NA	-0.265^{***}	-0.252^{***}	-0.222^{***}	-0.097^{***}	-0.060^{***}	-0.017^{***}	-0.006^{***}	-0.006^{**}	0.0005	-0.001	0.002
	(0.047)	(0.045)	(0.047)	(0.029)	(0.007)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
PREMIUM	0.149^{***}	0.058^{***}	0.042^{**}	-0.0003				0.079^{***}	0.021^{*}	0.012^{*}	-0.017
	(0.020)	(0.019)	(0.020)	(0.012)				(0.012)	(0.011)	(0.007)	(0.010)
CR.c	0.025^{***}	0.021^{***}	0.016^{***}	0.010^{***}	0.005	0.001	0.001	0.015^{***}	0.011^{***}	0.005^{***}	0.007^{***}
	(0.002)	(0.002)	(0.002)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
UNEMPL.c	-0.011^{***}	-0.008^{***}	-0.007^{***}	-0.005^{***}	-0.014^{***}	-0.004^{***}	-0.004^{***}	-0.006^{***}	-0.008^{***}	-0.006^{***}	-0.009^{***}
	(0.002)	(0.002)	(0.002)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
CR.p	0.033^{***}	0.027^{***}	0.019^{***}	0.012^{***}	0.002	0.0003	-0.0002	0.006^{***}	0.005^{***}	0.003^{***}	0.004^{***}
	(0.002)	(0.002)	(0.002)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
SCORE.num		-0.102^{***}	-0.084^{***}	-0.053^{***}					-0.117^{***}	-0.069^{***}	-0.087^{***}
		(0.002)	(0.002)	(0.001)					(0.001)	(0.001)	(0.001)
SCORE.num.NA		0.316^{***}	0.267^{***}	0.097^{***}					0.116^{***}	0.067^{***}	0.075^{***}
		(0.004)	(0.004)	(0.003)					(0.002)	(0.002)	(0.002)
INS.proc			-0.563^{***}	-0.350^{***}		-0.154^{***}	-0.127^{***}			-0.138^{***}	-0.188^{***}
			(0.011)	(0.007)		(0.009)	(0.007)			(0.005)	(0.006)
SINGL.ACC			-0.239^{***}	-0.152^{***}		-0.075^{***}	-0.054^{***}			0.085^{***}	0.115^{***}
			(0.003)	(0.002)		(0.007)	(0.005)			(0.002)	(0.003)
CR.OTHER			0.253^{***}	0.157^{***}		0.019^{***}	0.014^{***}			0.062^{***}	0.085^{***}
			(0.002)	(0.001)		(0.001)	(0.001)			(0.001)	(0.001)
TITLE				-0.158^{***}			-0.041^{***}				-0.059^{***}
				(0.003)			(0.003)				(0.003)
FIN.STATUS				-0.008^{*}			0.002				0.063^{***}
				(0.005)			(0.003)				(0.004)
R^2 adj.	0.123	0.242	0.401	0.426	0.095	0.156	0.183	0.144	0.248	0.317	0.32
Chi2	11,003.46	18,634.63	34,461.32	36,089.69	1,011.27	1,441.63	1,810.45	10,918.27	19,003.04	22,800.39	23,158.96
Chi2 Prob.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	182,880	182,880	182,880	182,880	16,623	16,623	16,623	126,015	126,015	126,015	126,015

*p<0.1; **p<0.05; ***p<0.01; Table 5 displays the fractional regression results for the Samples A, B, and C. For each sample, the results are presented for a stepwise inclusion of information gathered after the beginning of the third-party collection process: Spatial characteristics (1), credit bureau scores (2), customer relationship information (3), and financial status information (4). The table reports standard errors in brackets. The standard errors are clustered by year and postal code. The coefficients are calculated as the conditional marginal effects, which are given by the change in the outcome of the collection rate for a change of one standard deviation for continuous and a change of one unit for dichotomous variables. The marginal effects are calculated for the continuous variables at their median and for the dichotomous variables at zero.

Table 6

Matching results - financial status information

	Matching Results								
	Sample A			Sample B			Sample C		
	FIN.STATUS		Diff.	FIN.STATUS		Diff.	FIN.STATUS		Diff.
Variable	No	Yes	(t-value)	No	Yes	(t-value)	No	Yes	(t-value)
CR	—	—	0.233***	—	—	0.416***	—	—	0.210***
			(32.977)			(35.339)			(31.204)
EXP	241.662	241.556	-0.106	184.196	184.227	0.031	556.83	556.849	0.02
			(-0.018)			(0.006)			(0.002)
SCORE.num	6.882	6.882	exact				7.193	7.193	exact
N		8,279			2,574			7,863	

Table 6 reports results for a matching analysis of claims with and without financial status information. The matching is done on the exposure size and the credit score. The exposure size is matched using a Mahalanobis metric. The credit score is matched exactly. For each of the three Samples A, B, and C, there is one column for accounts with available financial status information and a second column for matched accounts where this information is not obtained. Each of these columns lists the means of the matching variables on the respective accounts. There is a third column for each sample with the differences in means and t-values for a two-tailed t-test. The bottom line states the number of matched pairs. Significance levels are indicated by *p<0.1; **p<0.05; ***p<0.01.

Table 7

Robustness checks - varying collection periods

	DV: Collection rate					
	Sample A		Sample B		Sample C	
	(1/4)	(1/2)	(1/4)	(1/2)	(1/4)	(1/2)
EXP	-0.018^{***}	-0.017^{***}	-0.004^{*}	0.0002	-0.146^{***}	-0.105^{***}
	(0.002)	(0.002)	(0.002)	(0.001)	(0.003)	(0.003)
AGE	-0.006^{***}	-0.009^{***}	-0.004^{***}	-0.004^{***}	-0.008^{***}	-0.011^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
AGE.NA	-0.046^{***}	-0.062^{***}	0.005	0.003	0.004	-0.008
	(0.007)	(0.008)	(0.003)	(0.003)	(0.007)	(0.007)
MALE	0.008^{***}	0.005^{**}	-0.005	-0.005^{**}	0.034^{***}	0.028^{***}
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
FIRM	-0.011	-0.026^{***}	-0.022^{***}	-0.026^{***}	0.002	-0.020^{**}
	(0.009)	(0.009)	(0.006)	(0.004)	(0.009)	(0.008)
INS.acqu	-0.254^{***}	-0.265^{***}	-0.151^{***}	-0.112^{***}	-0.071^{***}	-0.065^{***}
	(0.007)	(0.007)	(0.008)	(0.006)	(0.005)	(0.005)
TEL	0.046^{***}	0.091^{***}	0.051^{***}	0.049^{***}	0.138^{***}	0.153^{***}
	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
AGE.ACC	-0.046^{***}	-0.037^{***}	-0.016^{***}	-0.009^{***}	-0.006^{***}	-0.004^{***}
	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
AGE.ACC.NA	-0.084^{**}	-0.079^{**}	-0.018^{***}	-0.008^{***}	-0.015^{***}	-0.006^{**}
	(0.034)	(0.032)	(0.003)	(0.002)	(0.003)	(0.003)
PREMIUM	-0.036^{**}	0.010			-0.110^{***}	-0.051^{***}
	(0.016)	(0.015)			(0.014)	(0.011)
CR.c	0.009^{***}	0.010^{***}	0.001	0.001	0.008^{***}	0.007^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
UNEMPL.c	-0.006^{***}	-0.005^{***}	-0.005^{***}	-0.004^{***}	-0.010^{***}	-0.009^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
CR.p	0.012^{***}	0.013^{***}	0.001	0.00005	0.003^{**}	0.004^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
SCORE.num	-0.045^{***}	-0.058^{***}			-0.095^{***}	-0.091^{***}
	(0.001)	(0.001)			(0.001)	(0.001)
SCORE.num.NA	0.083^{***}	0.108^{***}			0.076^{***}	0.073^{***}
	(0.003)	(0.003)			(0.003)	(0.002)
INS.proc	-0.293^{***}	-0.353^{***}	-0.193^{***}	-0.137^{***}	-0.159^{***}	-0.164^{***}
	(0.007)	(0.008)	(0.011)	(0.008)	(0.007)	(0.006)
SINGL.ACC	-0.105^{***}	-0.152^{***}	-0.063^{***}	-0.055^{***}	0.177^{***}	0.139^{***}
	(0.002)	(0.002)	(0.006)	(0.005)	(0.003)	(0.003)
CR.OTHER	0.148^{***}	0.175^{***}	0.022^{***}	0.015^{***}	0.089^{***}	0.089^{***}
	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
TITLE	-0.281^{***}	-0.220^{***}	-0.104^{***}	-0.051^{***}	-0.175^{***}	-0.097^{***}
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
FIN.STATUS	-0.260^{***}	-0.093^{***}	-0.083^{***}	-0.005^{*}	-0.065^{***}	0.020^{***}
	(0.007)	(0.005)	(0.005)	(0.003)	(0.006)	(0.004)
R^2 adj.	0.507	0.476	0.367	0.205	0.34	0.34
Chi2	47,877.75	43,014.26	4,287.77	2,127.92	24,463.03	25,373.83
Chi2 Prob.	0.000	0.000	0.000	0.000	0.000	0.000
Observations	182,880	182,880	16,623	16,623	126,015	126,015

*p<0.1; **p<0.05; ***p<0.01; Table 7 displays checks of robustness for the full models in Table 5 using the collection periods as indicated in Figure 1. The first column for each sample shows results for a payment period that is one quarter as long as in the baseline case. The payment period in the second column is half as long. The table reports standard errors in brackets. The standard errors are clustered by year and postal code. The coefficients are calculated as the conditional marginal effects, which are given by the change in the outcome of the collection rate for a change of one standard deviation for continuous and a change of one unit for dichotomous variables. The marginal effects are calculated for the continuous variables at their median and for the dichotomous variables at zero.

Table 8

Robustness checks - single accounts, missing values

	DV: Collection rate					
	Single accounts			Complete accounts		
		Sample			Sample	
	(A)	(B)	(C)	(A)	(B)	(C)
EXP	−0.047***	0.005	−0.127***	−0.016***	0.005	−0.032***
	(0.007)	(0.004)	(0.004)	(0.006)	(0.014)	(0.005)
AGE	−0.023***	−0.013***	−0.024***	−0.0002	−0.050***	−0.005***
	(0.003)	(0.004)	(0.002)	(0.004)	(0.015)	(0.002)
AGE.NA	−0.093***	0.009	−0.028***			
	(0.014)	(0.009)	(0.010)			
MALE	0.005	−0.017**	0.039***	0.049***	0.031	0.013***
	(0.005)	(0.008)	(0.003)	(0.008)	(0.025)	(0.004)
FIRM	−0.004	−0.065***	−0.021*	2.334***		0.018
	(0.017)	(0.015)	(0.013)	(0.149)		(0.047)
INS.acqu	−0.389***	−0.337***	−0.100***	−0.115***	−0.772***	0.010
	(0.014)	(0.018)	(0.008)	(0.039)	(0.272)	(0.012)
TEL	0.204***	0.156***	0.247***	0.115***	0.062**	0.079***
	(0.005)	(0.010)	(0.003)	(0.008)	(0.028)	(0.004)
AGE.ACC	−0.041***	−0.032***	0.007***	−0.038***	−0.009**	−0.005
	(0.004)	(0.006)	(0.002)	(0.005)	(0.004)	(0.003)
AGE.ACC.NA	−0.216***	−0.034***	0.005			
	(0.053)	(0.007)	(0.004)			
PREMIUM	−0.074***		0.004	0.082*		−0.012
	(0.025)		(0.017)	(0.047)		(0.018)
CR.c	0.013***	0.004	0.008***	0.022***	0.002	0.009***
	(0.003)	(0.004)	(0.002)	(0.005)	(0.012)	(0.002)
UNEMPL.c	−0.015***	−0.014***	−0.015***	−0.005	0.013	−0.002
	(0.003)	(0.004)	(0.002)	(0.005)	(0.011)	(0.002)
CR.p	0.024***	−0.00005	0.006***	0.012***	0.006	0.001
	(0.003)	(0.004)	(0.002)	(0.004)	(0.011)	(0.002)
SCORE.num.NA	0.270***		0.124***			
	(0.007)		(0.004)			
SCORE.num	−0.096***		−0.136***	−0.139***		−0.042***
	(0.003)		(0.002)	(0.005)		(0.002)
INS.proc	−0.487***	−0.428***	−0.266***	−0.636***	−1.011***	−0.129***
	(0.013)	(0.025)	(0.010)	(0.042)	(0.323)	(0.014)
CR.OTHER				0.270***	0.041***	0.077***
				(0.005)	(0.015)	(0.002)
TITLE	−0.287***	−0.158***	−0.103***	−0.282***	−0.032	−0.018***
	(0.006)	(0.010)	(0.004)	(0.010)	(0.035)	(0.004)
FIN.STATUS	0.030***	0.026**	0.089***	0.036***	0.058	0.061***
	(0.010)	(0.011)	(0.007)	(0.013)	(0.058)	(0.007)
R^2 adj.	0.32	0.164	0.257	0.434	0.266	0.452
Chi2	14,654.31	1,283.54	17,137.96	10,306.26	583.62	4,295.82
Chi2 Prob.	0.000	0.000	0.000	0.000	0.000	0.000
Observations	63,312	11,465	86,754	31,172	543	18,154

*p<0.1; **p<0.05; ***p<0.01; Table 8 displays checks of robustness for only including one-time debtors and only including accounts with no missing values. The models correspond to the full models in Table 5. The table reports standard errors in brackets. The standard errors are clustered by year and postal code. The coefficients are calculated as the conditional marginal effects, which are given by the change in the outcome of the collection rate for a change of one standard deviation for continuous and a change of one unit for dichotomous variables. The marginal effects are calculated for the continuous variables at their median and for the dichotomous variables at zero.

Table 9

Robustness checks - single accounts with no hard negative credit information

	DV: Collection rate		
		Sample	
	(A)	(B)	(C)
EXP	-0.005^{*}	0.001	-0.046^{***}
	(0.003)	(0.003)	(0.002)
AGE	-0.044^{***}	-0.020^{***}	-0.013^{***}
	(0.003)	(0.005)	(0.001)
AGE.NA	-0.082^{***}	-0.027^{**}	-0.021^{***}
	(0.013)	(0.013)	(0.005)
MALE	-0.029^{***}	-0.006	0.013^{***}
	(0.006)	(0.009)	(0.002)
FIRM	-0.024	-0.028^{**}	-0.003
	(0.015)	(0.014)	(0.007)
TEL	0.138^{***}	0.086^{***}	0.109^{***}
	(0.005)	(0.012)	(0.002)
AGE.ACC	-0.046^{***}	-0.039^{***}	0.004^{***}
	(0.003)	(0.007)	(0.001)
AGE.ACC.NA	-0.243	-0.052^{***}	0.007^{**}
	(0.300)	(0.008)	(0.003)
PREMIUM	0.209		0.127
	(0.142)		(0.091)
CR.c	0.012^{***}	0.007^{*}	0.004^{***}
	(0.003)	(0.004)	(0.001)
UNEMPL.c	-0.011^{***}	-0.007^{*}	-0.005^{***}
	(0.003)	(0.004)	(0.001)
CR.p	0.017^{***}	0.001	0.002^{**}
	(0.003)	(0.004)	(0.001)
SCORE.num.NA	0.262^{***}		0.055^{***}
	(0.011)		(0.002)
SCORE.num	-0.022^{***}		-0.048^{***}
	(0.002)		(0.001)
R^2 adj.	0.092	0.045	0.152
Chi2	2,469.75	218.66	6,355.02
Chi2 Prob.	0.000	0.000	0.000
Observations	34,528	7,628	49,266

*p<0.1; **p<0.05; ***p<0.01; Table 9 displays checks of robustness for only including one-time debtors with no hard publicly available negative credit information (no enforceable claims and no insolvent debtors). The models correspond to the full models in Table 5. The table reports standard errors in brackets. The standard errors are clustered by year and postal code. The coefficients are calculated as the conditional marginal effects, which are given by the change in the outcome of the collection rate for a change of one standard deviation for continuous and a change of one unit for dichotomous variables. The marginal effects are calculated for the continuous variables at their median and for the dichotomous variables at zero.

Table 10

Robustness checks - individual and corporate debtors

	DV: Collection rate					
	Individuals			Firms		
		Sample			Sample	
	(A)	(B)	(C)	(A)	(B)	(C)
EXP	-0.023^{***}	-0.001	-0.084^{***}	0.010	0.007^{**}	-0.021^{**}
	(0.002)	(0.001)	(0.002)	(0.011)	(0.004)	(0.009)
AGE	-0.009^{***}	-0.005^{***}	-0.015^{***}			
	(0.001)	(0.001)	(0.001)			
AGE.NA	-0.056^{***}	0.006^{**}	-0.019^{***}			
	(0.007)	(0.003)	(0.007)			
MALE	0.005^{***}	-0.005^{*}	0.026^{***}			
	(0.002)	(0.002)	(0.002)			
INS.acqu	-0.201^{***}	-0.110^{***}	-0.052^{***}	-0.540^{***}	-0.130^{***}	-0.301^{***}
	(0.006)	(0.006)	(0.005)	(0.057)	(0.024)	(0.035)
TEL	0.096^{***}	0.053^{***}	0.168^{***}	0.061^{***}	0.016^{**}	0.059^{***}
	(0.002)	(0.003)	(0.002)	(0.008)	(0.006)	(0.008)
AGE.ACC	-0.023^{***}	-0.008^{***}	0.002	-0.041^{***}	-0.010^{***}	-0.012^{***}
	(0.001)	(0.001)	(0.001)	(0.005)	(0.002)	(0.004)
AGE.ACC.NA	-0.088^{***}	-0.005^{**}	0.003	-0.198	-0.017^{***}	-0.035^{***}
	(0.027)	(0.002)	(0.003)	(0.174)	(0.006)	(0.011)
PREMIUM	-0.020^{*}		-0.012	-0.033		0.017
	(0.011)		(0.011)	(0.109)		(0.023)
CR.c	0.009^{***}	0.001	0.007^{***}	0.009^{*}	0.0002	0.002
	(0.001)	(0.001)	(0.001)	(0.006)	(0.003)	(0.004)
UNEMPL.c	-0.004^{***}	-0.004^{***}	-0.008^{***}	-0.003	-0.003	-0.013^{**}
	(0.001)	(0.001)	(0.001)	(0.005)	(0.003)	(0.006)
CR.p	0.010^{***}	0.0002	0.004^{***}	0.016^{***}	-0.002	0.005
	(0.001)	(0.001)	(0.001)	(0.005)	(0.003)	(0.004)
SCORE.num	-0.050^{***}		-0.092^{***}			
	(0.001)		(0.001)			
SCORE.num.NA	0.089^{***}		0.075^{***}			
	(0.003)		(0.002)			
INS.proc	-0.314^{***}	-0.136^{***}	-0.190^{***}	-0.422^{***}	-0.126^{***}	-0.171^{***}
	(0.007)	(0.008)	(0.006)	(0.032)	(0.020)	(0.023)
SINGL.ACC	-0.143^{***}	-0.058^{***}	0.108^{***}	-0.060^{***}	-0.061^{***}	0.145^{***}
	(0.002)	(0.006)	(0.003)	(0.008)	(0.013)	(0.007)
CR.OTHER	0.139^{***}	0.013^{***}	0.080^{***}	0.206^{***}	0.033^{***}	0.118^{***}
	(0.001)	(0.001)	(0.001)	(0.007)	(0.005)	(0.005)
TITLE	-0.146^{***}	-0.046^{***}	-0.062^{***}	-0.047^{***}	-0.021^{**}	0.022^{*}
	(0.002)	(0.003)	(0.003)	(0.015)	(0.010)	(0.013)
FIN.STATUS	-0.005	0.004	0.064^{***}	-0.083^{***}	-0.020	-0.004
	(0.004)	(0.003)	(0.004)	(0.025)	(0.014)	(0.016)
R^2 adj.	0.428	0.171	0.321	0.453	0.316	0.317
Chi2	34,952.66	1,608.00	1,719.20	168.50	22,589.16	888.45
Chi2 Prob.	0.000	0.000	0.000	0.000	0.000	0.000
Observations	174,406	15,469	118,505	8,474	1,154	7,510

*p<0.1; **p<0.05; ***p<0.01; Table 10 displays checks of robustness for only including individual accounts and only including corporate accounts. The models correspond to the full models in Table 5. The reported standard errors (in brackets) are clustered by year and postal code. The coefficients are calculated as the conditional marginal effects, which are given by the change in the outcome of the collection rate for a change of one standard deviation for continuous and a change of one unit for dichotomous variables. The marginal effects are calculated for the continuous variables at their median and for the dichotomous variables at zero.

Table 11

Correlation table of the independent variables in the samples A, B, and C.

		EXP	AGE	AGE.NA	MALE	FIRM	INS.acqu	TEL	AGE.ACC	AGE.ACC.NA	PREMIUM	CR.c	UNEMPL.c	CR.p	SCORE.num.NA	SCORE.num	INS.proc	SINGL.ACC	CR.other	TITLE	FIN.STATUS
Sample A	EXP	1	.01	.07	.01	.07	.01	-.00	.20	.13	-.35	-.01	-.00	-.00	-.02	.01	.02	.01	-.03	.05	.04
	AGE	.01	1	-.00	.02	.00	.00	.11	-.09	.00	.03	.05	-.05	.05	.06	-.07	-.03	-.04	.06	-.11	-.03
	AGE.NA	.07	-.00	1	-.25	.87	.00	-.06	-.04	.01	-.00	-.03	.01	-.01	.15	.00	.01	.04	-.03	-.11	-.01
	MALE	.01	.02	-.25	1	-.30	-.01	.03	.02	.01	-.02	.02	-.03	.01	-.04	.08	-.01	-.03	.01	.02	-.03
	FIRM	.07	.00	.87	-.30	1	.00	-.03	-.04	-.00	.01	-.03	.01	-.01	.13	-.00	.01	.01	-.04	-.11	-.00
	INS.acqu	.01	.00	.00	-.01	.00	1	-.01	.02	.01	-.01	-.01	.01	-.02	.00	.05	-.02	.08	-.05	-.02	-.00
	TEL	-.00	.11	-.06	.03	-.03	-.01	1	-.05	-.01	.04	.05	-.04	.06	.18	-.10	-.03	-.13	.20	-.18	-.09
	AGE.ACC	.20	-.09	-.04	.02	-.04	.02	-.05	1	-.00	-.30	-.05	.05	-.04	-.13	.03	.02	.05	-.12	.21	.06
	AGE.ACC.NA	.13	.00	.01	.01	-.00	.01	-.01	-.00	1	-.33	-.00	.00	-.00	-.02	.01	.00	.03	-.01	.04	.02
	PREMIUM	-.35	.03	-.00	-.02	.01	-.01	.04	-.30	-.33	1	.01	-.01	.01	.08	-.03	-.01	-.06	.03	-.11	-.07
	CR.c	-.01	.05	-.03	.02	-.03	-.01	.05	-.05	-.00	.01	1	-.59	.48	.05	-.04	-.02	-.04	.09	-.05	-.03
	UNEMPL.c	-.00	-.05	.01	-.03	.01	.01	-.04	.05	.00	-.01	-.59	1	-.29	-.05	.03	.01	.02	-.07	.04	.02
	CR.p	-.00	.05	-.01	.01	-.01	-.02	.06	-.04	-.00	.01	.48	-.29	1	.05	-.04	-.02	-.04	.09	-.05	-.03
	SCORE.num.NA	-.02	.06	.15	-.04	.13	.00	.18	-.13	-.02	.08	.05	-.05	.05	1	.00	-.00	.01	.29	-.49	-.29
	SCORE.num	.01	-.07	.00	.08	-.00	.05	-.10	.03	.01	-.03	-.04	.03	-.04	.00	1	.02	.08	-.17	.06	-.16
	INS.proc	.02	-.03	.01	-.01	.01	-.02	-.03	.02	.00	-.01	-.02	.01	-.02	-.00	.02	1	.06	-.07	.06	.02
	SINGL.ACC	.01	-.04	.04	-.03	.01	.08	-.13	.05	.03	-.06	-.04	.02	-.04	.01	.08	.06	1	.00	.03	.09
	CR.OTHER	-.03	.06	-.03	.01	-.04	-.05	.20	-.12	-.01	.03	.09	-.07	.09	.29	-.17	-.07	.00	1	-.26	-.07
	TITLE	.05	-.11	-.11	.02	-.11	-.02	-.18	.21	.04	-.11	-.05	.04	-.05	-.49	.06	.06	.03	-.26	1	.28
	FIN.STATUS	.04	-.03	-.01	-.03	-.00	-.00	-.09	.06	.02	-.07	-.03	.02	-.03	-.29	-.16	.02	.09	-.07	.28	1
Sample B	EXP	1	.04	.06	-.11	.35	.02	.02	.01	.05		.00	-.01	.01			.03	.02	-.07	.05	.03
	AGE	.04	1	-.00	-.01	-.00	-.00	.04	-.01	-.09		.01	-.02	.02			-.03	-.00	-.02	-.04	-.04
	AGE.NA	.06	-.00	1	-.08	.23	-.01	-.60	-.03	-.08		.00	-.00	.00			-.05	.05	-.03	-.00	.05
	MALE	-.11	-.01	-.08	1	-.38	-.01	.02	.01	-.02		.01	-.01	.00			-.02	.01	.02	.03	.02
	FIRM	.35	-.00	.23	-.38	1	.02	-.03	-.00	-.05		-.02	-.00	.00			.06	-.04	-.07	-.07	-.06
	INS.acqu	.02	-.00	-.01	-.01	.02	1	-.04	.01	.02		-.01	.01	.00			-.02	.03	-.03	-.08	-.05
	TEL	.02	.04	-.60	.02	-.03	-.04	1	-.01	.01		.02	-.03	.01			-.03	-.07	.05	-.11	-.16
	AGE.ACC	.01	-.01	-.03	.01	-.00	.01	-.01	1	-.00		-.03	.02	-.01			.01	.04	-.06	.10	.09
	AGE.ACC.NA	.05	-.09	-.08	-.02	-.05	.02	.01	-.00	1		-.01	.08	-.02			.04	-.00	-.01	.19	.14
	CR.c	.00	.01	.00	.01	-.02	-.01	.02	-.03	-.01		1	-.25	.23			-.01	.00	.00	-.01	-.02
	UNEMPL.c	-.01	-.02	-.00	-.01	-.00	.01	-.03	.02	.08		-.25	1	-.27			-.00	.00	-.01	.02	.02
	CR.p	.01	.02	.00	.00	.00	.00	.01	-.01	-.02		.23	-.27	1			-.01	.01	.01	-.03	-.03
	INS.proc	.03	-.03	-.05	-.02	.06	-.02	-.03	.01	.04		-.01	-.00	-.01			1	.03	-.01	.02	-.00
	SINGL.ACC	.02	-.00	.05	.01	-.04	.03	-.07	.04	-.00		.00	.00	.01			.03	1	.00	.10	.10
	CR.OTHER	-.07	-.02	-.03	.02	-.07	-.03	.05	-.06	-.01		.00	-.01	.01			-.01	.00	1	-.08	-.06
	TITLE	.05	-.04	-.00	.03	-.07	-.08	-.11	.10	.19		-.01	.02	-.03			.02	.10	-.08	1	.72
	FIN.STATUS	.03	-.04	.05	.02	-.06	-.05	-.16	.09	.14		-.02	.02	-.03			-.00	.10	-.06	.72	1
Sample C	EXP	1	-.04	.05	.01	.03	-.01	-.02	.30	.07	-.35	-.03	.03	-.03	-.11	.04	.01	-.02	-.05	.21	.08
	AGE	-.04	1	-.00	.02	-.00	.02	.04	-.01	-.06	-.01	.02	-.02	.03	.08	-.03	-.01	.02	.02	-.07	-.02
	AGE.NA	.05	-.00	1	-.28	.86	-.01	-.04	-.00	-.04	-.01	-.01	.04	-.03	.35	-.00	.00	-.11	-.03	-.13	.00
	MALE	.01	.02	-.28	1	-.34	-.01	.03	.03	-.02	-.02	.01	-.01	.00	-.15	.07	-.00	.02	.02	.04	-.02
	FIRM	.03	-.00	.86	-.34	1	-.01	-.01	-.00	-.03	-.01	-.01	.03	-.03	.31	-.00	.01	-.14	-.03	-.11	.01
	INS.acqu	-.01	.02	-.01	-.01	-.01	1	-.01	-.00	.03	.00	-.00	-.01	-.00	.04	.14	-.04	.05	-.01	-.08	-.03
	TEL	-.02	.04	-.04	.03	-.01	-.01	1	-.07	-.06	.02	.04	-.04	.03	.02	-.20	-.01	-.00	.17	-.08	-.02
	AGE.ACC	.30	-.01	-.00	.03	-.00	-.00	-.07	1	-.00	-.61	.00	.00	-.00	-.04	-.01	-.00	.02	-.01	.08	-.01
	AGE.ACC.NA	.07	-.06	-.04	-.02	-.03	.03	-.06	-.00	1	.05	-.03	.01	-.01	-.12	.00	-.01	-.01	-.00	.07	-.01
	PREMIUM	-.35	-.01	-.01	-.02	-.01	.00	.02	-.61	.05	1	.01	-.02	.01	-.00	-.05	-.03	-.02	.01	-.17	-.03
	CR.c	-.03	.02	-.01	.01	-.01	-.00	.04	.00	-.03	.01	1	-.56	.35	.01	-.04	-.00	.03	.04	-.02	-.01
	UNEMPL.c	.03	-.02	.04	-.01	.03	-.01	-.04	.00	.01	-.02	-.56	1	-.36	.01	.02	.01	-.01	-.02	.02	.02
	CR.p	-.03	.03	-.03	.00	-.03	-.00	.03	-.00	-.01	.01	.35	-.36	1	.00	-.03	-.00	.03	.02	-.02	-.01
	SCORE.num.NA	-.11	.08	.35	-.15	.31	.04	.02	-.04	-.12	-.00	.01	.01	.00	1	-.00	-.00	.07	.04	-.43	-.15
	SCORE.num	.04	-.03	-.00	.07	-.00	.14	-.20	-.01	.00	-.05	-.04	.02	-.03	-.00	1	.03	-.10	-.18	.24	-.07
	INS.proc	.01	-.01	.00	-.00	.01	-.04	-.01	-.00	-.01	-.03	-.00	.01	-.00	-.00	.03	1	.03	-.01	.03	.01
	SINGL.ACC	-.02	.02	-.11	.02	-.14	.05	-.00	.02	-.01	-.02	.03	-.01	.03	.07	-.10	.03	1	.00	-.06	.05
	CR.OTHER	-.05	.02	-.03	.02	-.03	-.01	.17	-.01	-.00	.01	.04	-.02	.02	.04	-.18	-.01	.00	1	-.10	-.00
	TITLE	.21	-.07	-.13	.04	-.11	-.08	-.08	.08	.07	-.17	-.02	.02	-.02	-.43	.24	.03	-.06	-.10	1	.33
	FIN.STATUS	.08	-.02	.00	-.02	.01	-.03	-.02	-.01	-.01	-.03	-.01	.02	-.01	-.15	-.07	.01	.05	-.00	.33	1

Table 11 displays the correlation coefficients for the independent characteristics included in the regression models.