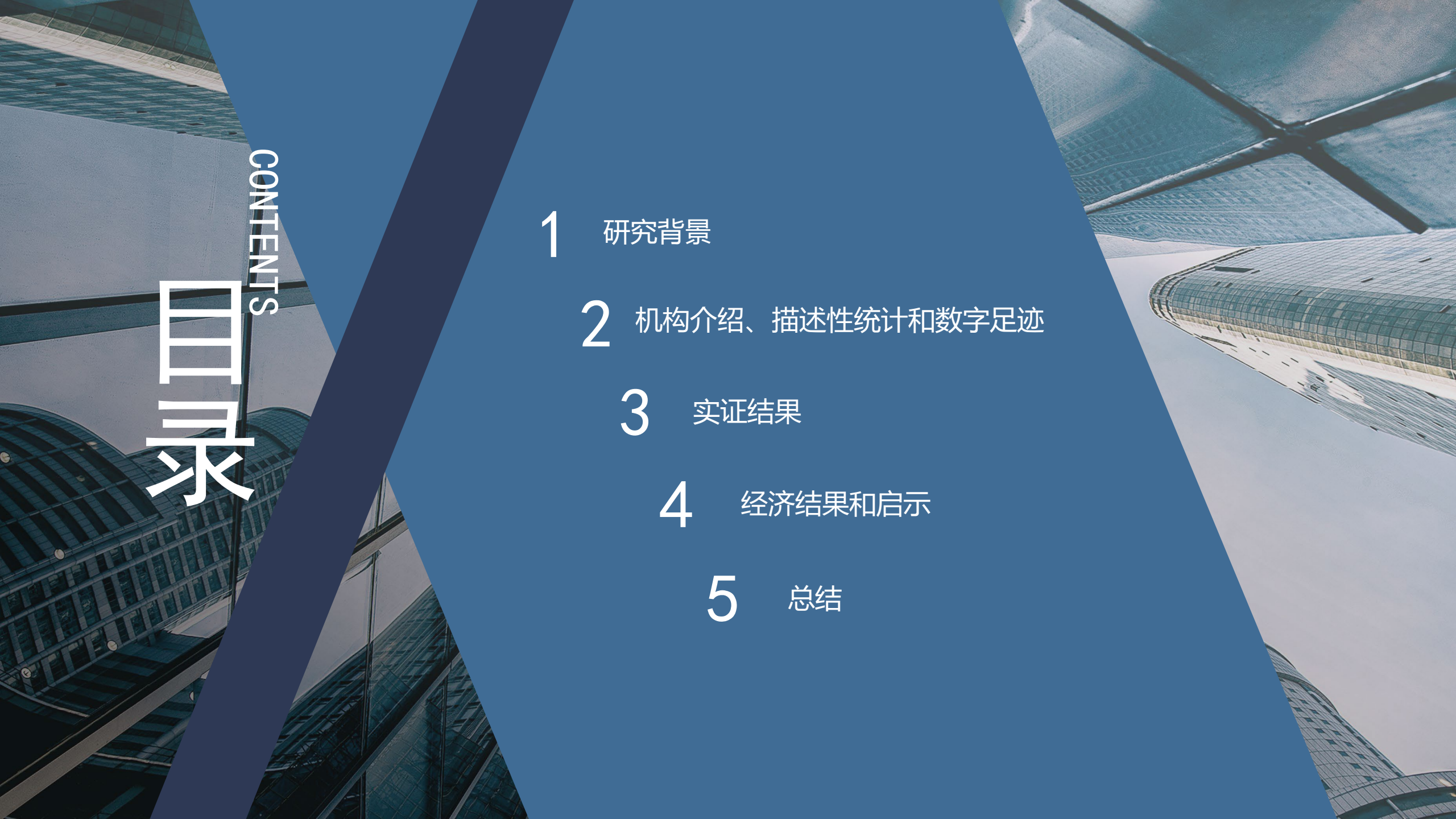
On the Rise of FinTechs: Credit Scoring Using Digital Footprints

基于数字足迹的个人信用评分

南京大学工程管理学院



CONTENTS

目录

PART 1

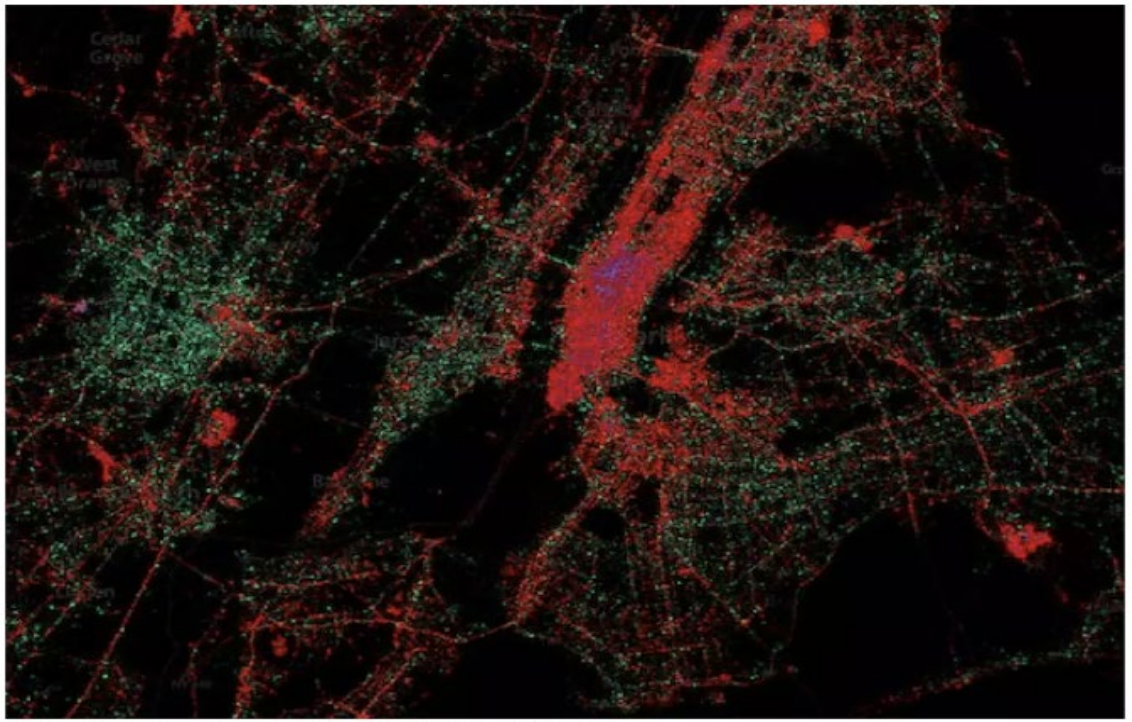
研究背景

Research Background

1 研究背景

Research Background

纽约：手机操作系统的使用分布
(设备是用户收入的代理指标)



Red = iOS, Green = Android, Purple = Blackberry

数字足迹

全球几乎每个互联网用户都在网上留下了痕迹，如手机操作系统类型、电子邮箱提供商等

这些痕迹是关于这些用户的简单、易于获取的信息，我们称之为“数字足迹”(Digital Footprint)。

1 研究背景

Research Background

研究问题

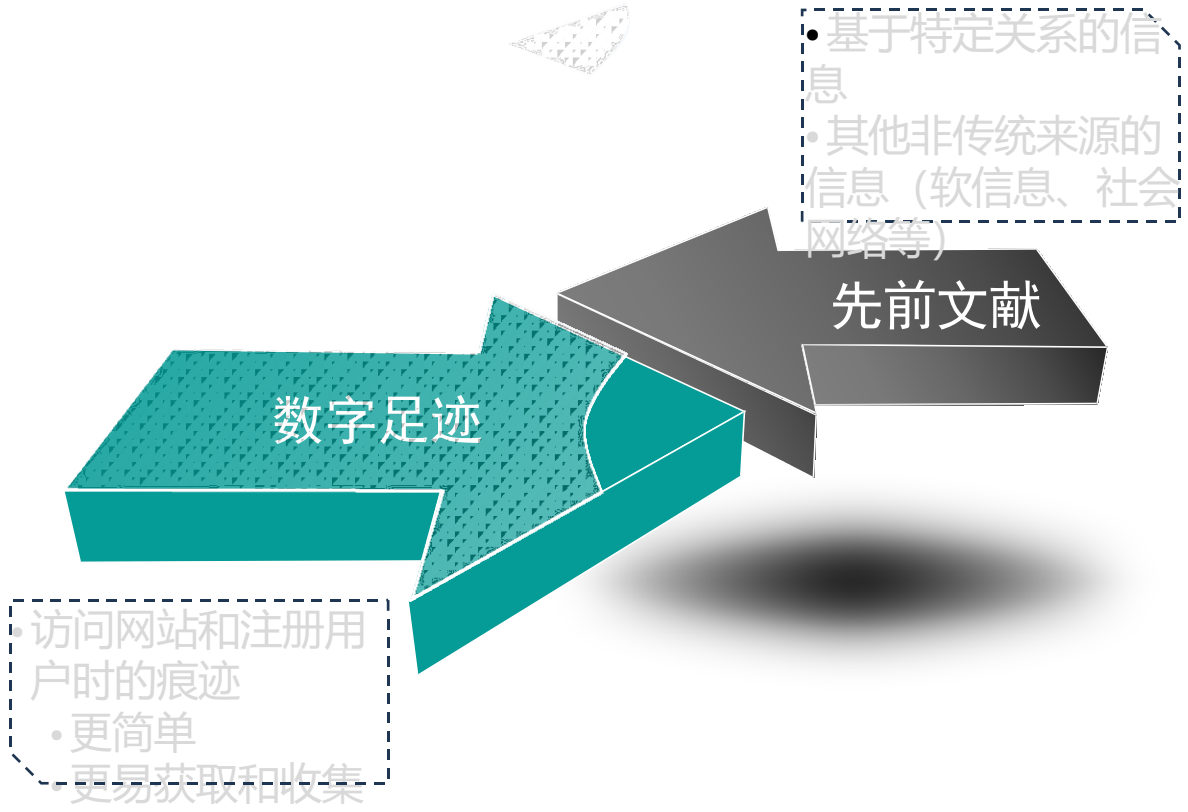
数字足迹能否预测借款人的违约概率？能否对信用局信用评分等传统的被用于违约预测的重要信息构成补充？

研究数字足迹的重要性

传统金融中介存在的关键原因是，它们与客户有长期关系，故有信息优势，能够筛选和监督借款人；

如果数字足迹可以预测违约，那么金融科技公司可能威胁金融中介的信息优势，从而挑战其商业模式。

创新点



1 研究背景

Research Background

研究问题

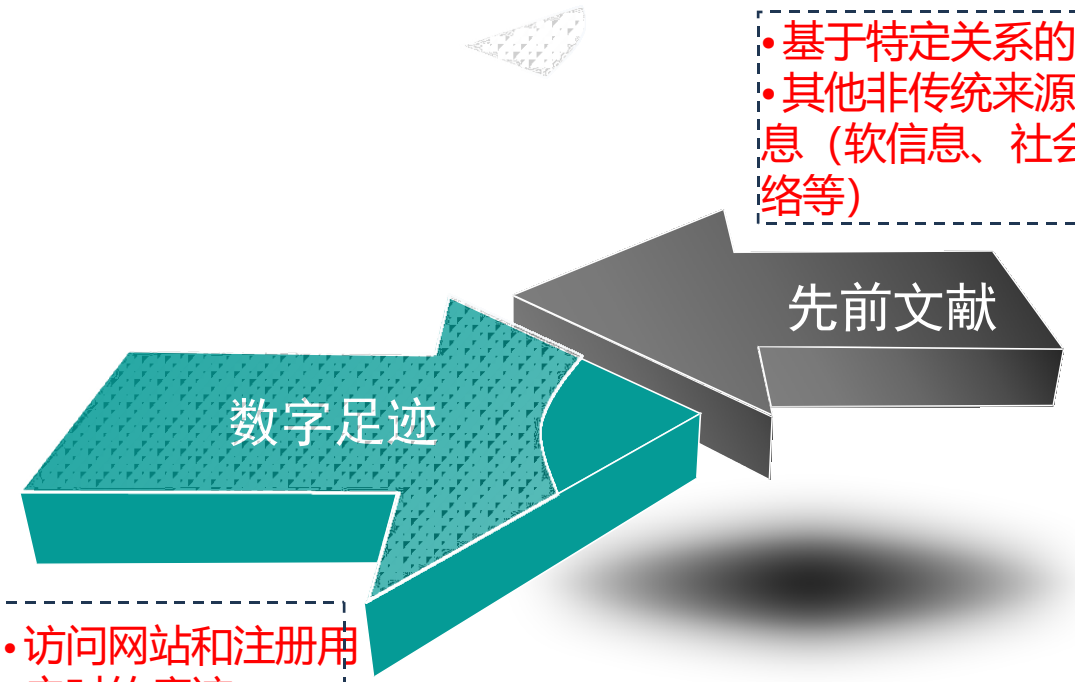
数字足迹能否预测借款人的违约概率？能否对信用局信用评分等传统的被用于违约预测的重要信息构成补充？

研究数字足迹的重要性

传统金融中介存在的关键原因是，它们与客户有长期关系，故有信息优势，能够筛选和监督借款人；

如果数字足迹可以预测违约，那么金融科技公司可能威胁金融中介的信息优势，从而挑战其商业模式。

创新点



1 研究背景：主要发现

Research Background

- 1. 简单、易于获取的数字足迹变量，其信息含量**与信用局评分的信息含量相当。**
- 2. 数字足迹可以**补充而不是替代**信用局评分。
- 3. 数字足迹不仅可以预测有信用局评分的客户的违约率，还**可以预测没有信用局评分的客户的违约率**，且效果一样好。

1 研究背景：主要发现

Research Background

- 1. 简单、易于获取的数字足迹变量，其信息含量**与信用局评分的信息含量相当。**
  - 对客户违约率的预测准确性相同
- 2. 数字足迹可以**补充而不是替代**信用局评分。
  - 两者联合时可提高客户违约率的预测准确性
- 3. 数字足迹不仅可以预测有信用局评分的客户的违约率，还**可以预测没有信用局评分的客户的违约率**，且效果一样好。
  - 可以帮助那些原本不能获得信贷服务的人获得信贷，从而促进金融包容

PART 2

机构介绍、描述性统计和数字足迹

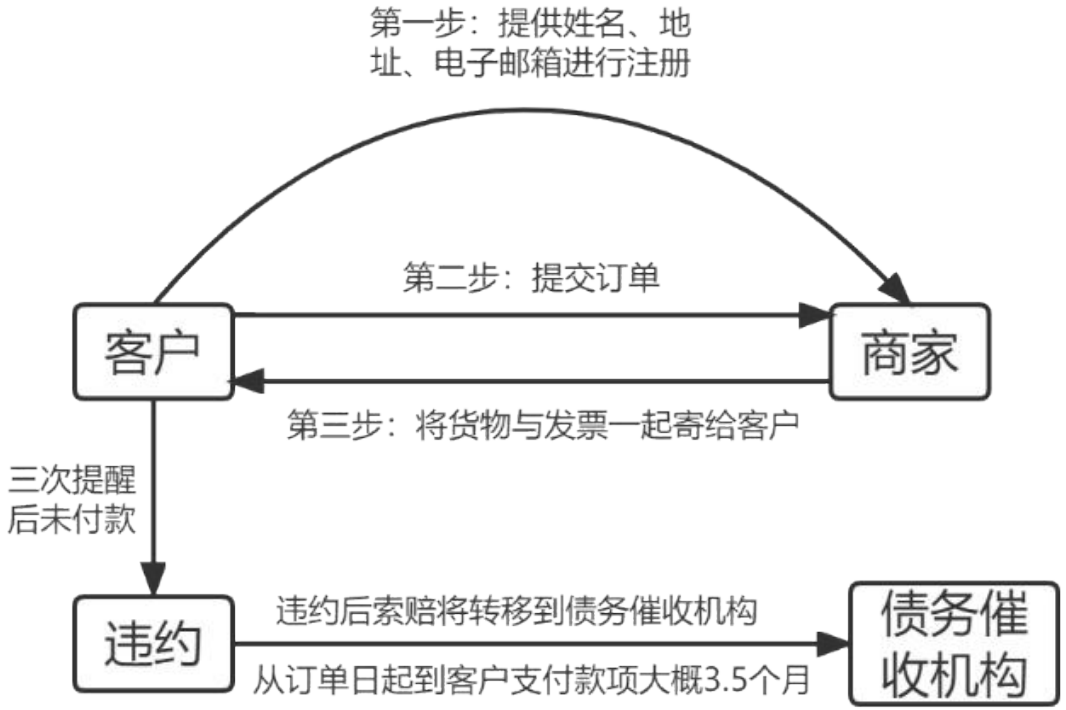
Institutional Setup, Descriptive Statistics, and the Digital Footprint

2.1 机构介绍

Institutional Setup

1) 机构的主营业务及其向客户提供的贷款

德国的一家电子商务公司，主营业务是在线销售家具，允许符合条件的客户货到付款（即提供短期消费贷款），及其对客户违约的定义：

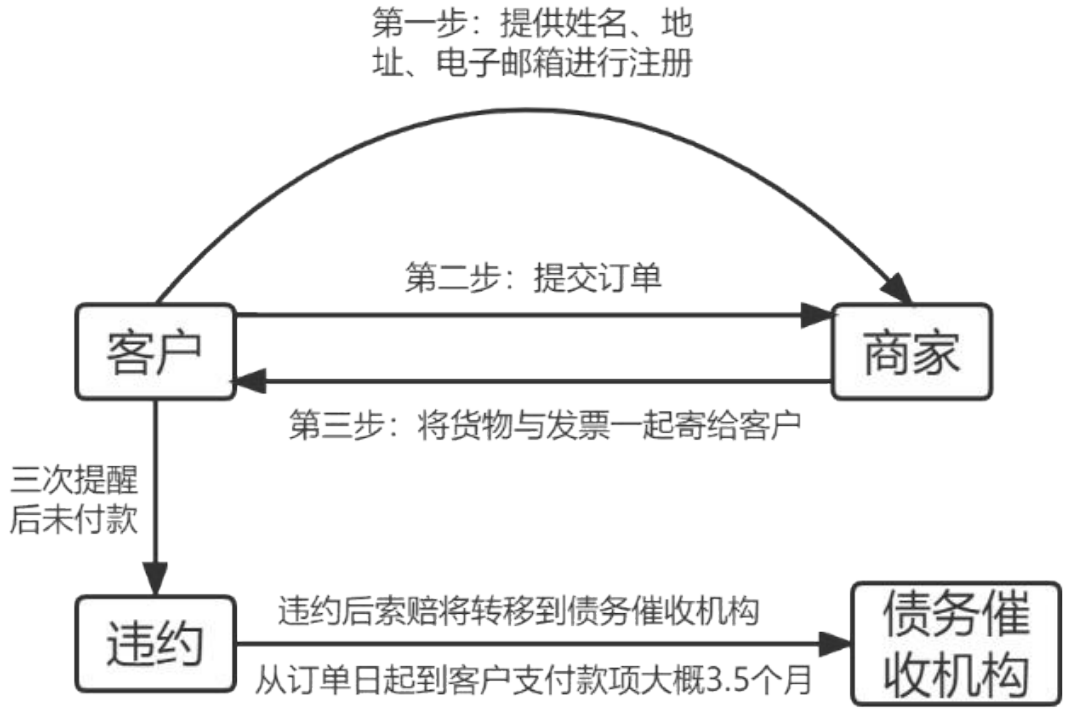


2.1 机构介绍

Institutional Setup

1) 机构的主营业务及其向客户提供的贷款

德国的一家电子商务公司，主营业务是在线销售家具，允许符合条件的客户货到付款（即提供短期消费贷款），及其对客户违约的定义：



(2) 机构用下面三类数据评估客户的信誉，进而决定是否允许客户货到付款

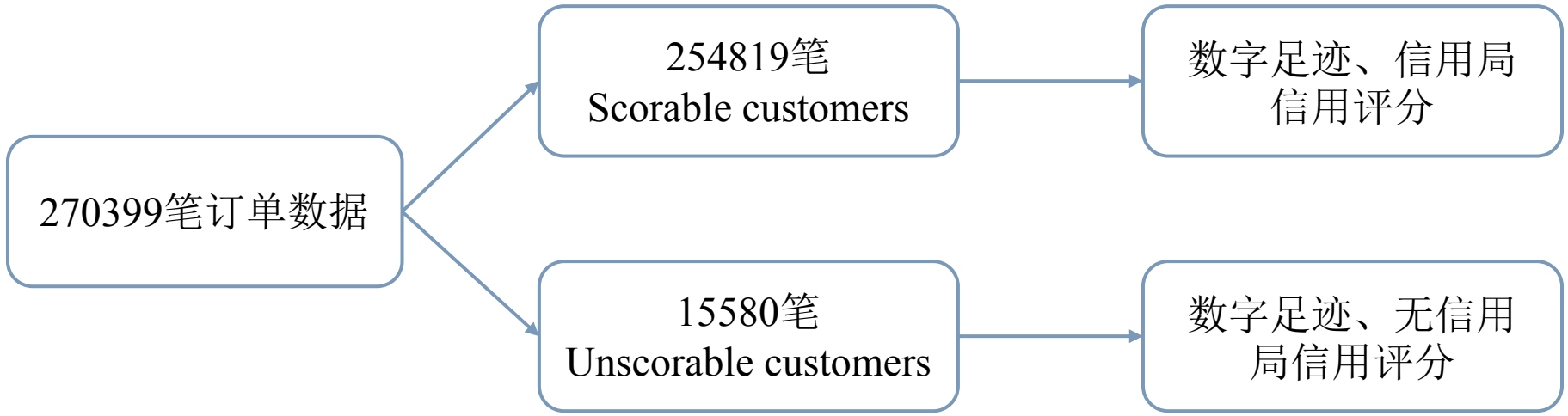
三类数据	信息
第一信用局评分 The first credit bureau	提供基本信息，如客户是否存在，客户是否目前处于破产或最近已经破产
第二信用局评分 The second credit bureau	不同银行的历史数据（信用卡债务和未偿还贷款、过去的支付行为、银行账户和信用卡的数量）、社会人口数据、各行业公司的支付行为数据
数字足迹	客户在该德国公司网站上留下的痕迹

2.1 机构介绍

Institutional Setup

(1) 时间跨度：2015.10 – 2016.12，该机构从2015.10开始引入数字足迹

(2) 交易总数：



主回归仅保留购买金额超过100欧元的订单（254819笔），只有这部分订单的客户需要提供信用评分；

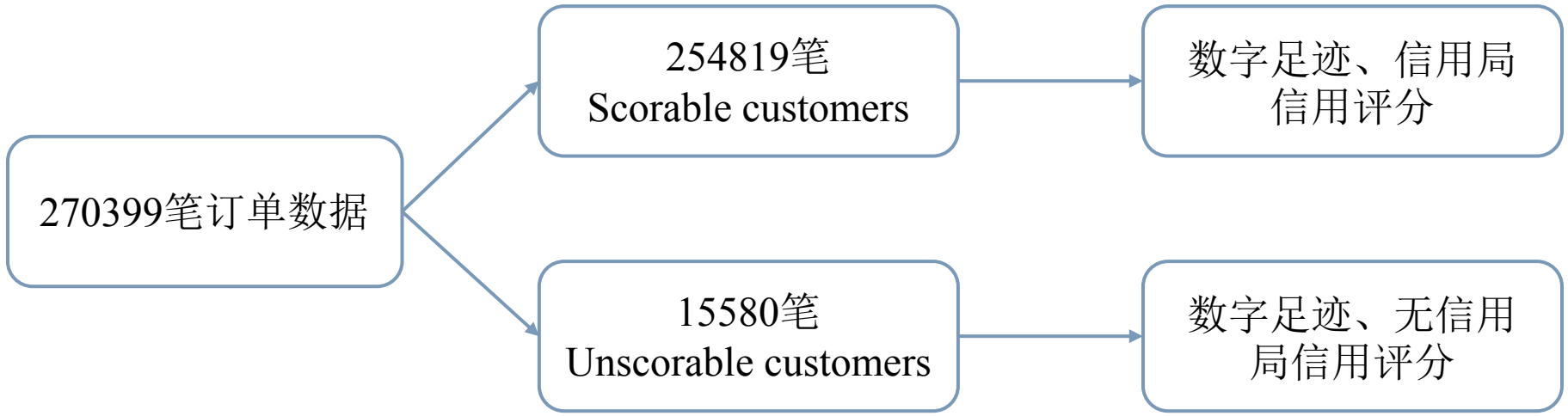
通过以上方式从样本中排除信誉非常低的客户，因此本文样本与以往研究信用卡、银行贷款或P2P贷款违约的数据集相似。

2.1 机构介绍

Institutional Setup

(1) 时间跨度：2015.10 – 2016.12，该机构从2015.10开始引入数字足迹

(2) 交易总数：



主回归仅保留购买金额超过100欧元的订单（254819笔），只有这部分订单的客户需要提供信用评分；

通过以上方式从样本中排除信誉非常低的客户，以使本文样本与以往文献用来研究信用卡、银行贷款或P2P贷款违约的数据集相似。

2.2 描述性统计

Descriptive Statistics

这些客户有第二信用局评分，被称为“*scorable customers*”

这些客户无第二信用局评分，被称为“*unscorable customers*”

Table 1
Descriptive statistics

A. Customers with credit bureau score

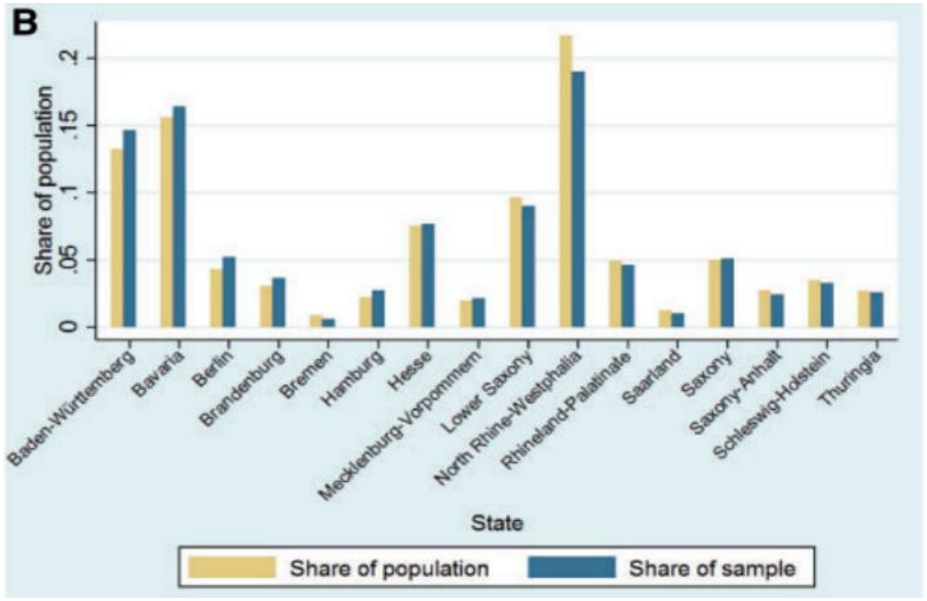
Variable	Unit	N	Mean	SD	P25	Median	P75
Order and customer							
Order amount	Euro	254,819	317.75	317.10	119.99	218.90	399.98
Gender	Dummy (0=male, 1=female)	254,819	0.66	0.47	0	1	1
Age[a]	Number	254,613	45.06	13.31	34	45	54
Credit bureau score	Number (0=worst, 100=best)	254,819	98.11	2.05	97.58	98.86	99.41
Payment behavior							
Default	Dummy (0/1)	254,819	0.009	0.096	0	0	0

B. Customers without credit bureau score

Variable	Unit	N	Mean	Std.	P25	Median	P75
Order and customer							
Order amount	Euro	15,580	324.57	319.22	119.99	221.60	399.99
Gender	Dummy (0=male, 1=female)	15,580	0.70	0.46	0	1	1
Age[a]	Number	555	38.20	10.46	30	35	46
Credit bureau score	Number (0=worst, 100=best)	15,580	na	na	na	na	na
Payment behavior							
Default	Dummy (0/1)	15,580	0.025	0.156	0	0	0

2.3 数据集的代表性

Representativeness of Data Set



A. Customers with credit bureau score

Variable	Unit	N	Mean
Order and customer			
Order amount	Euro	254,819	317.75
Gender	Dummy (0=male, 1=female)	254,819	0.66
Age[a]	Number	254,613	45.06
Credit bureau score	Number (0=worst, 100=best)	254,819	98.11
Payment behavior			
Default	Dummy (0/1)	254,819	0.009

订单的地理分布与德国人口的地理分布相似

客户的年龄分布与德国人口的年龄分布相似

2.3 数据集的代表性

Representativeness of Data Set

Table A2
Comparability of default rates to other retail data sets

Study	Sample	Default rate (%)	Time horizon	Default rate (annualized, %)
This study				
This study	270,399 purchases at a German E-commerce company between October 2015 and December 2016	1.0	~4 mo.	3.0
Germany				
Berg, Puri, and Rocholl (2017)	100,000 consumer loans at a large German private bank, 2008–2010	2.5	12 mo.	2.5
Puri, Rocholl, and Steffen (2017)	1 million consumer loans at 296 German savings banks, 2004–2008	1.1	12 mo.	1.1
Schufa 2017[a] (study by the major credit bureau in Germany	17.4 million consumer loans covered by the main credit bureau in Germany in 2016	2.2	12 mo.	2.2
Schufa (2016)[a] (study by the major credit bureau in Germany)	17.3 million consumer loans covered by the main credit bureau in Germany in 2015	2.4	12 mo.	2.4
Deutsche Bank (2016)[b]	All retail loans of Deutsche Bank (i.e., the largest German bank)	1.5 (Basel II PD estimate)	12 mo.	1.5
Commerzbank (2016)[c]	All retail loans of Commerzbank (i.e., the second-largest German bank)	2.0 (Basel II PD estimate)	12 mo.	2.0
United States				
Federal Reserve[d]	Charge-off rate on consumer loans, Q4/2016	2.09	12 mo. (annualized quarterly data)	2.09
Federal Reserve[d]	Charge-off rate on consumer loans, Q4/2015	1.76	12 mo. (annualized quarterly data)	1.76
Hertzberg, Liberman, and Paravisini (2016)	12,091 36-mo. Loans from Lending Club issued between December 2012 and February 2013	9.2	~26 mo.	4.2
Lending Club (own analysis)	375,803 36-mo. Loans from Lending Club issued between October 2015 and December 2016	5.11	12 mo.	5.11
Iyer et al. (2016)	17,212 36-mo. Loans from Prosper.com issued between February 2007 and October 2008	30.6	36 mo.	10.2
Hildebrandt, Puri, and Rocholl 2017	12,183 loans from Prosper.com between February 2007 and April 2008	10.8%–18.6	per 1,000 d.	3.9–6.8

4个月窗口期内平均违约率：

$$\frac{254819}{270399}\times 0.9\% + \frac{15580}{270399}\times 2.5\% = 1\% (= 3\%\text{年化})$$

表A2将本文的违约率与其他研究的违约率进行了比较，违约率与本文的样本相当

订单的违约率与德国典型消费贷款的违约率相似

2.4 数字足迹

Digital Footprint

Digital footprint variables

Device type	Device type. Main examples: Desktop, Tablet, Mobile.	Categorical variable	反应经济状况
Operating system	Operating system. Main examples: Windows, iOS, Android, Macintosh	Categorical variable	
E-mail host	E-mail host. Main examples: Gmx, Web, T-Online, Gmail, Yahoo, Hotmail	Categorical variable	
Channel	Channel through which customer comes to Web site. Main examples: Paid (including paid and retargeted clicks), Direct, Affiliate, Organic	Categorical variable	反映性格特征
Checkout time	Time of day of purchase	Numerical variable (0–24 hr)	
Do-not-track setting	Dummy equal to one if customer does not allow tracking of device and operating system information, and channel	Dummy variable	
Name in E-mail	Dummy equal to one if first or last name of customer is part of e-mail address	Dummy variable	反映声誉
Number in E-mail	Dummy equal to one if a number is part of e-mail address	Dummy variable	
Is lowercase	Dummy equal to one if first name, last name, street, or city are written in lowercase	Dummy variable	
E-mail error	Dummy equal to one if e-mail address contains an error in the first trial (Note: Clients can only order if they register with a correct e-mail address)	Dummy variable	

图：数字足迹分类

- 但一个变量也可能同时代表几个特征。

- 例如， iOS设备既是经济地位的预测因子（伯特兰和卡梅尼卡，2017年），也可能代表人物性格特征（寻求地位的用户更有可能购买iOS设备）。

PART 3

实证结果

Empirical Results

3.1 单变量结果

Univariate results

信用局评分：

Variable	Value	Observations	Proportion (%)	Default rate (%)	*t*-test against baseline
Credit bureau score (by quintile)	All	254,819	100	0.94	
	Q1 - lowest	50,980	20	2.12	Baseline
	Q2	50,949	20	1.02***	(−14.17)
	Q3	50,991	20	0.68***	(−19.51)
	Q4	51,181	20	0.47***	(−23.37)
	Q5 - highest	50,718	20	0.39***	(−24.89)

- 信用评分最低的客户的违约率为2.12%，超过平均违约率的2倍，超过信用评分最高的客户的5倍

3.1 单变量结果

Univariate results

数字足迹：反映收入和财务的变量

Variable	Value	Observations	Proportion (%)	Default rate (%)	*t*-test against baseline
Device	All	254,819	100	0.94	
	Desktop	145,879	57	0.74	Baseline
	Tablet	45,575	18	0.91***	(3.62)
	Mobile	26,808	11	2.14***	(21.84)
	Do-not-track setting	36,557	14	0.88***	(2.90)
Operating system	All	254,819	100	0.94	
	Windows	124,605	49	0.74	Baseline
	iOS	41,478	16	1.07***	(6.35)
	Android	29,089	11	1.79***	(16.64)
	Macintosh	21,163	8	0.69	(−0.79)
	Other	1,927	1	1.09*	(1.74)
	Do-not-track setting	36,557	14	0.88***	(2.66)
E-mail host	All	254,819	100	0.94	
	Gmx (partly paid)	58,609	23	0.82	Baseline
	Web (partly paid)	54,867	22	0.86	(0.70)
	T-Online (affluent customers)	30,279	12	0.51***	(−5.32)
	Gmail (free)	27,845	11	1.25***	(6.02)
	Yahoo (free, older service)	11,923	5	1.96***	(11.33)
	Hotmail (free, older service)	10,241	4	1.45***	(6.11)
	Other	61,055	24	0.90	(1.38)

- 代表收入、财富的数字足迹变量和违约率显著相关
- 手机订单的违约率是台式电脑订单违约率的3倍
- 安卓系统的订单违约率是IOS系统的2倍 (与购买iPhone的消费者通常比购买其他智能手机的更富有的观点一致)
- 来自高端互联网服务的客户 (T-online) 的违约率更低

3.1 单变量结果

Univariate results

数字足迹：反映性格特征的变量

Variable	Value	Observations	Proportion (%)	Default rate (%)	*t*-test against baseline
Channel	All	254,819	100	0.94	
	Paid	111,399	44	1.11	Baseline
	Direct	45,183	18	0.84***	(−4.78)
	Affiliate	24,770	10	0.64***	(−6.68)
	Organic	18,295	7	0.86***	(−3.00)
	Other	18,615	7	0.69***	(−5.24)
	Do-not-track setting	36,557	14	0.88***	(−3.69)
Checkout time	All	254,819	100	0.94	
	Evening (6 p.m.-midnight)	108,549	43	0.85	Baseline
	Night (midnight-6 a.m.)	6,913	3	1.97***	(9.49)
	Morning (6 a.m.-noon)	46,601	18	1.09***	(4.55)
	Afternoon (noon-6 p.m.)	92,756	36	0.89	(0.91)

- 性格特征与违约率有显著关系
- 通过付费广告进入主页的客户的违约率最高(1.11%) (可能是广告诱使客户购买无法负担的产品)，通过联属网站链接进入的客户以及直接在浏览器输入网站的客户的违约率较低
- 在夜间下单的客户大概是平均违约率的2倍

3.1 单变量结果

Univariate results

数字足迹：反映声誉的变量

Variable	Value	Observations	Proportion (%)	Default rate (%)	*t*-test against baseline
Do-not-track setting	All	254,819	100	0.94	
	No	218,262	86	0.94	Baseline
	Yes	36,557	14	0.88	(−1.12)
Name in e-mail	All	254,819	100	0.94	
	No	71,017	28	1.24	Baseline
	Yes	183,802	72	0.82***	(−9.99)
Number in e-mail	All	254,819	100	0.94	
	No	213,649	84	0.84	Baseline
	Yes	41,170	16	1.41***	(10.95)
Is lowercase	All	254,819	100	0.94	
	No	235,569	92	0.84	Baseline
	Yes	19,250	8	2.14***	(18.07)
E-mail error	All	254,819	100	0.94	
	No	251,319	99	0.88	Baseline
	Yes	3,500	1	5.09***	(25.71)

- 在电子邮件中使用自己名字的客户几乎不太可能违约 (以自己名字命名的公司表现更好（Chatterji and Daley, 2017))
- 在输入电子邮件地址时有数字的客户的违约率更高 (与非欺诈违约相比，欺诈案件的电子邮件地址中有数字的占比更高，可能是欺诈者使用数字创建了大量的电子邮件地址)
- 在输入姓名和送货地址时只使用小写字母的客户违约可能性是首字母大写的两倍多
- 在输入电子邮件地址时有拼写错误的客户的违约率更高

3.2 变量之间相关性分析

Measure of Association Between Variables

表3：信用局评分、数字足迹变量及控制变量（年龄、性别、日期、金额、购买项目的类型）的相关矩阵

Correlation/association between credit bureau score, digital footprint, and control variables (scorable customers)

	Credit bureau score	Device type	Operating system	E-mail host	Channel	Checkout time	Name in e-mail	Number in e-mail	Is lowercase	E-mail error	Age	Order amount	Item category	Month
Main variables														
Credit bureau score[a]	1.00***	0.07***	0.05***	0.07***	0.03***	0.03***	0.01***	0.07***	0.02***	0.00	0.20***	0.01***	0.05***	0.01**
Device type		1.00***	0.71***[b]	0.07***	0.06***[b]	0.04***	0.05***	0.06***	0.07***	0.01***	0.12***	0.03***	0.05***	0.06**
Operating system			1.00***	0.08***	0.06***[b]	0.04***	0.06***	0.08***	0.06***	0.01***	0.10***	0.02***	0.04***	0.03**
E-mail host				1.00***	0.03***	0.03***	0.08***	0.18***	0.04***	0.06***	0.16***	0.02***	0.02***	0.01**
Channel					1.00***	0.02***	0.01***	0.02***	0.04***	0.02***	0.09***	0.04***	0.06***	0.13**
Checkout time[a]						1.00***	0.01***	0.01***	0.01***	0.01*	0.06***	0.01***	0.03***	0.02**
Name in e-mail							1.00***	0.22***	0.01***	0.02***	0.04***	0.01	0.03***	0.01
Number in e-mail								1.00***	0.02***	0.00**	0.06***	0.01***	0.04***	0.01**
Is lowercase									1.00***	0.03***	0.03***	0.02***	0.02***	0.02**
E-mail error										1.00***	0.03***	0.01**	0.01***	0.01*
Control variables														
Age[a]											1.00***	0.05***	0.11***	0.03**
Order amount[a]												1.00***	0.27***	0.02**
Item category													1.00***	0.11**
Month														1.00**

信用局评分和数字足迹变量之间的克莱姆系数的值很小，介于0.01到0.07之间。

初步说明：数字足迹变量可以作为信用局评分的补充，而不是替代品。

3.2 变量之间相关性分析

Measure of Association Between Variables

表3：信用局评分、数字足迹变量及控制变量（年龄、性别、日期、金额、购买项目的类型）的相关矩阵

Correlation/association between credit bureau score, digital footprint, and control variables (scorable customers)

	Credit bureau score	Device type	Operating system	E-mail host	Channel	Checkout time	Name in e-mail	Number in e-mail	Is lowercase	E-mail error	Age	Order amount	Item category	Month
Main variables														
Credit bureau score[a]	1.00***	0.07***	0.05***	0.07***	0.03***	0.03***	0.01***	0.07***	0.02***	0.00	0.20***	0.01***	0.05***	0.01**
Device type		1.00***	0.71***[b]	0.07***	0.06***[b]	0.04***	0.05***	0.06***	0.07***	0.01***	0.12***	0.03***	0.05***	0.06**
Operating system			1.00***	0.08***	0.06***[b]	0.04***	0.06***	0.08***	0.06***	0.01***	0.10***	0.02***	0.04***	0.03**
E-mail host				1.00***	0.03***	0.03***	0.08***	0.18***	0.04***	0.06***	0.16***	0.02***	0.02***	0.01**
Channel					1.00***	0.02***	0.01***	0.02***	0.04***	0.02***	0.09***	0.04***	0.06***	0.13**
Checkout time[a]						1.00***	0.01***	0.01***	0.01***	0.01*	0.06***	0.01***	0.03***	0.02**
Name in e-mail							1.00***	0.22***	0.01***	0.02***	0.04***	0.01	0.03***	0.01
Number in e-mail								1.00***	0.02***	0.00**	0.06***	0.01***	0.04***	0.01**
Is lowercase									1.00***	0.03***	0.03***	0.02***	0.02***	0.02**
E-mail error										1.00***	0.03***	0.01**	0.01***	0.01*
Control variables														
Age[a]											1.00***	0.05***	0.11***	0.03**
Order amount[a]												1.00***	0.27***	0.02**
Item category													1.00***	0.11**
Month														1.00**

数字足迹变量组合（除设备类型与操作系统外）Cramer'sV都小于0. 25

初步说明：数字足迹变量提供相互独立的信息

数字足迹变量的组合在预测违约方面明显比单个变量更有效

3.3 主回归：数字足迹的预测有效性

Multivariate results: predictive effectiveness of digital footprint

Table 4
Default regressions (scorable customers)

	(1) Credit bureau bureau score		(2) Digital footprint		(3) Credit bureau score & digital footprint		(4) Credit bureau score & digital footprint, further controls	
Variables	Coef.	z-stat	Coef.	z-stat	Coef.	z-stat	Coef.	z-stat
Credit bureau score	−0.17***	(−7.89)			−0.15***	(−6.67)	−0.14***	(−5.90)
Device type & operating system[a]								
Desktop/Windows			Baseline		Baseline		Baseline	
Desktop/Macintosh			−0.07	(−0.53)	−0.13	(−1.03)	−0.19	(−1.52)
Tablet/Android			0.29***	(3.19)	0.29***	(3.06)	0.33***	(3.44)
Tablet/iOS			0.08	(1.05)	0.08	(0.97)	0.07	(0.89)
Mobile/Android			1.05***	(17.25)	0.95***	(15.34)	1.01***	(16.13)
Mobile/iOS			0.72***	(9.07)	0.57***	(6.73)	0.61***	(7.26)
E-mail Host [a]								
Gmx (partly paid)			Baseline		Baseline		Baseline	
Web (partly paid)			0.00	(0.00)	−0.02	(−0.22)	−0.01	(−0.08)
T-Online (affluent customers)			−0.40***	(−3.90)	−0.35***	(−3.35)	−0.27**	(−2.47)
Gmail (free)			0.34***	(3.81)	0.29***	(3.09)	0.27***	(2.86)
Yahoo (free, older service)			0.75***	(9.19)	0.72***	(8.98)	0.70***	(8.28)
Hotmail (free, older service)			0.35***	(3.70)	0.28***	(2.72)	0.25**	(2.32)
Channel								
Paid			Baseline		Baseline		Baseline	
Affiliate			−0.49***	(−5.35)	−0.54***	(−5.58)	−0.61***	(−6.31)
Direct			−0.27***	(−4.25)	−0.28***	(−4.44)	−0.26***	(−4.30)
Organic			−0.15*	(−1.79)	−0.15*	(−1.74)	−0.15*	(−1.82)
Other			−0.47***	(−4.50)	−0.48***	(−4.36)	−0.39***	(−3.43)
Checkout time								
Evening (6 p.m.-midnight)			Baseline		Baseline		Baseline	
Morning (6 a.m.-noon)			0.28***	(4.50)	0.28***	(4.60)	0.29***	(4.75)
Afternoon (noon-6 p.m.)			0.08	(1.42)	0.08	(1.47)	0.10*	(1.92)
Night (midnight-6 a.m.)			0.79***	(7.73)	0.75***	(7.09)	0.72***	(6.68)
Do-not-track setting			−0.02	(−0.25)	−0.07	(−0.91)	−0.09	(−1.19)
Name in e-mail			−0.28***	(−5.67)	−0.29***	(−5.70)	−0.29***	(−5.59)
Number in e-mail			0.26***	(4.50)	0.23***	(3.91)	0.22***	(3.85)
Is lowercase			0.76***	(13.10)	0.74***	(13.20)	0.74***	(13.24)
E-mail error			1.66***	(20.00)	1.67***	(20.36)	1.70***	(20.37)
Constant	12.42***	(5.76)	−4.92***	(−62.87)	9.97***	(4.48)	9.04***	(4.06)

3.3 主回归：数字足迹的预测有效性

Multivariate results: predictive effectiveness of digital footprint

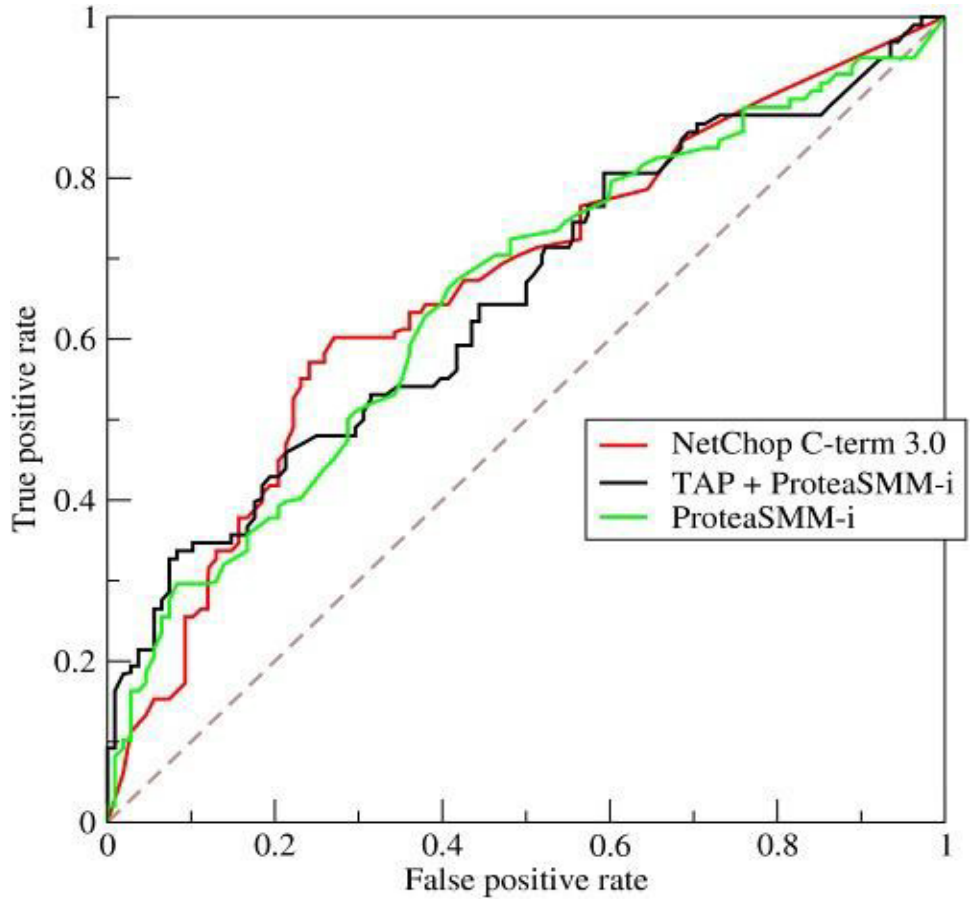
Table 4
Default regressions (scorable customers)

	(1) Credit bureau bureau score		(2) Digital footprint		(3) Credit bureau score & digital footprint		(4) Credit bureau score & digital footprint, further controls	
Variables	Coef.	z-stat	Coef.	z-stat	Coef.	z-stat	Coef.	z-stat
Credit bureau score	−0.17***	(−7.89)			−0.15***	(−6.67)	−0.14***	(−5.90)
Device type & operating system[a]								
Desktop/Windows			Baseline		Baseline		Baseline	
Desktop/Macintosh			−0.07	(−0.53)	−0.13	(−1.03)	−0.19	(−1.52)
Tablet/Android			0.29***	(3.19)	0.29***	(3.06)	0.33***	(3.44)
Tablet/iOS			0.08	(1.05)	0.08	(0.97)	0.07	(0.89)
Mobile/Android			1.05***	(17.25)	0.95***	(15.34)	1.01***	(16.13)
Mobile/iOS			0.72***	(9.07)	0.57***	(6.73)	0.61***	(7.26)
E-mail Host [a]								
Gmx (partly paid)			Baseline		Baseline		Baseline	
Web (partly paid)			0.00	(0.00)	−0.02	(−0.22)	−0.01	(−0.08)
T-Online (affluent customers)			−0.40***	(−3.90)	−0.35***	(−3.35)	−0.27**	(−2.47)
Gmail (free)			0.34***	(3.81)	0.29***	(3.09)	0.27***	(2.86)
Yahoo (free, older service)			0.75***	(9.19)	0.72***	(8.98)	0.70***	(8.28)
Hotmail (free, older service)			0.35***	(3.70)	0.28***	(2.72)	0.25**	(2.32)
Channel								
Paid			Baseline		Baseline		Baseline	
Affiliate			−0.49***	(−5.35)	−0.54***	(−5.58)	−0.61***	(−6.31)
Direct			−0.27***	(−4.25)	−0.28***	(−4.44)	−0.26***	(−4.30)
Organic			−0.15*	(−1.79)	−0.15*	(−1.74)	−0.15*	(−1.82)
Other			−0.47***	(−4.50)	−0.48***	(−4.36)	−0.39***	(−3.43)
Checkout time								
Evening (6 p.m.-midnight)			Baseline		Baseline		Baseline	
Morning (6 a.m.-noon)			0.28***	(4.50)	0.28***	(4.60)	0.29***	(4.75)
Afternoon (noon-6 p.m.)			0.08	(1.42)	0.08	(1.47)	0.10*	(1.92)
Night (midnight-6 a.m.)			0.79***	(7.73)	0.75***	(7.09)	0.72***	(6.68)
Do-not-track setting			−0.02	(−0.25)	−0.07	(−0.91)	−0.09	(−1.19)
Name in e-mail			−0.28***	(−5.67)	−0.29***	(−5.70)	−0.29***	(−5.59)
Number in e-mail			0.26***	(4.50)	0.23***	(3.91)	0.22***	(3.85)
Is lowercase			0.76***	(13.10)	0.74***	(13.20)	0.74***	(13.24)
E-mail error			1.66***	(20.00)	1.67***	(20.36)	1.70***	(20.37)
Constant	12.42***	(5.76)	−4.92***	(−62.87)	9.97***	(4.48)	9.04***	(4.06)

3.3 主回归：数字足迹的预测有效性

Multivariate results: predictive effectiveness of digital footprint

有效性评价指标：AUC



AUC（Area Under Curve）

- 每种预测方法都有：

  假阳性率：指在实际阴性例中，被分类器错误判定为阳性例的样本比例。（类似于假设检验中二类错误）

  真阳性率：指在实际阳性例中，被分类器正确判定为阳性例的样本比例。

- AUC是ROC曲线下与坐标轴围成的面积

  ROC曲线是一种显示分类器在不同阈值下的真阳性率和假阳性率之间权衡的曲线。横轴是“假阳性率“，纵轴是“真阳性率”。

根据AUC判断某种预测方法的预测能力

- 0.5 < AUC <1
- AUC越接近1.0，检测方法真实性越高；等于0.5时，则真实性最低(纯随机预测)，无应用价值。

3.3 再看主回归：数字足迹的预测有效性

Multivariate results: predictive effectiveness of digital footprint

Table 4
Default regressions (scorable customers)

Variables	(1) Credit bureau bureau score Coef.	z-stat	(2) Digital footprint Coef.	z-stat	(3) Credit bureau score & digital footprint Coef.	z-stat	(4) Credit bureau score & digital footprint, further controls Coef.	z-stat
Credit bureau score	−0.17***	(−7.89)			−0.15***	(−6.67)	−0.14***	(−5.90)
Device type & operating system[a]								
.....								
Constant	12.42***	(5.76)	−4.92***	(−62.87)	9.97***	(4.48)	9.04***	(4.06)
Control for *Age*, *Gender*, *Item category*, *Loan amount*, and month and region fixed effects	No		No		No		Yes	
Observations	254,819		254,819		254,819		254,613	
Pseudo R^2	**.0244**		**.0524**		**.0717**		**.0921**	
AUC	**0.683**		**0.696**		**0.736**		**0.762**	
(SE)	**(0.006)**		**(0.006)**		**(0.005)**		**(0.005)**	
Difference to AUC=50%	**0.183*** **		**0.196*** **		**0.236*** **		**0.262*** **	
Difference AUC to (1)			**0.013***		**0.053*** **		**0.080*** **	

	(1)		(2)		(3)		(4)	
system[a]								
Desktop/Windows	Baseline		Baseline		Baseline			
Desktop/Macintosh	−0.07	(−0.53)	−0.13	(−1.03)	−0.19	(−1.52)		
Tablet/Android	0.29***	(3.19)	0.29***	(3.06)	0.33***	(3.44)		
Tablet/iOS	0.08	(1.05)	0.08	(0.97)	0.07	(0.89)		
Mobile/Android	1.05***	(17.25)	0.95***	(15.34)	1.01***	(16.13)		
Mobile/iOS	0.72***	(9.07)	0.57***	(6.73)	0.61***	(7.26)		
E-mail Host [a]								
Gmx (partly paid)	Baseline		Baseline		Baseline			
Web (partly paid)	0.00	(0.00)	−0.02	(−0.22)	−0.01	(−0.08)		
T-Online (affluent customers)	−0.40***	(−3.90)	−0.35***	(−3.35)	−0.27**	(−2.47)		
Gmail (free)	0.34***	(3.81)	0.29***	(3.09)	0.27***	(2.86)		
Yahoo (free, older service)	0.75***	(9.19)	0.72***	(8.98)	0.70***	(8.28)		
Hotmail (free, older service)	0.35***	(3.70)	0.28***	(2.72)	0.25**	(2.32)		
Channel								
Paid	Baseline		Baseline		Baseline			
Affiliate	−0.49***	(−5.35)	−0.54***	(−5.58)	−0.61***	(−6.31)		
Direct	−0.27***	(−4.25)	−0.28***	(−4.44)	−0.26***	(−4.30)		
Organic	−0.15*	(−1.79)	−0.15*	(−1.74)	−0.15*	(−1.82)		
Other	−0.47***	(−4.50)	−0.48***	(−4.36)	−0.39***	(−3.43)		
Checkout time								
Evening (6 p.m.-midnight)	Baseline		Baseline		Baseline			
Morning (6 a.m.-noon)	0.28***	(4.50)	0.28***	(4.60)	0.29***	(4.75)		
Afternoon (noon-6 p.m.)	0.08	(1.42)	0.08	(1.47)	0.10*	(1.92)		
Night (midnight-6 a.m.)	0.79***	(7.73)	0.75***	(7.09)	0.72***	(6.68)		
Do-not-track setting	−0.02	(−0.25)	−0.07	(−0.91)	−0.09	(−1.19)		
Name in e-mail	−0.28***	(−5.67)	−0.29***	(−5.70)	−0.29***	(−5.59)		
Number in e-mail	0.26***	(4.50)	0.23***	(3.91)	0.22***	(3.85)		
Is lowercase	0.76***	(13.10)	0.74***	(13.20)	0.74***	(13.24)		
E-mail error	1.66***	(20.00)	1.67***	(20.36)	1.70***	(20.37)		

信用局评分对违约率有较强的预测能力

3.3 再看主回归：数字足迹的预测有效性

Multivariate results: predictive effectiveness of digital footprint

Table 4
Default regressions (scorable customers)

Variables	(1) Credit bureau bureau score Coef.	z-stat	(2) Digital footprint Coef.	z-stat	(3) Credit bureau score & digital footprint Coef.	z-stat	(4) Credit bureau score & digital footprint, further controls Coef.	z-stat
Credit bureau score	−0.17***	(−7.89)			−0.15***	(−6.67)	−0.14***	(−5.90)
Device type & operating system[a]								
.....								
Constant	12.42***	(5.76)	−4.92***	(−62.87)	9.97***	(4.48)	9.04***	(4.06)
Control for *Age*, *Gender*, *Item category*, *Loan amount*, and month and region fixed effects	No		No		No		Yes	
Observations	254,819		254,819		254,819		254,613	
Pseudo R^2	**.0244**		**.0524**		**.0717**		**.0921**	
AUC	**0.683**		**0.696**		**0.736**		**0.762**	
(SE)	**(0.006)**		**(0.006)**		**(0.005)**		**(0.005)**	
Difference to AUC=50%	**0.183*** **		**0.196*** **		**0.236*** **		**0.262*** **	
Difference AUC to (1)			**0.013***		**0.053*** **		**0.080*** **	

	(1)		(2)		(3)		(4)	
system[a]								
Desktop/Windows	Baseline		Baseline		Baseline			
Desktop/Macintosh	−0.07	(−0.53)	−0.13	(−1.03)	−0.19	(−1.52)		
Tablet/Android	0.29***	(3.19)	0.29***	(3.06)	0.33***	(3.44)		
Tablet/iOS	0.08	(1.05)	0.08	(0.97)	0.07	(0.89)		
Mobile/Android	1.05***	(17.25)	0.95***	(15.34)	1.01***	(16.13)		
Mobile/iOS	0.72***	(9.07)	0.57***	(6.73)	0.61***	(7.26)		
E-mail Host [a]								
Gmx (partly paid)	Baseline		Baseline		Baseline			
Web (partly paid)	0.00	(0.00)	−0.02	(−0.22)	−0.01	(−0.08)		
T-Online (affluent customers)	−0.40***	(−3.90)	−0.35***	(−3.35)	−0.27**	(−2.47)		
Gmail (free)	0.34***	(3.81)	0.29***	(3.09)	0.27***	(2.86)		
Yahoo (free, older service)	0.75***	(9.19)	0.72***	(8.98)	0.70***	(8.28)		
Hotmail (free, older service)	0.35***	(3.70)	0.28***	(2.72)	0.25**	(2.32)		
Channel								
Paid	Baseline		Baseline		Baseline			
Affiliate	−0.49***	(−5.35)	−0.54***	(−5.58)	−0.61***	(−6.31)		
Direct	−0.27***	(−4.25)	−0.28***	(−4.44)	−0.26***	(−4.30)		
Organic	−0.15*	(−1.79)	−0.15*	(−1.74)	−0.15*	(−1.82)		
Other	−0.47***	(−4.50)	−0.48***	(−4.36)	−0.39***	(−3.43)		
Checkout time								
Evening (6 p.m.-midnight)	Baseline		Baseline		Baseline			
Morning (6 a.m.-noon)	0.28***	(4.50)	0.28***	(4.60)	0.29***	(4.75)		
Afternoon (noon-6 p.m.)	0.08	(1.42)	0.08	(1.47)	0.10*	(1.92)		
Night (midnight-6 a.m.)	0.79***	(7.73)	0.75***	(7.09)	0.72***	(6.68)		
Do-not-track setting	−0.02	(−0.25)	−0.07	(−0.91)	−0.09	(−1.19)		
Name in e-mail	−0.28***	(−5.67)	−0.29***	(−5.70)	−0.29***	(−5.59)		
Number in e-mail	0.26***	(4.50)	0.23***	(3.91)	0.22***	(3.85)		
Is lowercase	0.76***	(13.10)	0.74***	(13.20)	0.74***	(13.24)		
E-mail error	1.66***	(20.00)	1.67***	(20.36)	1.70***	(20.37)		

简单、易获得的数字足迹变量对违约率的预测效果与信用局评分相当

3.3 再看主回归：数字足迹的预测有效性

Multivariate results: predictive effectiveness of digital footprint

Table 4
Default regressions (scorable customers)

Variables	(1) Credit bureau bureau score Coef.	z-stat	(2) Digital footprint Coef.	z-stat	(3) Credit bureau score & digital footprint Coef.	z-stat	(4) Credit bureau score & digital footprint, further controls Coef.	z-stat
Credit bureau score	−0.17***	(−7.89)			−0.15***	(−6.67)	−0.14***	(−5.90)
Device type & operating system[a]								
.....								
Constant	12.42***	(5.76)	−4.92***	(−62.87)	9.97***	(4.48)	9.04***	(4.06)
Control for *Age*, *Gender*, *Item category*, *Loan amount*, and month and region fixed effects	No		No		No		Yes	
Observations	254,819		254,819		254,819		254,613	
Pseudo R^2	**.0244**		**.0524**		**.0717**		**.0921**	
AUC	**0.683**		**0.696**		**0.736**		**0.762**	
(SE)	**(0.006)**		**(0.006)**		**(0.005)**		**(0.005)**	
Difference to AUC=50%	**0.183*** **		**0.196*** **		**0.236*** **		**0.262*** **	
Difference AUC to (1)			**0.013***		**0.053*** **		**0.080*** **	

	(1)		(2)		(3)		(4)	
system[a]								
Desktop/Windows	Baseline		Baseline		Baseline			
Desktop/Macintosh	−0.07	(−0.53)	−0.13	(−1.03)	−0.19	(−1.52)		
Tablet/Android	0.29***	(3.19)	0.29***	(3.06)	0.33***	(3.44)		
Tablet/iOS	0.08	(1.05)	0.08	(0.97)	0.07	(0.89)		
Mobile/Android	1.05***	(17.25)	0.95***	(15.34)	1.01***	(16.13)		
Mobile/iOS	0.72***	(9.07)	0.57***	(6.73)	0.61***	(7.26)		
E-mail Host [a]								
Gmx (partly paid)	Baseline		Baseline		Baseline			
Web (partly paid)	0.00	(0.00)	−0.02	(−0.22)	−0.01	(−0.08)		
T-Online (affluent customers)	−0.40***	(−3.90)	−0.35***	(−3.35)	−0.27**	(−2.47)		
Gmail (free)	0.34***	(3.81)	0.29***	(3.09)	0.27***	(2.86)		
Yahoo (free, older service)	0.75***	(9.19)	0.72***	(8.98)	0.70***	(8.28)		
Hotmail (free, older service)	0.35***	(3.70)	0.28***	(2.72)	0.25**	(2.32)		
Channel								
Paid	Baseline		Baseline		Baseline			
Affiliate	−0.49***	(−5.35)	−0.54***	(−5.58)	−0.61***	(−6.31)		
Direct	−0.27***	(−4.25)	−0.28***	(−4.44)	−0.26***	(−4.30)		
Organic	−0.15*	(−1.79)	−0.15*	(−1.74)	−0.15*	(−1.82)		
Other	−0.47***	(−4.50)	−0.48***	(−4.36)	−0.39***	(−3.43)		
Checkout time								
Evening (6 p.m.-midnight)	Baseline		Baseline		Baseline			
Morning (6 a.m.-noon)	0.28***	(4.50)	0.28***	(4.60)	0.29***	(4.75)		
Afternoon (noon-6 p.m.)	0.08	(1.42)	0.08	(1.47)	0.10*	(1.92)		
Night (midnight-6 a.m.)	0.79***	(7.73)	0.75***	(7.09)	0.72***	(6.68)		
Do-not-track setting	−0.02	(−0.25)	−0.07	(−0.91)	−0.09	(−1.19)		
Name in e-mail	−0.28***	(−5.67)	−0.29***	(−5.70)	−0.29***	(−5.59)		
Number in e-mail	0.26***	(4.50)	0.23***	(3.91)	0.22***	(3.85)		
Is lowercase	0.76***	(13.10)	0.74***	(13.20)	0.74***	(13.24)		
E-mail error	1.66***	(20.00)	1.67***	(20.36)	1.70***	(20.37)		

数字足迹可以作为信用局评分的补充而非替代

3.3 再看主回归：数字足迹的预测有效性

Multivariate results: predictive effectiveness of digital footprint

Table 4
Default regressions (scorable customers)

Variables	(1) Credit bureau bureau score Coef.	z-stat	(2) Digital footprint Coef.	z-stat	(3) Credit bureau score & digital footprint Coef.	z-stat	(4) Credit bureau score & digital footprint, further controls Coef.	z-stat
Credit bureau score	−0.17***	(−7.89)			−0.15***	(−6.67)	−0.14***	(−5.90)
Device type & operating system[a]								
.....								
Constant	12.42***	(5.76)	−4.92***	(−62.87)	9.97***	(4.48)	9.04***	(4.06)
Control for *Age*, *Gender*, *Item category*, *Loan amount*, and month and region fixed effects	No		No		No		Yes	
Observations	254,819		254,819		254,819		254,613	
Pseudo R^2	**.0244**		**.0524**		**.0717**		**.0921**	
AUC	**0.683**		**0.696**		**0.736**		**0.762**	
(SE)	**(0.006)**		**(0.006)**		**(0.005)**		**(0.005)**	
Difference to AUC=50%	**0.183*** **		**0.196*** **		**0.236*** **		**0.262*** **	
Difference AUC to (1)			**0.013***		**0.053*** **		**0.080*** **	

	(1)		(2)		(3)		(4)	
system[a]								
Desktop/Windows	Baseline		Baseline		Baseline			
Desktop/Macintosh	−0.07	(−0.53)	−0.13	(−1.03)	−0.19	(−1.52)		
Tablet/Android	0.29***	(3.19)	0.29***	(3.06)	0.33***	(3.44)		
Tablet/iOS	0.08	(1.05)	0.08	(0.97)	0.07	(0.89)		
Mobile/Android	1.05***	(17.25)	0.95***	(15.34)	1.01***	(16.13)		
Mobile/iOS	0.72***	(9.07)	0.57***	(6.73)	0.61***	(7.26)		
E-mail Host [a]								
Gmx (partly paid)	Baseline		Baseline		Baseline			
Web (partly paid)	0.00	(0.00)	−0.02	(−0.22)	−0.01	(−0.08)		
T-Online (affluent customers)	−0.40***	(−3.90)	−0.35***	(−3.35)	−0.27**	(−2.47)		
Gmail (free)	0.34***	(3.81)	0.29***	(3.09)	0.27***	(2.86)		
Yahoo (free, older service)	0.75***	(9.19)	0.72***	(8.98)	0.70***	(8.28)		
Hotmail (free, older service)	0.35***	(3.70)	0.28***	(2.72)	0.25**	(2.32)		
Channel								
Paid	Baseline		Baseline		Baseline			
Affiliate	−0.49***	(−5.35)	−0.54***	(−5.58)	−0.61***	(−6.31)		
Direct	−0.27***	(−4.25)	−0.28***	(−4.44)	−0.26***	(−4.30)		
Organic	−0.15*	(−1.79)	−0.15*	(−1.74)	−0.15*	(−1.82)		
Other	−0.47***	(−4.50)	−0.48***	(−4.36)	−0.39***	(−3.43)		
Checkout time								
Evening (6 p.m.-midnight)	Baseline		Baseline		Baseline			
Morning (6 a.m.-noon)	0.28***	(4.50)	0.28***	(4.60)	0.29***	(4.75)		
Afternoon (noon-6 p.m.)	0.08	(1.42)	0.08	(1.47)	0.10*	(1.92)		
Night (midnight-6 a.m.)	0.79***	(7.73)	0.75***	(7.09)	0.72***	(6.68)		
Do-not-track setting	−0.02	(−0.25)	−0.07	(−0.91)	−0.09	(−1.19)		
Name in e-mail	−0.28***	(−5.67)	−0.29***	(−5.70)	−0.29***	(−5.59)		
Number in e-mail	0.26***	(4.50)	0.23***	(3.91)	0.22***	(3.85)		
Is lowercase	0.76***	(13.10)	0.74***	(13.20)	0.74***	(13.24)		
E-mail error	1.66***	(20.00)	1.67***	(20.36)	1.70***	(20.37)		

数字足迹和信用局评分都不是年龄、性别、日期、金额、购买项目类型等控制变量的替代

3.4 稳健性检验

Robustness tests

1. 前面回归（表4）的AUC是样本内的，是否存在过拟合?

2. “违约”被粗暴地定义为坏账移交至催款机构。结论对这一定义敏感吗?

3. 在主要的子样本上，结论还能成立吗?

4. 结论是在电商公司消费贷这个特例下得出的，它是否能外推到其他的贷款去?

3.4 稳健性检验

Robustness tests

①过拟合？——样本外

100×2折叠交叉验证

2015.10~2016.02——训练集
2016.03~2016.07——不使用(观察违约/不违约结果需要时间)
2016.08~2016.12——测试集

Table 5
Out-of-sample estimates

	(1) Baseline (in-sample)	(2) Out-of-sample	(3) Out-of-sample/out-of-time
AUC credit bureau score	0.683	0.681	0.691
N	254,819	254,819	74,543
AUC digital footprint	0.696	0.688	0.692
N	254,819	254,819	74,543
AUC credit bureau score + Digital footprint	0.736	0.728	0.739
N	254,819	254,819	74,543
AUC credit bureau score + Digital footprint, fixed effects	0.762	0.734	0.730
N	254,613	254,613	74,543

样本内回归和样本外回归的AUC非常相似

3.4 稳健性检验

Robustness tests

②改变违约的定义

Table 6
Robustness tests (scorable customers)

A. Default definition	(1)	(2)	(3)	(4)
	Baseline (default = transfer to collection agency)	Default = Write-down	Exclude cases of fraud (9% of defaults)	Loss given default (R^2 reported)
AUC credit bureau score	0.683	0.692	0.681	0.013
AUC Digital footprint	0.696	0.723	0.691	0.062
AUC credit bureau score + digital footprint	0.736	0.757	0.730	0.069
N	254,819	254,819	254,604	2,384
B. Sample splits	(1) Small orders < EUR 218.91	(2) Large orders ≥ EUR 218.91	(3) Female	(4) Male
AUC credit bureau score	0.688	0.678	0.689	0.670
AUC Digital footprint	0.711	0.689	0.697	0.700
AUC credit bureau score + digital footprint	0.749	0.729	0.743	0.724
N	127,410	127,409	168,374	86,445

面板A提供了使用替代“违约”定义的结果

列（1）
default定义为“转移到债务催收机构”

AUC增加

列（2）
default定义为“经催收机构催债后仍未偿还”

列（3）
不包括欺诈案件

表现：AUC值并未大幅降低

结论：数字足迹的预测能力及其相对较好的性能并不仅仅是由欺诈案件驱动的

列（4）
只使用违约贷款的样本

结论：数字足迹能够比信用局评分更好地预测违约造成的损失

3.4 稳健性检验

Robustness tests

③在子样本上是否成立

Table 6
Robustness tests (scorable customers)

A. Default definition	(1)	(2)	(3)	(4)
	Baseline (default = transfer to collection agency)	Default = Write-down	Exclude cases of fraud (9% of defaults)	Loss given default (R^2 reported)
AUC credit bureau score	0.683	0.692	0.681	0.013
AUC Digital footprint	0.696	0.723	0.691	0.062
AUC credit bureau score + digital footprint	0.736	0.757	0.730	0.069
N	254,819	254,819	254,604	2,384

B. Sample splits	(1) Small orders < EUR 218.91	(2) Large orders ≥ EUR 218.91	(3) Female	(4) Male
AUC credit bureau score	0.688	0.678	0.689	0.670
AUC Digital footprint	0.711	0.689	0.697	0.700
AUC credit bureau score + digital footprint	0.749	0.729	0.743	0.724
N	127,410	127,409	168,374	86,445

面板B提供了各种子样本的回归结果

表现：小订单和大订单（按中位数分割）以及女性和男性客户的结果非常相似。

结论：数字足迹的预测能力对于各种样本分割都是稳健的，且都优于信用局评分的预测能力。

④外部有效性(External Validity)

能否将“数字足迹”推广到其他类型贷款？

——验证今天的数字足迹是否可以预测机构信用评分的未来变化。

信用机构评分的变化　　使用数字足迹和用信用机构评分预测的违约率的差异

$$\Delta(CreditScore_{t+1}, CreditScore_t) = \beta_0 + \beta_1 \Delta(DF_t, CreditScore_t) + X + \varepsilon,$$

Table 7
Predicting changes in the credit bureau score with the digital footprint

Dependent variable	(1) Δ (CreditScore$_{t+1}$, CreditScore$_t$)	(2) Δ (CreditScore$_{t+1}$, CreditScore$_t$)	(3) Δ (CreditScore$_{t+1}$, CreditScore$_t$)	(4) Δ (CreditScore$_{t+1}$, CreditScore$_t$)	(5) Δ (CreditScore$_{t+1}$, CreditScore$_t$)
Δ (DigitalFootprint$_t$, CreditBureauScore$_t$)	−75.86***	−28.43***	−30.11***		
	(−11.86)	(−4.64)	(−5.05)		
Q1 (-100% to -0.49%)				0.40**	
				(2.52)	
Q2 (-0.49% to -0.25%)				0.15*	
				(1.75)	
Q3 (-0.25% to -0.05%)				baseline	
Q4 (-0.05% to +0.35%)				0.08	
Q4 (-0.05% to +0.35%)				(0.91)	
Q5 (+0.35% to +100%)				−0.39***	
				(−3.04)	
DigitalFootprint-Better-Than-CreditBureauScore (0/1)					0.33**
DigitalFootprint-Better-Than-CreditBureauScore (0/1) x					(2.14)
LowCreditBureauScore					0.86**
					(2.36)
Q2					0.03
					(0.13)
Q3					baseline
Q4					−0.13
					(−0.71)
HighCreditBureauScore					0.00
					(0.01)
CreditBureauScore$_t$		−0.43***	−0.42***	−0.42***	FE for each credit score quintile
		(−13.47)	(−13.28)	(−10.05)	
Constant	0.37***	41.99***			
	(8.75)	(13.51)	absorbed	absorbed	absorbed
Month & region fixed effects	No	No	Yes	Yes	Yes
Observations	17,646	17,646	17,646	17,646	17,646
Adj. R^2	.028	.071	.081	.081	.074

PART 4

经济结果和启示

Economic Outcomes and Implications

4.1 经济机制：数字足迹预测能力的来源

Economic mechanism

Table 8——Panel A

将数字足迹的总体信息内容**分解为单独的变量**：

Table 8

Marginal AUC for digital footprint variables and combinations of digital footprint variables

A. Individual digital footprint variables (dependent variable: default (0/1))

Variable	Stand-alone AUC (%)	Marginal AUC (PP)
Computer & operating system	59.03	+1.71***
E-mail host	59.78	+2.44***
E-mail Host: paid versus nonpaid dummy	53.80	+0.98***
E-mail Host: Variation within nonpaid e-mail hosts	57.82	+1.79***
Channel	54.95	+0.70***
Checkout time	53.56	+0.63***
Do not track setting	50.40	+0.14*
Name in e-mail	54.61	+0.30**
Number in e-mail	54.15	+0.19**
Is lowercase	54.91	+1.15***
E-mail error	53.08	+1.78***

- **列表中没有一个变量占主导地位**

注：$Marginal\ AUC_x$ = 使用所有数字足迹变量的完整模型的AUC − 使用除变量x以外的所有变量的模型的AUC

4.1 经济机制：数字足迹预测能力的来源

Economic mechanism

Table 8——Panel A

将数字足迹的总体信息内容**分解为单独的变量**：

Table 8
Marginal AUC for digital footprint variables and combinations of digital footprint variables

A. Individual digital footprint variables (dependent variable: default (0/1))

Variable	Stand-alone AUC (%)	Marginal AUC (PP)
Computer & operating system	59.03	+1.71***
E-mail host	59.78	+2.44***
E-mail Host: paid versus nonpaid dummy	53.80	+0.98***
E-mail Host: Variation within nonpaid e-mail hosts	57.82	+1.79***
Channel	54.95	+0.70***
Checkout time	53.56	+0.63***
Do not track setting	50.40	+0.14*
Name in e-mail	54.61	+0.30**
Number in e-mail	54.15	+0.19**
Is lowercase	54.91	+1.15***
E-mail error	53.08	+1.78***

- **列表中没有一个变量占主导地位**

Table 8——Panel B

将数字足迹变量进行**组合**：

B. Combinations of digital footprint variables (dependent variable: default (0/1))

Variables	Stand-alone AUC (%)	Marginal AUC (PP)
Potential proxy for income		
Potential proxy for income, financially costly to change (computer & operating system, e-mail host: paid vs. nonpaid dummy)	61.03	+2.20
Unlikely to be a proxy for income, not financially costly to change (nonpaid e-mail host, channel, checkout time, do not track setting, name in e-mail, number in e-mail, is lowercase, e-mail error)	67.35	+8.52
Impact on everyday behavior		
Requires one-time action only (computer & operating system, e-mail host, do not track setting, name in e-mail, number in e-mail)	64.92	+7.25
Requires thinking about how to behave during every individual purchase (channel, checkout time, is lowercase, e-mail error)	62.30	+4.63

- **分类标准1：根据是否反映财务特征分类**
  - 不太可能代表收入的变量有更高的独立AUC和边际AUC
  - 数字足迹包含的信息超过了纯粹的财务特征
- **分类标准2：根据对日常行为、性格的影响分类**
  - 由单一行为决定的变量和每次购买中重新决定的变量都显著促进了数字足迹的信息性

注：$Marginal\ AUC_x$ = 使用所有数字足迹变量的完整模型的AUC − 使用除变量x以外的所有变量的模型的AUC

4.1 经济机制：数字足迹预测能力的来源

Economic mechanism

Table 8——Panel A

将数字足迹的总体信息内容**分解为单独的变量**：

Table 8
Marginal AUC for digital footprint variables and combinations of digital footprint variables

A. Individual digital footprint variables (dependent variable: default (0/1))

Variable	Stand-alone AUC (%)	Marginal AUC (PP)
Computer & operating system	59.03	+1.71***
E-mail host	59.78	+2.44***
E-mail Host: paid versus nonpaid dummy	53.80	+0.98***
E-mail Host: Variation within nonpaid e-mail hosts	57.82	+1.79***
Channel	54.95	+0.70***
Checkout time	53.56	+0.63***
Do not track setting	50.40	+0.14*
Name in e-mail	54.61	+0.30**
Number in e-mail	54.15	+0.19**
Is lowercase	54.91	+1.15***
E-mail error	53.08	+1.78***

- **列表中没有一个变量占主导地位**

Table 8——Panel B

将数字足迹变量进行**组合**：

B. Combinations of digital footprint variables (dependent variable: default (0/1))

Variables	Stand-alone AUC (%)	Marginal AUC (PP)
Potential proxy for income		
Potential proxy for income, financially costly to change (computer & operating system, e-mail host: paid vs. nonpaid dummy)	61.03	+2.20
Unlikely to be a proxy for income, not financially costly to change (nonpaid e-mail host, channel, checkout time, do not track setting, name in e-mail, number in e-mail, is lowercase, e-mail error)	67.35	+8.52
Impact on everyday behavior		
Requires one-time action only (computer & operating system, e-mail host, do not track setting, name in e-mail, number in e-mail)	64.92	+7.25
Requires thinking about how to behave during every individual purchase (channel, checkout time, is lowercase, e-mail error)	62.30	+4.63

- **分类标准1：根据是否反映财务特征分类**
  - 不太可能代表收入的变量有更高的独立AUC和边际AUC
  - 数字足迹包含的信息超过了纯粹的财务特征
- **分类标准2：根据对日常行为、性格的影响分类**
  - 由单一行为决定的变量和每次购买中重新决定的变量都显著促进了数字足迹的信息性

注：$Marginal\ AUC_x$ = 使用所有数字足迹变量的完整模型的AUC − 使用除变量x以外的所有变量的模型的AUC

4.1 经济机制：数字足迹预测能力的来源

Economic mechanism

- 重要启示：
  - ✓ 一些调查显示，贷款申请人在网上申请贷款时，甚至不愿提供非常基本的信息（尤其是财务信息），如他们的银行账号或信用卡号码；
  - ✓ 但是数字足迹很容易收集(不需要申请人提供和核实收入或银行账户信息，只需通过访问或注册网站来收集这些信息)；
  - ✓ 为提高客户体验和节约信息收集成本提供了显著优势，这对于小批量/大数量零售业务尤为重要。

4.2 引入数字足迹对机构的影响

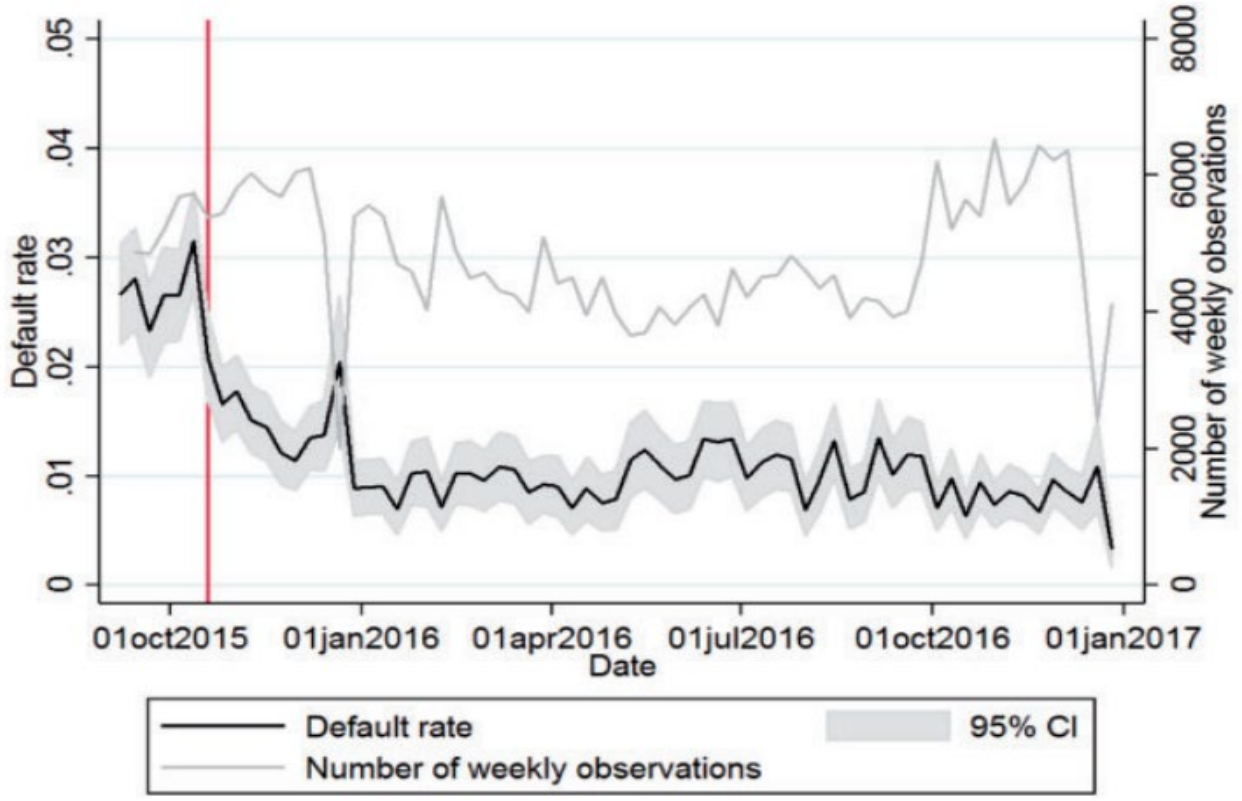
Access to credit and default rates at the E-commerce firm

- 是否提高了机构的利润?
  两个角度：①降低违约率 ②提高信贷接受率（机构接受货到付款的概率）

- 数据集的时间维度：2015.10.19向更早时间拓展，包含未使用数字足迹时的相关数据

4.2 引入数字足迹对机构的影响

Access to credit and default rates at the E-commerce firm



违约率显著下降

订单数量维持稳定

4.2 引入数字足迹对机构的影响

Access to credit and default rates at the E-commerce firm

- 是否提高了机构的利润?
  两个角度：①降低违约率 ②提高信贷接受率（机构接受货到付款的概率）

- 数据集的时间维度：2015.10.19向更早时间拓展，包含未使用数字足迹时的相关数据

- 电商接受货到付款的政策调整：
  - 2015年10月19日前，没有引入数字足迹，
    - 100~1100欧元：不应用信用评分判别，1100欧元以上：应用信用评分判别；
  - 2015年10月19日后，凡100欧元以上的订单均经过信用评分和数字足迹审查。

  - 这就自然地形成了两个子样本：
    - 100~1100欧元： 10月19日前后的接受率和违约率变动是信用评分和数字足迹的双重影响， “ScoreAndDFAdded”；
    - 1100欧元以上： 10月19日前后仅受数字足迹冲击， “DFAdded”。

4.2 引入数字足迹对机构的影响

Access to credit and default rates at the E-commerce firm

- 是否提高了机构的利润?
  两个角度：①降低违约率 ②提高信贷接受率（机构接受货到付款的概率）

- 数据集的时间维度：2015.10.19向更早时间拓展，包含未使用数字足迹时的相关数据

- 电商接受货到付款的政策调整：
  - 2015年10月19日前，没有引入数字足迹，
    - 100~1100欧元：不应用信用评分判别，1100欧元以上：应用信用评分判别；
  - 2015年10月19日后，凡100欧元以上的订单均经过信用评分和数字足迹审查。

  - 这就自然地形成了两个子样本：
    - 100~1100欧元： 10月19日前后的接受率和违约率变动是信用评分和数字足迹的双重影响， “ScoreAndDFAdded”；
    - 1100欧元以上： 10月19日前后仅受数字足迹冲击， “DFAdded”。

4.2 引入数字足迹对机构的影响

Access to credit and default rates at the E-commerce firm

Table 9
Development of default rates and access to credit around the introduction of the digital footprint (Univariate results)

		Default rate (%)			Invoice offered (%)			Credit bureau score		
	N	Pre	Post	Δ	Pre	Post	Δ	Pre	Post	Δ
A. Categories										
Sample 1: ScoreAndDFAdded	33,896	2.54	1.19	−1.36***	96.65	90.05	−6.60***	na	98.26	na
Sample 2: DFAdded	10,807	3.62	2.33	−1.29***	39.00	40.11	1.11***	97.82	97.84	0.02
B. Subcategories of "DFAdded"										
DFAdded / High score	3,614	0.84	0.88	0.04	90.00	90.94	0.95	99.42	99.42	0.00
DFAdded / Medium score	4,023	1.82	2.14	0.33	85.21	87.72	2.50***	98.17	98.16	0.00
DFAdded / Low score	2,088	6.33	3.75	−2.57***	31.59	27.52	−4.07***	94.45	94.41	−0.04
DFAdded / Unscorable	1,082	11.65	6.44	−5.22***	10.14	9.59	−0.54	na	na	na

- Sample1(ScoreAndDFAdded)：在10月19日后加入信用局分数和数字足迹
  - ✓ 违约率下降约53%，利润：$0.9665 \times (0.1 - 0.0254) \rightarrow 0.9005 \times (0.1 - 0.0119)$，即从约7.2%变化至约8.9%（电商毛利预估为10%），也增加；
  - ✓ 但无法区分信用局评分和数字足迹对于此样本的影响。

- Sample2(DFAdded)：在10余19日后仅加入数字足迹
  - ✓ 违约率下降约42%，较低的违约率，加上较高的信贷接受率，故利润必然增加。

4.2 引入数字足迹对机构的影响

Access to credit and default rates at the E-commerce firm

Table 10——稳健性检验

	重现Table 9 Panel A 单变量结果		控制变量时间变化	划分子样本探究违约率降低的驱动因素	时间范围缩小	一年后的安慰剂检验
Dependent variable	(1) Default (0/1)	(2) Default (0/1)	(3) Default (0/1)	(4) Default (0/1)	(5) Default (0/1)	(6) Default (
Method	Difference post vs. pre	Difference post vs pre, add categories	Add time trend, controls and FEs	Add subcategories	Narrower window around Oct. 19, 2015	Placebo t 1-year la
Sample	±6 weeks	±6 weeks	±6 weeks	±6 weeks	±4 weeks	±4 wee
Post	−0.014***					
	(−9.12)					
Post x ScoreAndDFAdded		−0.014***	−0.014***	−0.015***	−0.015***	0.001
		(−8.55)	(−5.88)	(−6.13)	(−4.30)	(0.29)
Post x DFAdded		−0.013***	−0.012***			
		(−3.85)	(−3.04)			
Post x "DFAdded / High score"				−0.001	0.000	0.002
				(−0.19)	(0.00)	(0.78)
Post x "DFAdded / Medium score"				0.003	0.003	0.004
				(0.65)	(0.46)	(1.07)
Post x "DFAdded / Low score"				−0.026**	−0.021*	−0.015
				(−2.51)	(−1.70)	(−1.50)
Post x "DFAdded / Unscorable"				−0.052***	−0.059***	0.007
				(−2.72)	(−2.66)	(0.43)
Time trend	*No*	*No*	0.000	0.001	0.001	−0.002
			(0.29)	(0.53)	(0.15)	(−0.80)
Category FE (=variables from interaction terms as noninteracted variables)	*No*	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes	Yes
Fixed effects	No	No	Yes	Yes	Yes	Yes
Observations	44,703	44,703	44,703	44,703	30,322	28,905
Adj. R^2	.002	.003	.012	.021	.020	.012

假设前提

- 如果没有筛查技术的改变（即引入数字足迹变量和信用局评分）违约率将**保持稳定**

列（1）&列（2）

- 重现表9面板A单变量结果
- 表明引入数字足迹后，违约率下降了1.3-1.4pp

4.2 引入数字足迹对机构的影响

Access to credit and default rates at the E-commerce firm

Table 10——稳健性检验

	重现Table 9 Panel A 单变量结果		控制变量时间变化	划分子样本探究违约率降低的驱动因素	时间范围缩小	一年后的安慰剂检验
Dependent variable	(1) Default (0/1)	(2) Default (0/1)	(3) Default (0/1)	(4) Default (0/1)	(5) Default (0/1)	(6) Default (
Method	Difference post vs. pre	Difference post vs. pre, add categories	Add time trend, controls and FEs	Add subcategories	Narrower window around Oct. 19, 2015	Placebo t 1-year la
Sample	±6 weeks	±6 weeks	±6 weeks	±6 weeks	±4 weeks	±4 wee
Post	−0.014***					
	(−9.12)					
Post x ScoreAndDFAdded		−0.014***	−0.014***	−0.015***	−0.015***	0.001
		(−8.55)	(−5.88)	(−6.13)	(−4.30)	(0.29)
Post x DFAdded		−0.013***	−0.012***			
		(−3.85)	(−3.04)			
Post x "DFAdded / High score"				−0.001	0.000	0.002
				(−0.19)	(0.00)	(0.78)
Post x "DFAdded / Medium score"				0.003	0.003	0.004
				(0.65)	(0.46)	(1.07)
Post x "DFAdded / Low score"				−0.026**	−0.021*	−0.015
				(−2.51)	(−1.70)	(−1.50)
Post x "DFAdded / Unscorable"				−0.052***	−0.059***	0.007
				(−2.72)	(−2.66)	(0.43)
Time trend	*No*	*No*	0.000	0.001	0.001	−0.002
			(0.29)	(0.53)	(0.15)	(−0.80)
Category FE (=variables from interaction terms as noninteracted variables)	*No*	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes	Yes
Fixed effects	No	No	Yes	Yes	Yes	Yes
Observations	44,703	44,703	44,703	44,703	30,322	28,905
Adj. R^2	.002	.003	.012	.021	.020	.012

列（3）

- 采购构成的变化（不同地区、不同商品类别、不同性别的采购）
- 经济的整体改善
- **控制可观察特征**（购买项目类别、性别以及区域固定效应）
- **引入时间趋势**
- **无影响**

4.2 引入数字足迹对机构的影响

Access to credit and default rates at the E-commerce firm

Table 10——稳健性检验

	重现Table 9 Panel A 单变量结果		控制变量时间变化	划分子样本探究违约率降低的驱动因素	时间范围缩小	一年后的安慰剂检验
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Default (0/1)	Default (0/1)	Default (0/1)	Default (0/1)	Default (0/1)	Default (
Method	Difference post vs. pre	Difference post vs. pre, add categories	Add time trend, controls and FEs	Add subcategories	Narrower window around Oct. 19, 2015	Placebo t 1-year la
Sample	±6 weeks	±6 weeks	±6 weeks	±6 weeks	±4 weeks	±4 wee
Post	−0.014***					
	(−9.12)					
Post x ScoreAndDFAdded		−0.014***	−0.014***	−0.015***	−0.015***	0.001
		(−8.55)	(−5.88)	(−6.13)	(−4.30)	(0.29)
Post x DFAdded		−0.013***	−0.012***			
		(−3.85)	(−3.04)			
Post x "DFAdded / High score"				−0.001	0.000	0.002
				(−0.19)	(0.00)	(0.78)
Post x "DFAdded / Medium score"				0.003	0.003	0.004
				(0.65)	(0.46)	(1.07)
Post x "DFAdded / Low score"				−0.026**	−0.021*	−0.015
				(−2.51)	(−1.70)	(−1.50)
Post x "DFAdded / Unscorable"				−0.052***	−0.059***	0.007
				(−2.72)	(−2.66)	(0.43)
Time trend	*No*	*No*	0.000	0.001	0.001	−0.002
			(0.29)	(0.53)	(0.15)	(−0.80)
Category FE (=variables from interaction terms as noninteracted variables)	*No*	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes	Yes
Fixed effects	No	No	Yes	Yes	Yes	Yes
Observations	44,703	44,703	44,703	44,703	30,322	28,905
Adj. R^2	.002	.003	.012	.021	.020	.012

列（4）

- 划分子样本探究违约率降低的驱动因素

- 由于**不可评分**的客户和**评分较低**的客户**驱动**的

4.2 引入数字足迹对机构的影响

Access to credit and default rates at the E-commerce firm

Table 10——稳健性检验

	重现Table 9 Panel A 单变量结果		控制变量时间变化	划分子样本探究违约率降低的驱动因素	时间范围缩小	一年后的安慰剂检验
Dependent variable	(1) Default (0/1)	(2) Default (0/1)	(3) Default (0/1)	(4) Default (0/1)	(5) Default (0/1)	(6) Default (
Method	Difference post vs. pre	Difference post vs. pre, add categories	Add time trend, controls and FEs	Add subcategories	Narrower window around Oct. 19, 2015	Placebo t 1-year la
Sample	±6 weeks	±6 weeks	±6 weeks	±6 weeks	±4 weeks	±4 wee
Post	−0.014***					
	(−9.12)					
Post x ScoreAndDFAdded		−0.014***	−0.014***	−0.015***	−0.015***	0.001
		(−8.55)	(−5.88)	(−6.13)	(−4.30)	(0.29)
Post x DFAdded		−0.013***	−0.012***			
		(−3.85)	(−3.04)			
Post x "DFAdded / High score"				−0.001	0.000	0.002
				(−0.19)	(0.00)	(0.78)
Post x "DFAdded / Medium score"				0.003	0.003	0.004
				(0.65)	(0.46)	(1.07)
Post x "DFAdded / Low score"				−0.026**	−0.021*	−0.015
				(−2.51)	(−1.70)	(−1.50)
Post x "DFAdded / Unscorable"				−0.052***	−0.059***	0.007
				(−2.72)	(−2.66)	(0.43)
Time trend	*No*	*No*	0.000	0.001	0.001	−0.002
			(0.29)	(0.53)	(0.15)	(−0.80)
Category FE (=variables from interaction terms as noninteracted variables)	*No*	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes	Yes
Fixed effects	No	No	Yes	Yes	Yes	Yes
Observations	44,703	44,703	44,703	44,703	30,322	28,905
Adj. R^2	.002	.003	.012	.021	.020	.012

列（5）

- **较长窗口期**
- 部分经济变量可能发生明显变化
- **较短窗口期**
- 经济变量的变化可以忽略
- 将窗口期缩小为前后4周
- 更短的窗口期排除了所有基于缓慢变化的经济变量的替代解释

4.2 引入数字足迹对机构的影响

Access to credit and default rates at the E-commerce firm

Table 10——稳健性检验

	重现Table 9 Panel A 单变量结果		控制变量时间变化	划分子样本探究违约率降低的驱动因素	时间范围缩小	一年后的安慰剂检验
Dependent variable	(1) Default (0/1)	(2) Default (0/1)	(3) Default (0/1)	(4) Default (0/1)	(5) Default (0/1)	(6) Default (0/1)
Method	Difference post vs. pre	Difference post vs. pre, add categories	Add time trend, controls and FEs	Add subcategories	Narrower window around Oct. 19, 2015	Placebo test: 1-year later
Sample	±6 weeks	±6 weeks	±6 weeks	±6 weeks	±4 weeks	±4 weeks
Post	−0.014*** (−9.12)					
Post x ScoreAndDFAdded		−0.014*** (−8.55)	−0.014*** (−5.88)	−0.015*** (−6.13)	−0.015*** (−4.30)	0.001 (0.29)
Post x DFAdded		−0.013*** (−3.85)	−0.012*** (−3.04)			
Post x "DFAdded / High score"				−0.001 (−0.19)	0.000 (0.00)	0.002 (0.78)
Post x "DFAdded / Medium score"				0.003 (0.65)	0.003 (0.46)	0.004 (1.07)
Post x "DFAdded / Low score"				−0.026** (−2.51)	−0.021* (−1.70)	−0.015 (−1.50)
Post x "DFAdded / Unscorable"				−0.052*** (−2.72)	−0.059*** (−2.66)	0.007 (0.43)
Time trend	*No*	*No*	0.000 (0.29)	0.001 (0.53)	0.001 (0.15)	−0.002 (−0.80)
Category FE (=variables from interaction terms as noninteracted variables)	*No*	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes	Yes
Fixed effects	No	No	Yes	Yes	Yes	Yes
Observations	44,703	44,703	44,703	44,703	30,322	28,905
Adj. R^2	.002	.003	.012	.021	.020	.012

列（6）

- **不同的时间支付行为不同**
- **安慰剂测试**
- 结果表明：仅在2015年10.19前后违约率下降，但是2016年（筛查技术保持不变）并没有

4.2 引入数字足迹对机构的影响

Access to credit and default rates at the E-commerce firm

Table 10——稳健性检验

	重现Table 9 Panel A 单变量结果		控制变量时间变化	划分子样本探究违约率降低的驱动因素	时间范围缩小	一年后的安慰剂检验
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Default (0/1)	Default (0/1)	Default (0/1)	Default (0/1)	Default (0/1)	Default (
Method	Difference post vs. pre	Difference post vs. pre, add categories	Add time trend, controls and FEs	Add subcategories	Narrower window around Oct. 19, 2015	Placebo t 1-year la
Sample	±6 weeks	±6 weeks	±6 weeks	±6 weeks	±4 weeks	±4 wee
Post	−0.014***					
	(−9.12)					
Post x ScoreAndDFAdded		−0.014***	−0.014***	−0.015***	−0.015***	0.001
		(−8.55)	(−5.88)	(−6.13)	(−4.30)	(0.29)
Post x DFAdded		−0.013***	−0.012***			
		(−3.85)	(−3.04)			
Post x "DFAdded / High score"				−0.001	0.000	0.002
				(−0.19)	(0.00)	(0.78)
Post x "DFAdded / Medium score"				0.003	0.003	0.004
				(0.65)	(0.46)	(1.07)
Post x "DFAdded / Low score"				−0.026**	−0.021*	−0.015
				(−2.51)	(−1.70)	(−1.50)
Post x "DFAdded / Unscorable"				−0.052***	−0.059***	0.007
				(−2.72)	(−2.66)	(0.43)
Time trend	*No*	*No*	0.000	0.001	0.001	−0.002
			(0.29)	(0.53)	(0.15)	(−0.80)
Category FE (=variables from interaction terms as noninteracted variables)	*No*	Yes	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes	Yes
Fixed effects	No	No	Yes	Yes	Yes	Yes
Observations	44,703	44,703	44,703	44,703	30,322	28,905
Adj. R^2	.002	.003	.012	.021	.020	.012

引入数字足迹，违约率确实会下降，并且确实是因为引入数字足迹，违约率才下降。

附录：电子商务公司不是一个特殊案例！

4.3 帮助没有银行账户的人获得信贷

Access to credit for the unbanked

过去

- 由于缺乏信贷局评分，没有银行账户的人口无法参与金融服务。

数字足迹变量具有易得性，因此分析借款人的数字行为可能会提供一个**促进金融包容性**的机会。

4.3 帮助没有银行账户的人获得信贷

Access to credit for the unbanked

Table 12——划分不可评分客户和可评分客户

测试数字足迹是否可以为没有信用局评分的客户提供融资机会

不可评分客户的平均违约率为2.49%（Table12），

明显超过了可评分客户的违约率0.94%(Table2)

Table 12
Digital footprint variables and default rates (unscorable customers)

Variable	Value	Observations	Proportion (%)	Default rate (%)	*t*-test agair baseline
Device	All	15,580	100	2.49	
	Desktop	9,183	59	2.16	Baseline
	Tablet	2,618	17	1.64	(−1.64)
	Mobile	1,546	10	6.21***	(9.07)
	Do-not-track setting	2,233	14	2.28	(0.37)
Operating system	All	15,580	100	2.49	
	Windows	7,763	50	2.19	Baseline
	iOS	2,424	16	2.35	(0.47)
	Android	1,646	11	4.80***	(6.00)
	Macintosh	1,420	9	1.69	(−1.20)
	Other	94	1	7.45***	(3.42)
	Do-not-track setting	2,233	14	2.28	(0.27)
E-mail host	All	15,580	100	2.49	
	Gmx (partly paid)	3,681	24	2.42	Baseline
	Web (partly paid)	3,349	21	2.63	(0.56)
	T-Online (affluent customers)	1,709	11	1.52**	(−2.12)
	Gmail (free)	1,691	11	3.61**	(2.46)
	Yahoo (free, older service)	731	5	3.15	(1.14)
	Hotmail (free, older service)	546	4	2.75	(0.46)
	Other	3,873	25	2.22	(−0.57)
Channel	All	15,580	100	2.49	
	Paid	6,446	41	2.89	Baseline
	Direct	3,257	21	1.87***	(−2.99)
	Affiliate	1,394	9	2.65	(−0.47)
	Organic	1,178	8	2.55	(−0.64)
	Other	1,072	7	2.15	(−1.36)
	Do-not-track setting	2,233	14	2.28	(−1.50)

Credit bureau score, digital footprint variables, and default rates (scorable customers)

Variable	Value	Observations	Proportion (%)	Default rate (%)	*t*-test agains baseline
Credit bureau score (by quintile)	All	254,819	100	0.94	
	Q1 - lowest	50,980	20	2.12	Baseline
	Q2	50,949	20	1.02***	(−14.17)
	Q3	50,991	20	0.68***	(−19.51)
	Q4	51,181	20	0.47***	(−23.37)
	Q5 - highest	50,718	20	0.39***	(−24.89)
Device	All	254,819	100	0.94	
	Desktop	145,879	57	0.74	Baseline
	Tablet	45,575	18	0.91***	(3.62)
	Mobile	26,808	11	2.14***	(21.84)
	Do-not-track setting	36,557	14	0.88***	(2.90)
Operating system	All	254,819	100	0.94	
	Windows	124,605	49	0.74	Baseline
	iOS	41,478	16	1.07***	(6.35)
	Android	29,089	11	1.79***	(16.64)
	Macintosh	21,163	8	0.69	(−0.79)
	Other	1,927	1	1.09*	(1.74)
	Do-not-track setting	36,557	14	0.88***	(2.66)
E-mail host	All	254,819	100	0.94	
	Gmx (partly paid)	58,609	23	0.82	Baseline
	Web (partly paid)	54,867	22	0.86	(0.70)
	T-Online (affluent customers)	30,279	12	0.51***	(−5.32)
	Gmail (free)	27,845	11	1.25***	(6.02)
	Yahoo (free, older service)	11,923	5	1.96***	(11.33)
	Hotmail (free, older service)	10,241	4	1.45***	(6.11)
	Other	61,055	24	0.90	(1.38)
Channel	All	254,819	100	0.94	
	Paid	111,399	44	1.11	Baseline
	Direct	45,183	18	0.84***	(−4.78)
	Affiliate	24,770	10	0.64***	(−6.68)
	Organic	18,295	7	0.86***	(−3.00)
	Other	18,615	7	0.69***	(−5.24)

4.3 帮助没有银行账户的人获得信贷

Access to credit for the unbanked

Table 13——划分不可评分客户和可评分客户

测试数字足迹是否可以为没有信用局评分的客户提供融资机会

Table 13
Default regressions (unscorable customers)

Variables	(1) Digital footprint for unscorable customers Coef.	z-stat	(2) For comparison: Digital footprint for scorable customers (Column 2 of Table 4) Coef.	z-stat	(3) Digital footprint for unscorable customers, fixed effects Coef.	z-stat
Computer & operating system						
Desktop/Windows	Baseline		Baseline		Baseline	
Desktop/Macintosh	−0.26	(−1.10)	−0.07	(−0.53)	−0.26	(−1.06)
Tablet/Android	−0.22	(−0.86)	0.29***	(3.19)	−0.11	(−0.44)
Tablet/iOS	−0.45*	(−1.72)	0.08	(1.05)	−0.45*	(−1.67)
Mobile/Android	1.07***	(5.97)	1.05***	(17.25)	1.08***	(5.38)
Mobile/iOS	0.63***	(2.69)	0.72***	(9.07)	0.69***	(2.76)
E-mail host[a]						
Gmx	Baseline		Baseline		Baseline	
Web	0.02	(0.11)	0.00	(0.00)	0.01	(0.04)
T-Online	−0.39	(−1.14)	−0.40***	(−3.90)	−0.42	(−1.21)
Gmail	0.33	(1.36)	0.34***	(3.81)	0.31	(1.34)
Yahoo	0.17	(0.61)	0.75***	(9.19)	0.11	(0.36)
Hotmail	−0.02	(−0.06)	0.35***	(3.70)	−0.13	(−0.41)
Channel						
Paid	Baseline		Baseline		Baseline	
Affiliate	−0.08	(−0.39)	−0.49***	(−5.35)	−0.07	(−0.34)
Direct	−0.42**	(−2.34)	−0.27***	(−4.25)	−0.52***	(−2.66)
Organic	−0.05	(−0.24)	−0.15*	(−1.79)	0.03	(0.13)
Other	−0.27	(−1.21)	−0.47***	(−4.50)	−0.18	(−0.82)
Checkout time						
Evening (6 p.m.-midnight)	Baseline		Baseline		Baseline	
Morning (6 a.m.-noon)	0.30*	(1.81)	0.28***	(4.50)	0.32*	(1.88)
Afternoon (noon-6 p.m.)	0.39***	(2.70)	0.08	(1.42)	0.40***	(2.76)
Night (midnight-6 a.m.)	0.44	(1.38)	0.79***	(7.73)	0.45	(1.38)
Do-not-track setting	−0.16	(−0.83)	−0.02	(−0.25)	−0.23	(−1.18)
Name in e-mail	−0.59***	(−4.67)	−0.28***	(−5.67)	−0.54***	(−4.24)
Number in e-mail	0.63***	(4.31)	0.26***	(4.50)	0.61***	(4.07)
Is lowercase	0.95***	(5.45)	0.76***	(13.10)	0.91***	(4.71)
E-mail error	1.66***	(7.81)	1.66***	(20.00)	1.67***	(6.85)
Constant	−3.80***	(−19.20)	−4.92***	(−62.87)	−6.00***	(−11.32)
Control for *Gender*, *Item category*, *Loan amount*, and month and region fixed effects			No		No	Yes
Observations			15,580		254,819	15,580
Pseudo R^2			.0906		.0524	.1645
AUC			0.722		0.696	0.803
(SE)			(0.014)		(0.006)	(0.011)
Difference to AUC=50%			0.222***		0.196***	0.302***
AUC (OOS)			0.684		0.688	0.659

样本外AUC

列1: 使用数字足迹变量作为自变量的结果

列2：与可评分客户样本的结果的比较

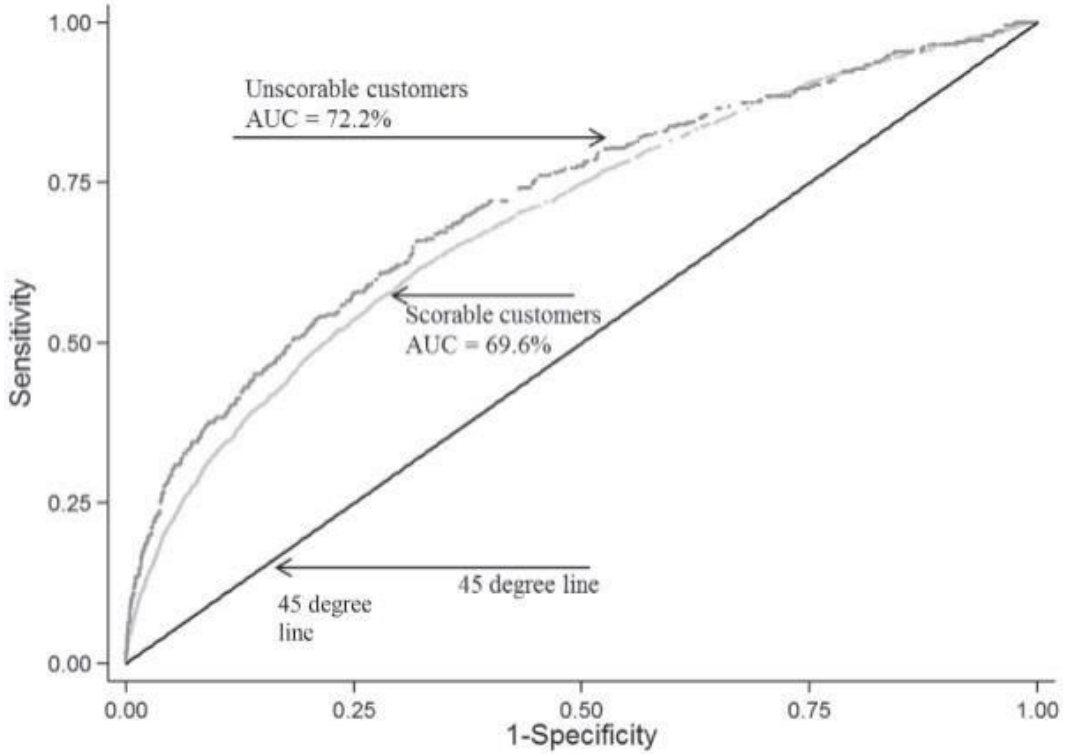
列3：增加了额外的控制（性别、贷款金额、项目类型）和月份和区域固定效应

可评分用户的AUC与不可评分客户的AUC相似(72.2%vs.69.6%)

4.3 帮助没有银行账户的人获得信贷

Access to credit for the unbanked

Figure 6——可评分客户和不可评分客户的AUC



可评分用户的AUC与不可评分客户的AUC相似

(72.2%vs.69.6%)

PART 5

总结

Conclusion

5.1 研究发现

Research Findings

- 1. 简单、易于访问的数字足迹变量，其信息含量**与信用局评分的信息含量相当。**
  - 对客户违约率的预测准确性相同
- 2. 数字足迹可以**补充而不是替代**信用局评分。
  - 两者联合时可提高客户违约率的预测准确性
- 3. 数字足迹既可以预测有信用局评分的客户的违约率，也**可以预测没有信用局评分的客户的违约率**，且效果一样好。
  - 可以帮助那些原本不能获得信贷服务的人获得信贷，从而促进金融包容

5.2 Lucas（1976）的批评

Advice

01

批评

如果数字足迹广泛用于贷款决策，客户可能会改变他们的在线行为。

解释

02

管理一个人的数字足迹会广泛地影响一个人的日常生活；同时，它与管理一个人的信用局评分有至关重要的不同，后者与谨慎的财务行为有关，而不是日常生活中的选择和习惯。

5.3 对监管的影响

Impact on Regulation

- 监管机构很可能会密切关注数字足迹的使用情况
- 使用数字足迹的贷款机构可能面临审查，信息的数字足迹代理是否违反公平贷款法案。
- 同时，现有的金融机构受到竞争对手使用数字足迹的威胁。

THANK YOU!