Data vs Collateral

Leonardo Gambacorta♦, Yiping Huang[†], Zhenhua Li[*], Han Qiu♦ and Shu Chen[*]

♦Bank for International Settlements, [†]Institute of Digital Finance & National School of Development, Peking University, [*]Ant Group.

Webinar Series - Capital Market Development: China and Asia, 16 September 2021

Disclaimer

The views in this paper are those of the authors only and do not necessarily reflect those of the Bank for International Settlements nor Ant Group. The authors highlight that the data and analysis reported in this paper may contain errors and are not suited for the purpose of company valuation or to deduce conclusions about the business success and/or commercial strategy of Ant Group or other firms. All statements made reflect the private opinions of the authors and do not express any official position of Ant Group and its management. The analysis was undertaken in strict observance of the Chinese law on privacy. Gambacorta, Huang and Qiu declare that they have no relevant or material financial interests that relate to the research described in this paper. Zhenhua Li and Shu Chen disclose having an employment relationship in Ant Group. Ant Group did not exercise any influence on the content of this paper, but has ensured confidentiality of the (raw) data.

Outline of the presentation

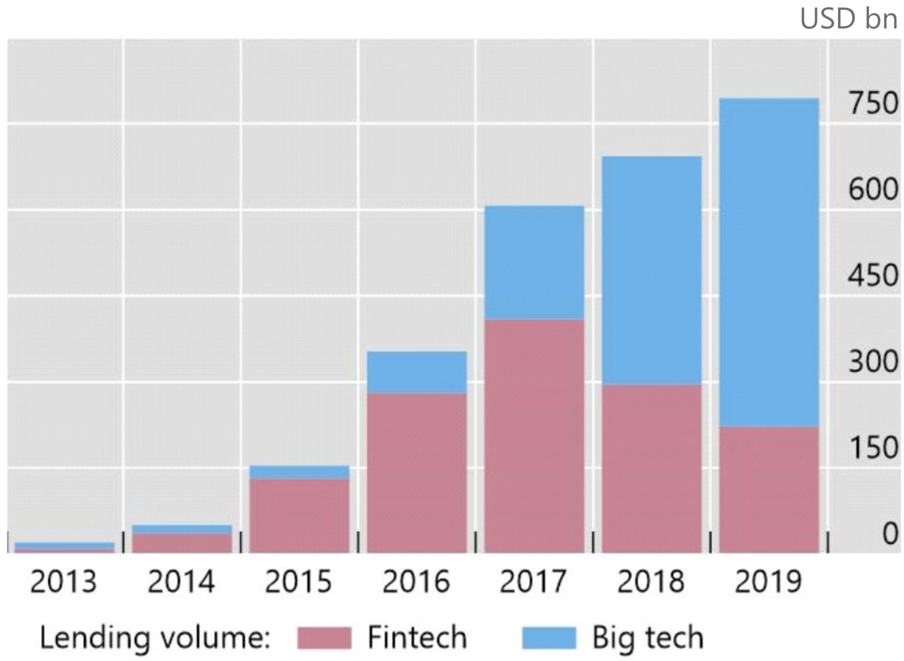
- Motivation and research questions
- Data and stylised facts
- Econometric strategy
- Results
- Conclusions

BIS 北京大学数字金融研究中心 Institute of Digital Finance Peking University

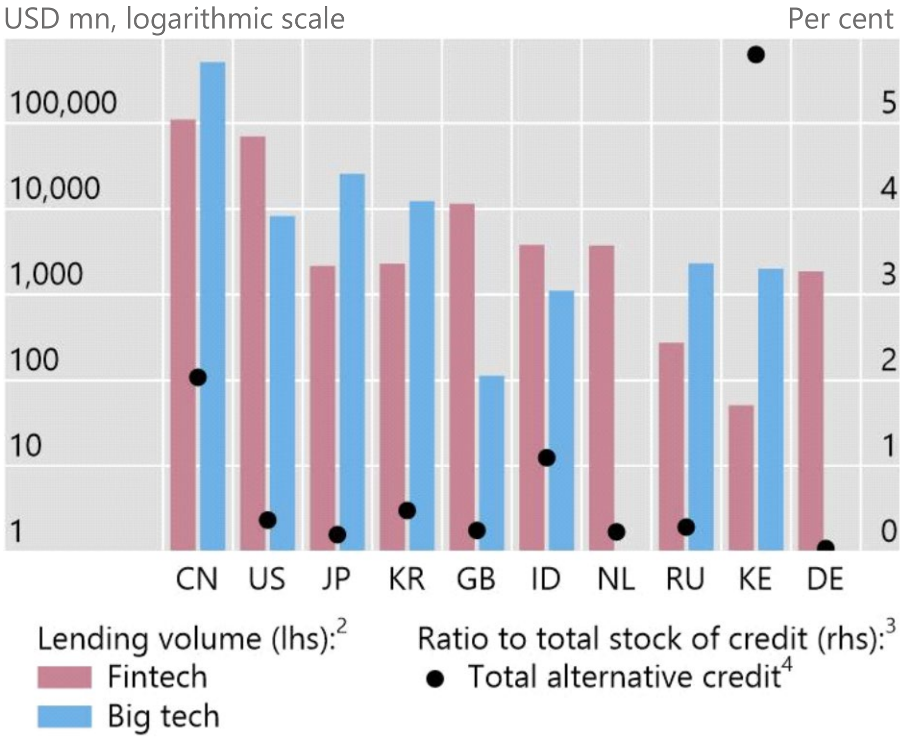
Motivation and research questions

Big tech credit is booming – reaching USD 572 bn in 2019

Big tech credit is overtaking fintech credit[1]



These alternative forms of lending are becoming a significant portion of total credit in a few economies

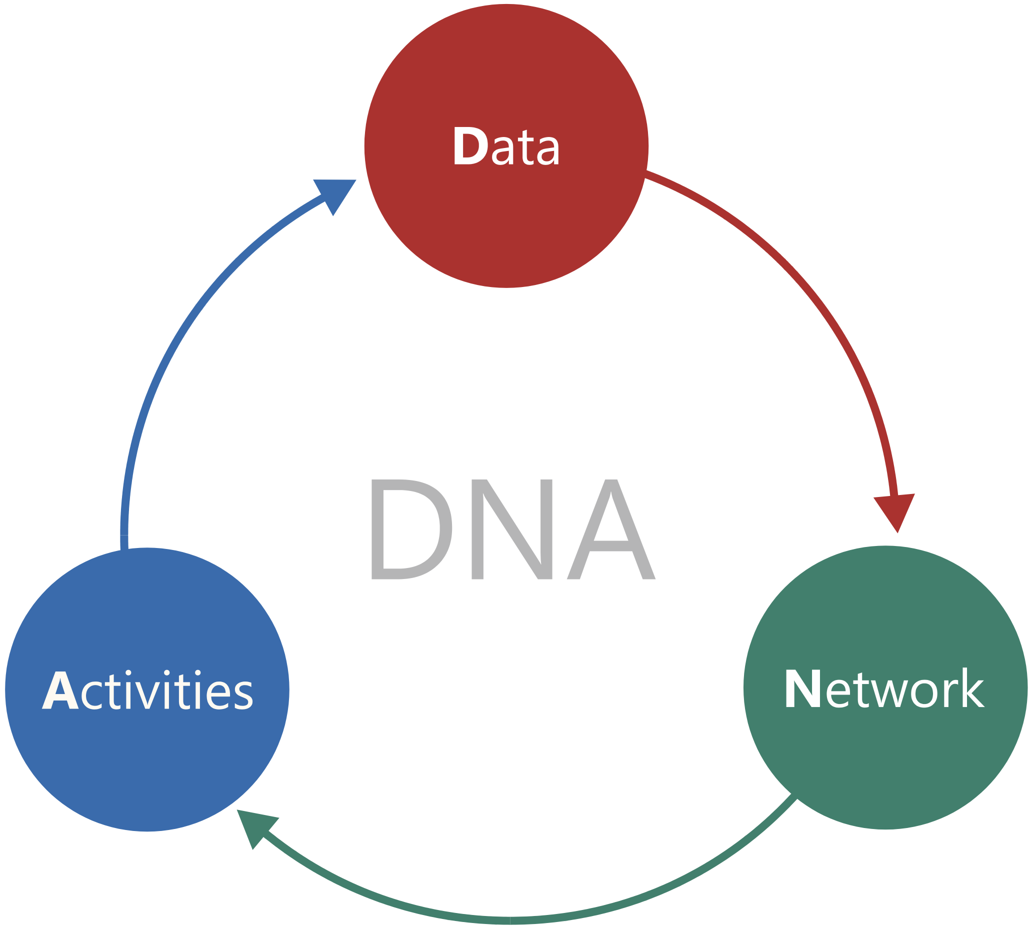


Figures include estimates. CN = China, US = United States, JP = Japan, KR = Korea, GB = Great Britain, ID = Indonesia, NL = Netherlands, RU = Russia, KE = Kenya, DE = Germany.

[1] 2019 fintech lending volume figures are estimated on AU, CN, EU, GB, NZ and US. [2] Data for 2019. [3] Domestic credit provided by the financial sector. Data for 2018. [4] Total alternative credit is defined as the sum of fintech and big tech credit. Data for 2019.

Source: Cornelli et al (2020), "Fintech and big tech credit: a new database", BIS WP 887.

Data-Network-Activities loop



Data can replace collateral

Big techs could address AI problems differently from banks (Hau et al, 2018).

They can use machine learning and big data to infer the credit quality of a borrower more precisely in real time (Berg et al 2019; Bazarbash, 2019; Frost at al 2019; Huang et al 2020).

Collateral is used in debt contracts to mitigate agency problems arising from asymmetric information.

Banks usually require their borrowers to pledge tangible assets, such as real estate, to:

- lessen ex-ante adverse selection problems (Bester 1985, Chan and Kanatas, 1985; Besanko and Thakor, 1987)
- as a way to reduce ex-post frictions, such as: i) moral hazard; ii) costly state verification; iii) imperfect contract enforcement

Research questions

Do big tech and bank credit react differently to collateral value, local economic conditions and firm-specific characteristics?

How could the increased use of big data and machine learning in solving asymmetric information problems, in lieu of collateral, impact the collateral channel?

Do big tech platforms matter? Are there differences between credit granted to firms that operate in the ecommerce platform (online) and credit granted to firms that operate on traditional business channels (offline)?

北京大学数字金融研究中心
Institute of Digital Finance Peking University

Data and stylised facts

Data

Unique dataset on credit from MYBank (Ant Group) and Chinese banks.

Random sample of more than 2 million Chinese firms in 2017:01-2019:04

Most of the firms have access only to big tech credit, however, 47,000 have access to secured bank credit and 120,000 to unsecured bank credit

On-line firms and off-line firms

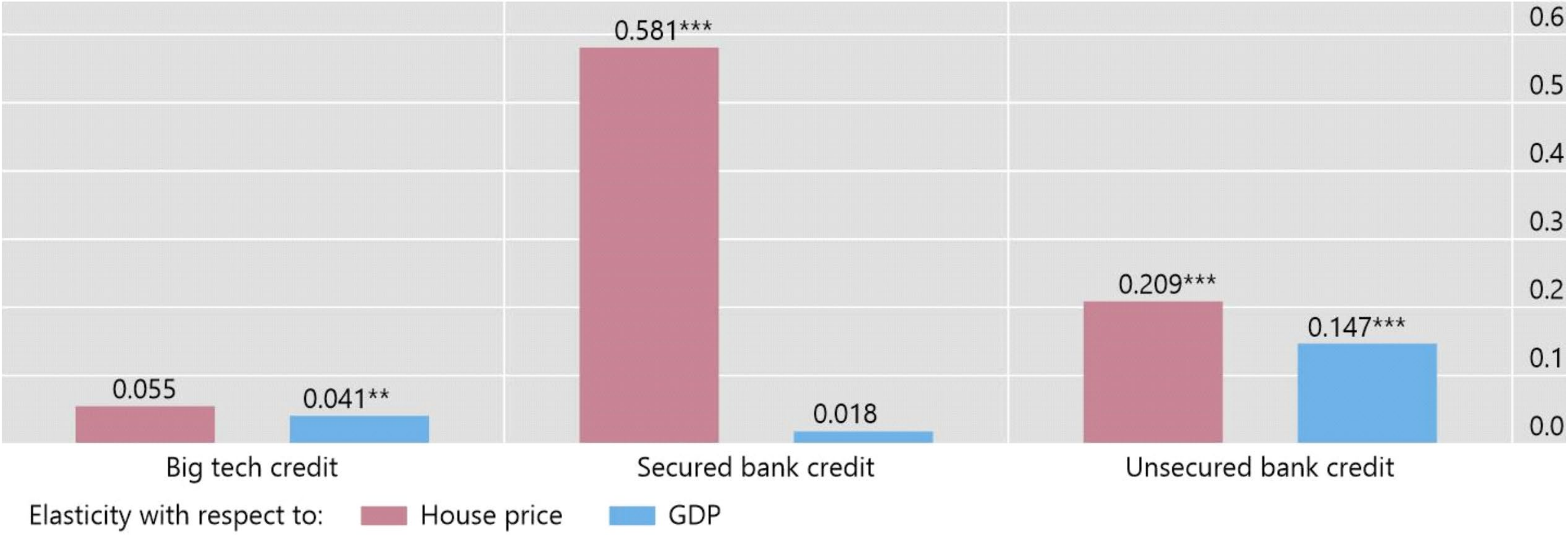
Firm level information at the monthly frequency:

- transaction volumes and network score
- personal characteristics such as age, information on car and house property, and total amount of funds into Alipay wallet (proxy for income)

House price at the city-month level

GDP at the city-quarter level

Elasticity of credit with respect to house prices and GDP



The figure reports the coefficient of three different regressions (one for each credit types) in which the log of credit is regressed with respect to the log of house prices at the city level, the log of GDP at the city level and a complete set of time dummies. Significance level: ** $p<0.05$; *** $p<0.01$.

BIS 北京大学数字金融研究中心 Institute of Digital Finance Peking University

Econometric strategy

Baseline model

$$ln\left(credit_{i,j,t}\right) = \mathrm{A}'X_{i,j,t} + \Gamma'Y_{j,t} + \mu_j + \mu_T + \varepsilon_{i,t}$$

$ln\left(credit_{i,j,t}\right)$ is the logarithm of the credit granted by MYbank or traditional banks (secured and unsecured) to firm *i*, headquartered in city *j*, in time *t*

$X_{i,j,t}$ is a vector that contains time-variant firm characteristics (transaction volume, network score) and owners' time-invariant characteristics (age, income)

$Y_{j,t}$ are the city-level indicators to capture regional conditions, including log of house price and local GDP

Time (μ_T) and city (μ_j) fixed effects

Standard errors are clustered at the city-month level

Model with firm fixed effects

$$ln(credit_{i,j,t}) = \mathrm{A}'X_{i,j,t} + \Gamma'Y_{j,t} + \mu_i + \mu_T + \varepsilon_{i,t}$$

$ln(credit_{i,j,t})$ is the logarithm of the credit granted by MYbank or traditional banks (secured and unsecured) to firm i, headquartered in city j, in time t

$Y_{j,t}$ are the city-level indicators to capture regional conditions, including log of house price and local GDP

Firm fixed effects (μ_i)

Time (μ_T) and city (μ_j) fixed effects

Standard errors are clustered at the city-month level

Nested model for firms with both forms of credit

$$ln(credit_{i,j,t}) = A'X_{i,j,t} + \Gamma'Y_{j,t} + B'X_{i,j,t} * credit_{type} + \text{K}'Y_{j,t} * credit_{type} + \mu_{iC} + \mu_{TC} + \varepsilon_{i,t}$$

$ln(credit_{i,j,t})$ is the logarithm of the credit granted by MYbank or traditional banks (secured and unsecured) to firm i, headquartered in city j, in time t

$Y_{j,t}$ are the city-level indicators to capture regional conditions, including log of house price and local GDP

$credit\_type$ dummy: takes the value of one for bank secured (or bank unsecured) credit and 0 for big tech credit

Time*credit type fixed effects (μ_{TC})

Borrower*credit type fixed effects (μ_{iC})

BIS 北京大学数字金融研究中心 Institute of Digital Finance Peking University

Results

Main drivers of credit

Explanatory variables	Dependent variable: Log (MYbank credit) (I)	Dependent variable: Log (Secured bank credit) (II)	Dependent variable: Log (Unsecured bank credit) (III)
Log House Price	0.052	0.371**	0.239*
	(0.047)	(0.151)	(0.138)
Log GDP	-0.028	0.053	0.229*
	(0.034)	(0.101)	(0.121)
Log Transaction Volume	0.064***	0.036***	0.064***
	(0.001)	(0.002)	(0.001)
Log Network Score	0.465***	0.096***	0.128***
	(0.004)	(0.011)	(0.008)
Age of the borrower	0.026***	0.009***	0.037***
	(0.0003)	(0.001)	(0.001)
Middle income	0.569***	0.129***	0.075***
	(0.003)	(0.012)	(0.007)
High income	1.281***	0.325***	0.312***
	(0.005)	(0.014)	(0.008)
Time and City FE	Yes	Yes	Yes
Number of observations	6,299,630	91,767	365,344
Adjusted R-squared	0.242	0.098	0.115

Notes: Standard errors in brackets are clustered at the city-month level. Significance level: *p<0.1; ** p<0.05; *** p<0.01.

Drivers of big tech credit: borrower specific fixed effects

Explanatory variables	Dependent variable: Log (MYbank credit used)			Dependent variable: Log (MYbank credit line granted)		
	All	Offline	Online	All	Offline	Online
	(I)	(II)	(III)	(IV)	(V)	(VI)
Log House Price	0.002	0.046	0.019	-0.041	-0.089	0.154
	(0.038)	(0.047)	(0.042)	(0.065)	(0.066)	(0.096)
Log GDP	0.046**	0.064***	-0.003	0.0004	-0.009	0.018
	(0.019)	(0.022)	(0.027)	(0.017)	(0.017)	(0.030)
Log Transaction Volume	0.046***	0.021***	0.160***	0.0005**	0.0003*	0.001**
	(0.001)	(0.0004))	(0.002)	(0.0002)	(0.0002)	(0.001)
Log Network Score	0.567***	0.092***	0.818***	0.205***	0.193***	0.179***
	(0.012)	(0.009)	(0.010)	(0.010)	(0.010)	(0.024)
Time FE (month)	Yes	Yes	Yes	Yes	Yes	Yes
Borrower FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	7,096,381	4,885,482	2,210,899	1,354,461	1,254,503	163,025
Adjusted R-squared	0.635	0.642	0.619	0.918	0.917	0.930

Notes: Standard errors reported in brackets are clustered at the city-month level. Significance level: *p<0.1; ** p<0.05; *** p<0.01.

Drivers of bank credit

Explanatory variables	Dependent variable: Log (Secured bank credit)			Dependent variable: Log (Unsecured bank credit)		
	All	Offline	Online	All	Offline	Online
Log House Price (1)	0.591***	0.480***	1.129***	0.212***	0.249***	0.036
	(0.145)	(0.148)	(0.319)	(0.081)	(0.084)	(0.169)
Log GDP (2)	-0.002	-0.024	0.117	0.144**	0.118*	0.292**
	(0.107)	(0.112)	(0.286)	(0.057)	(0.062)	(0.119)
Log Transaction Volume	0.003	0.001	0.013**	0.004***	0.004***	0.004
	(0.002)	(0.003)	(0.006)	(0.001)	(0.001)	(0.003)
Log Network Score (3)	-0.039	-0.051	-0.001	0.016	0.022	-0.029
	(0.029)	(0.032)	(0.082)	(0.012)	(0.014)	(0.031)
Time FE (month)	Yes	Yes	Yes	Yes	Yes	Yes
Borrower FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	94,948	81,799	13,149	398,789	369,287	67,103
Adjusted R-squared	0.579	0.578	0.583	0.651	0.651	0.650

Notes: (1) At the city-monthly level. (2) At the city-quarterly level. (3) Network score measures users' centrality in the network and is based on users' payment and funds information and social interactions. The user who has more connections gets a higher network score. Standard errors reported in brackets are clustered at the city-month level. Significance level: *p<0.1; ** p<0.05; *** p<0.01.

Potential endogeneity problems

Large firms may have a non-negligible impact through the demand for local labour and locally produced goods on housing price. (Chaney et al., 2012)

The characters of BigTech credit users and traditional bank credit users are different. (Tang, 2019)

Expansion in credit may also have effects on house prices. (Favara and Imbs, 2015)

It could be that our measure of housing prices proxies for local demand shocks that are not fully captured by local GDP conditions. (Jimenez et al., 2014)

Big tech credit vs bank secured credit: nested model

Explanatory variables	Dependent variable: Log (credit)		
	All	Offline	Online
Log House Price (1)	-0.061	-0.297	0.092
	(0.149)	(0.185)	(0.230)
Log GDP (2)	0.074	0.007	0.178
	(0.092)	(0.110)	(0.173)
Log Transaction Volume	0.037***	0.015***	0.145***
	(0.002)	(0.002)	(0.007)
Log Network Score (3)	0.456***	0.079*	0.830***
	(0.040)	(0.048)	(0.063)
Log House Price * Bank secured (4)	0.633***	0.689***	1.106***
	(0.217)	(0.238)	(0.395)
Log GDP* Bank secured (4)	-0.090	-0.024	-0.217
	(0.157)	(0.178)	(0.336)
Log Transaction Volume* Bank secured (4)	-0.035***	-0.016***	-0.128***
	(0.003)	(0.004)	(0.009)
Log Network Score * Bank secured (4)	-0.505***	-0.148**	-0.840***
	(0.052)	(0.061)	(0.099)
Time*credit type FE/ Borrower*credit type FE	Yes	Yes	Yes
Number of observations	168,518	128,163	40,355
Adjusted R-squared	0.737	0.741	0.732

Notes: Standard errors in brackets are clustered at the city-month level. Significance level: *p<0.1; ** p<0.05; *** p<0.01.

Big tech credit vs bank unsecured credit: nested model

Explanatory variables	Dependent variable: Log (credit) All	Offline	Online
Log House Price	-0.038	-0.101	0.108
	(0.071)	(0.076)	(0.120)
Log GDP	0.114**	0.074	0.175*
	(0.050)	(0.051)	(0.096)
Log Transaction Volume	0.037***	0.014***	0.113***
	(0.001)	(0.001)	(0.003)
Log Network Score	0.389***	0.059***	0.776***
	(0.022)	(0.022)	(0.034)
Log House Price * Bank unsecured	0.188*	0.224*	-0.098
	(0.111)	(0.102)	(0.219)
Log GDP* Bank unsecured	0.043	0.028	0.136
	(0.075)	(0.075)	(0.156)
Log Transaction Volume* Bank unsecured	-0.034***	-0.011***	-0.110***
	(0.002)	(0.002)	(0.004)
Log Network Score * Bank unsecured	-0.371***	-0.047*	-0.782***
	(0.027)	(0.027)	(0.049)
Time*credit type FE / Borrower*credit type FE	Yes	Yes	Yes
Number of observations	699,755	525,118	174,657
Adjusted R-squared	0.679	0.689	0.651

Notes: Standard errors in brackets are clustered at the city-month level. Significance level: *p<0.1; ** p<0.05; *** p<0.01.

Results using Instrumented values of the housing price (1)

First stage regression

Explanatory variables	Dependent variable: Log House Price
Lagged Land Supply (1)	-0.0370***
	(0.0131)
Lagged Land Supply (1) * mortgage rate (2)	0.00667**
	(0.00269)
Time FE	Yes
City FE	Yes
Number of observations	2,688
Adjusted R-squared	0.9952

Notes: (1) Lagged land supply are calculated using annual land supply scaled by urban construction land lagged by 12 months. (2) Nationwide interest rate at which banks refinance their home loans at the quarterly level. Standard errors in brackets are clustered at the city level. Significance level: *p<0.1; ** p<0.05; *** p<0.01.

Results using Instrumented values of the housing price (2)

Second stage regression

Explanatory variables	Dependent variable: Log (MYbank credit used)	Dependent variable: Log (Secured bank credit)	Dependent variable: Log (Unsecured bank credit)
Log House Price IV (1)	0.074	0.605**	0.393*
	(0.074)	(0.292)	(0.211)
Log GDP (2)	0.043**	0.008	0.144***
	(0.017)	(0.087)	(0.053)
Log Transaction Volume	0.039***	0.004**	0.004***
	(0.001)	(0.002)	(0.001)
Log Network Score (3)	0.462***	-0.037	0.015
	(0.011)	(0.024)	(0.014)
Time FE	Yes	Yes	Yes
Borrower FE	Yes	Yes	Yes
Number of observations	7,096,381	94,948	398,789
Adjusted R-squared	0.635	0.583	0.641

Notes: (1) Log House Prices are instrumented using the model described in Table 6. (2) At the city-quarterly level. (3) Network score measures users' centrality in the network and is based on users' payment and funds information and social interactions. The user who has more connections gets a higher network score. (4) Dummy variable that takes the value of one for bank unsecured credit and 0 for big tech credit. Standard errors in brackets are clustered at the city-month level. Significance level: *p<0.1; ** p<0.05; *** p<0.01.

Results controlling for demand shifts

Explanatory variables	Dependent variable: Log (credit)			
	Big tech credit vs Bank secured credit	Big tech credit vs Bank unsecured credit	Big tech credit vs Bank secured credit	Big tech credit vs Bank unsecured credit
Log Transaction Volume	0.041***	0.037***		
	(0.002)	(0.001)		
Log Network Score (3)	0.447***	0.396***		
	(0.040)	(0.022)		
Log House Price (1) * Bank credit (4)	0.583***	0.161	1.551**	0.892***
	(0.217)	(0.119)	(0.602)	(0.292)
Log GDP (2) * Bank credit (4)	−0.100	0.064	0.102	0.017
	(0.167)	(0.080)	(0.375)	(0.203)
Log Transaction Volume* Bank credit (4)	-0.039***	-0.034***	-0.057***	-0.008**
	(0.004)	(0.002)	(0.006)	(0.003)
Log Network Score (3) * Bank credit (4)	-0.484***	-0.369**	-0.443***	-0.210***
	(0.052)	(0.027)	(0.043)	(0.022)
Time*City FE	Yes	Yes	No	No
Time*Borrower FE	No	No	Yes	Yes
Borrower*credit type FE	Yes	Yes	No	No
City*credit type FE	No	No	Yes	Yes
Time*credit type FE	Yes	Yes	Yes	Yes
Number of observations	168,518	699,775	48,574	221,082
Adjusted R-squared	0.737	0.680	0.577	0.488

北京大学数字金融研究中心
Institute of Digital Finance Peking University

Conclusions

Main takeaways

Big tech credit does not correlate with local business conditions and house prices, but reacts strongly to firm characteristics.

An increased use of big tech credit could weaken the collateral channel.

Big tech credit to online firms, fully integrated in the e-commerce platform, is more strongly correlated with transaction volumes and network scores than it is in the case of offline firms. Big tech credit to offline firms shows some sign of correlation with local demand conditions.