BANK FOR INTERNATIONAL SETTLEMENTS

BIS Working Papers

No 970

Big techs in finance: on the new nexus between data privacy and competition

by Frederic Boissay, Torsten Ehlers,
Leonardo Gambacorta and Hyun Song Shin

Monetary and Economic Department

October 2021

JEL classification: E51, G23, O31.

Keywords: digital platforms, big techs, finance, data privacy, competition.

ISSN 1020-0959 (print)
ISSN 1682-7678 (online)

Big techs in finance: on the new nexus between data privacy and competition

Frederic Boissay, Torsten Ehlers, Leonardo Gambacorta and Hyun Song Shin*

Abstract

The business model of big techs rests on enabling direct interactions among a large number of users on digital platforms, such as in e-commerce, search and social media. An essential by-product is their large stock of user data, which they use to offer a wide range of services and exploit natural network effects, generating further user activity. Increased user activity completes the circle, as it generates yet more data. Building on the self-reinforcing nature of the data-network-activities loop, some big techs have ventured into financial services, including payments, money management, insurance and lending. The entry of big techs into finance promises efficiency gains and greater financial inclusion. At the same time, it introduces new risks associated with market power and data privacy. The nature of the new trade-off between efficiency and privacy will depend on societal preferences, and will vary across jurisdictions. This increases the need to coordinate policies both at the domestic and international level.

Keywords: digital platforms, big techs, finance, data privacy, competition.

JEL Codes: E51, G23, O31.

* All authors are from the Bank for International Settlements (BIS). This paper has been published as Chapter 31 in The Palgrave Handbook of Technological Finance, edited by Raghavendra Rau, Robert Wardrop and Luigi Zingales. We would like to thank Stijn Claessens, Sebastian Doerr and Jon Frost for comments and suggestions. We thank Giulio Cornelli for excellent research assistance. The views in this paper are those of the authors only and do not necessarily reflect those of the Bank for International Settlements.

Introduction

Large technology companies such as Alibaba, Amazon, Facebook, Google and Tencent have begun to enter financial services. Their entry builds on their established digital platforms in e-commerce, search and social media, and holds the prospect of efficiency gains and greater financial inclusion (BIS, 2019). The business model of these "big techs" rests on enabling direct interactions among a large number of users. An essential by-product of their business is their large stock of user data, which are used as an input for a range of services that exploit natural network effects, generating further user activity. Increased user activity then completes the circle, as it generates yet more data. The self-reinforcing loop between **D**ata, **N**etwork externalities and **A**ctivities, is the DNA of big techs.

Personal data lie at the heart of this new, digital economy. Building on the advantages of the self-reinforcing nature of the data-network-activities loop, some big techs have ventured into financial services, including payments, money management, insurance and lending. As yet, financial services are only a small part of big techs' business globally. But given their size and customer reach, big techs' entry into finance has the potential to spark rapid change in the financial industry. The entry of big techs promises many potential benefits. With their low-cost structures, big techs can easily scale up their businesses to provide basic financial services, especially where a large part of the population remains unbanked. Using big data and analysis of the network structure of user connections in their platforms, big techs can better assess the riskiness of borrowers, reducing the need for collateral to assure repayment. As such, big techs promise to enhance the efficiency of financial services provision, promote financial inclusion and allow associated gains in economic activity.

The benefits notwithstanding, widespread use of data in combination with the new technologies and applications can have adverse side effects. Big techs have the potential to become dominant through the advantages afforded by the data-network-activities loop, raising competition and data privacy issues. Complex questions arise about how best to organise access to personal data, not only to limit side effects, but also to protect people's privacy to the degrees desired. How to define and regulate the use of data has become an important policy issue for authorities at the domestic and international level.

This chapter starts by explaining big techs' business model, their life cycle and the characteristics of the data-network externalities-activities feedback loop. A second section analyses how big techs' DNA could produce potential benefits in the provision of financial services, both increasing financial inclusion and reducing asymmetric information problems in the supply of credit. A third section evaluates the potential costs caused by the entry of big techs into finance, including the new risks of price discrimination, abuse of market power, anti-competitive behaviour and limits to data privacy. A last section lays out the complex public policy trade-off between the objectives of efficiency and privacy, and discusses the policy options, as well as the case for policy coordination at the domestic and international level.

Big techs’ business model

Big techs’ business models can be best described as online platforms that allow different types of user (eg buyers and sellers) to interact. This creates network externalities: the more users interact, the more attractive the platform. Other industries (eg telecommunication networks, credit card payments networks, etc) feature network externalities. But big techs’ online-focused business models allow them to reach dominant market positions at unprecedented speed. Further, the systematic accumulation of user data and new ways of analysing it (eg artificial intelligence such as machine learning solutions) allow them to exploit these network externalities in a very effective way.

Big techs as multi-sided platforms and their life cycle

Big techs initially create value as online “multi-sided platforms” (MSPs), by enabling and catalysing direct interactions between two or more groups of users (eg buyers and sellers). The three main types of online platform are social networks, e-commerce platforms and search engines.

In contrast to traditional bilateral exchanges, users on each side transact with each other through the platform – not necessarily with the platform itself. Social platforms, for example, allow people to connect to each other, so that each member benefits from a larger community. E-commerce websites enable their users to buy and sell a wide variety of goods and services worldwide. The larger number of sellers reduces buyers’ search costs, and a larger number of buyers expands sellers’ business opportunities. A typical feature of MSPs is the presence of network externalities: the very fact that users participate on one side of the platform (eg buyers) increases users’ benefits on the other side (eg sellers). One challenge is to attract users onto both sides at the same time – a “chicken and egg” problem. Successful platforms solve this problem by using specific price structures, which essentially consist in charging a lower fee to the side that creates the most network externalities – and letting the side that benefits the most from the network subsidise the other (see, eg, Jullien (2004)).

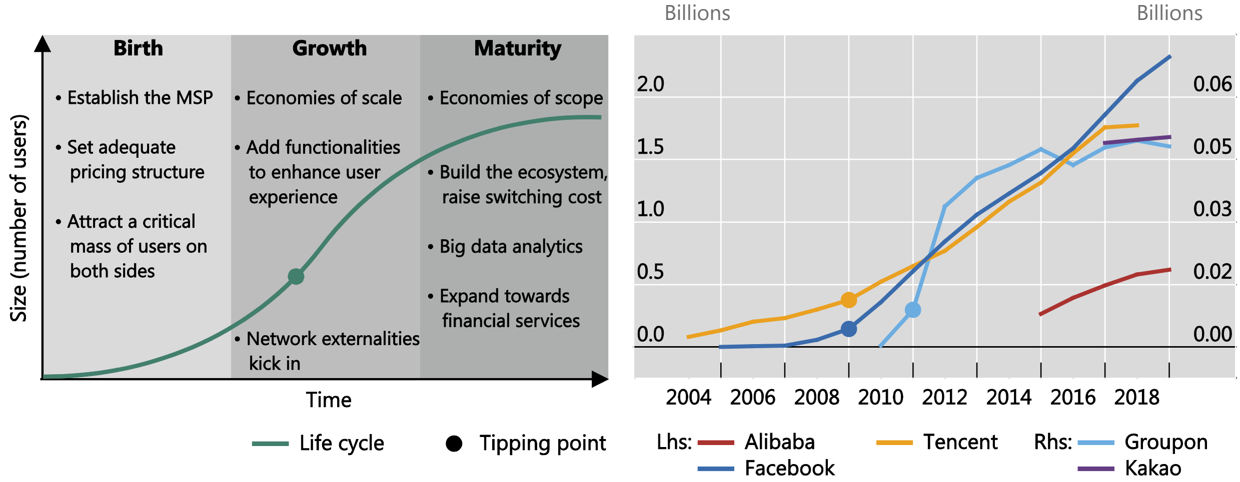
Big techs have so far followed a rather traditional corporate life cycle with three phases: birth, growth and maturity (Graph 1, left-hand panel). What makes them unique is the coincidence of several factors (ie the collection of personal data on a large scale, network effects and a large number of activities) and the rapidity with which they reach maturity. Petralia et al. (2019), for example, report that social networks such as Facebook or Tencent’s WeChat took less than five years to reach 50 million users (see also Graph 1, right-hand panel). In terms of user numbers, these firms are much larger, and have grown much faster, than any financial firm. In particular, Nguyen Trieu (2017) argues that big techs have scaled up between 10 and 100 times faster than traditional financial institutions.

Big techs' life cycle: theory and practice

Graph 1

From big techs' birth to maturity[1]

Big techs' monthly active users



[1] The firm's life cycle described in the left-hand panel borrows from the synthesis of the literature by Miller and Friesen (1984). Given that big techs are still new and rising firms, we purposely ignore the usual "decline" phase. MSP = multi-sided platform.

Sources: BIS (2019); Miller and Friesen (1984); S&P Capital IQ; Authors' calculations.

Even after an MSP has attracted a sufficient mass of users on both sides, the emphasis remains on further increasing the number of users, with the aim of reaching a tipping point where adoption rates accelerate and network effects kick in. Beyond this point, growth can be very fast. More buyers bring more sellers – and vice versa – so that the MSP enjoys increasing returns to scale. The average cost of serving a user falls as the total number of users rises. And users are willing to pay more for access to a bigger network. As a result, the platform's margins improve.

The DNA of big techs

Data analytics, **n**etwork externalities and interwoven **a**ctivities ("DNA") comprise the key features of big techs' business models. These three elements reinforce each other. Network externalities generate additional users and added value for users. They allow the big tech to generate more data, the key input into data analytics. The analysis of large troves of data enhances existing services and attracts further users. More users, in turn, provide the critical mass of customers, so that a wider range of activities can be offered, which yield even more data. Accordingly, network externalities are stronger on platforms that offer a broader range of services, thus representing an essential element of big techs' life cycle.

Financial services both benefit from and fuel the DNA feedback loop. Offering financial services can complement and reinforce big techs' commercial activities. The typical example is payment services, which facilitate secure transactions on e-commerce platforms, or make it possible to send money to other users on social media platforms. Payment transactions also generate data detailing the network of links between fund senders and recipients. These data can be used both to enhance existing (eg targeted advertising) and other financial services, such as credit scoring.

The source and type of data and the related DNA synergies vary across big tech platforms. Those with a dominant presence in e-commerce collect data from vendors, such as sales and profits, combining financial and consumer habit information. Big techs with a focus on social media collect data on individuals and their preferences, as well as their network of connections. Big techs with search engines do not observe connections directly, but typically have a broad base of users and can infer their preferences from their online searches.

The type of synergy varies with the nature of the data collected. Data from e-commerce platforms can be a valuable input into credit scoring models, especially for SME and consumer loans. Big techs with a large user base in social media or internet search can use the information on users' preferences to market, distribute and price third-party financial services (eg insurance).

Potential benefits from big techs

Building on the advantages of the reinforcing nature of the data-network-activities loop, some big techs have ventured into financial services, including payments, money management, insurance and lending. As yet, financial services are only a small part of their business globally (around 11% of their total revenues). But given their size and customer reach, big techs' entry into finance has the potential to spark rapid change in the industry. It offers many potential benefits. Thanks to their low-cost structures, the businesses of big techs can easily be scaled up to provide basic financial services, especially in places where a large part of the population remains unbanked.

Payments were the first financial service to be offered by big techs, mainly to help overcome the lack of trust between buyers and sellers on e-commerce platforms. Buyers want goods to be delivered, but sellers are only willing to deliver after being assured of payment. Payment services, such as those provided by Alipay (owned by Alibaba) or PayPal (owned by eBay), allow guaranteed settlement at delivery and/or reclaims by buyers and are fully integrated into e-commerce platforms. In some regions with less developed retail payment systems, new payment services have emerged through mobile network operators (eg M-Pesa in several African countries). Over time, big techs' payments services have become more widely used as an alternative to other electronic payment means such as credit and debit cards.

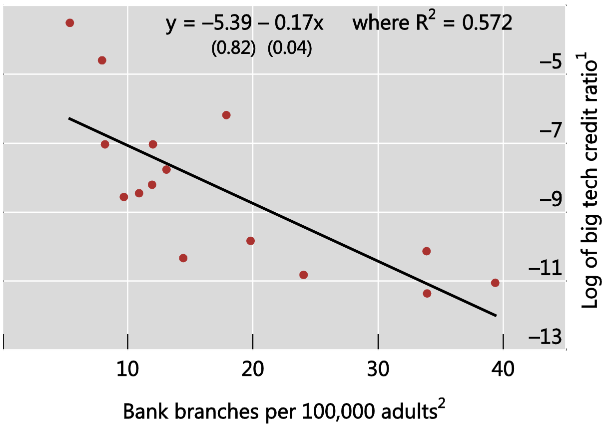
Benefits are also evident in lending provision. Using big data and by analysing the network structure of user connections in their platforms, big techs can assess the riskiness of borrowers, reducing the need for collateral to assure repayment. As such, big techs can promote financial inclusion.

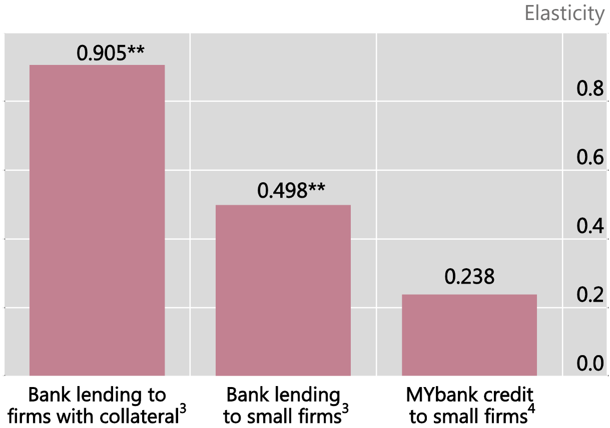
Financial inclusion

Financial inclusion allows individuals and businesses to have access to financial products and services in a responsible and sustainable way. In particular, the provision of credit or saving products to unbanked individuals allows them to use resources more optimally over time. Meanwhile, insurance products can serve as a cushion against shocks such as bad harvests, illness, or the death of the main wage earner.

Big tech credit, asset prices and bank development

Graph 2





Robust standard error in parentheses. ** indicates significance at the 5% level.

[1] The ratio is calculated for 2017 and is defined as big tech credit divided by total credit to the private non-financial sector (including total fintech credit). [2] Average over the period 2010–15. [3] Period of estimation: 2005–13. [4] Period of estimation: 2011–17.

Sources: BIS (2019); Frost et al (2019); Gambacorta et al (2019); World Bank; Cambridge Centre for Alternative Finance and research partners; company reports; Authors' calculations.

The greater use of data can foster greater convenience, more tailored and personalised products, and greater financial inclusion. Big techs may have a competitive advantage over banks and serve firms and households that would otherwise remain unbanked (Graph 2, left-hand panel).[1] This is because they can tap different but relevant information from their digital platforms.[2] For example, Ant Financial and Mercado Libre claim that their credit quality assessment and lending decisions typically involve more than 1,000 data series per loan applicant.

There is evidence that the advent of fintech and big tech lenders and their use of alternative data have been a boon for borrowers who are unserved or underserved by banks. In China, for example, the major platforms have provided access to credit for hundreds of millions of new personal and business borrowers. In many emerging markets, the use of data on transactions, payment of utility bills, or platform reviews is driving greater access to and personalisation of financial services. These benefits exist even in countries with advanced systems of financial services provision. For example, in the United States, better use of personal transaction data promises to help the 45–60 million "thin credit file" Americans, ie those who have inadequate credit history, to obtain loans. In other words, the efficient use of data allows big techs to overcome some of the traditional information problems typically encountered by traditional intermediaries with respect to more opaque borrowers.

[1] More generally, big techs' market penetration rate tends to be higher in areas where banks are absent or their branch networks sparser.

[2] See Hau et al (2018) and Huang et al (2018) for the case of China.

Reduction of financial frictions in lending

The sheer amount of data collected by big techs and their intelligent use have the potential to reduce financial frictions, in particular borrower screening, monitoring and collateral requirements. Potential borrowers who cannot be served by regular banks due to prohibitive administrative costs could potentially obtain credit on the basis of credit ratings or scores built on a broader set of data processed in novel ways.

Borrower information and screening

Information costs on credit markets may sometimes be so prohibitive that banks refrain from serving borrowers – or do so only at very high spreads. Big techs' processing of large quantities of information (big data) using advanced analytical methods such as machine learning and network analysis (artificial intelligence) can reduce such costs. Big data relevant for financial services obtained directly from big tech platforms include (i) transactions (sales volumes and average selling prices); (ii) reputation-related information (claim ratio, handling time, reviews and complaints); and (iii) industry-specific characteristics (sales seasonality, demand trend and macroeconomic sensitivity). This can be also enriched by using non-traditional data obtained via social media and other channels.

Frost et al (2019) suggest that, when applied to small vendors, big techs' credit scoring outperforms models based on credit bureau ratings and traditional borrower characteristics. The predictive power of the big techs' scoring systems arises in large part from exploiting the network structure. For instance, MYbank (Ant Financial group) uses network analysis of transactions to evaluate whether an entrepreneur separates personal funds from business funds, which is one of the basic principles of good business conduct.

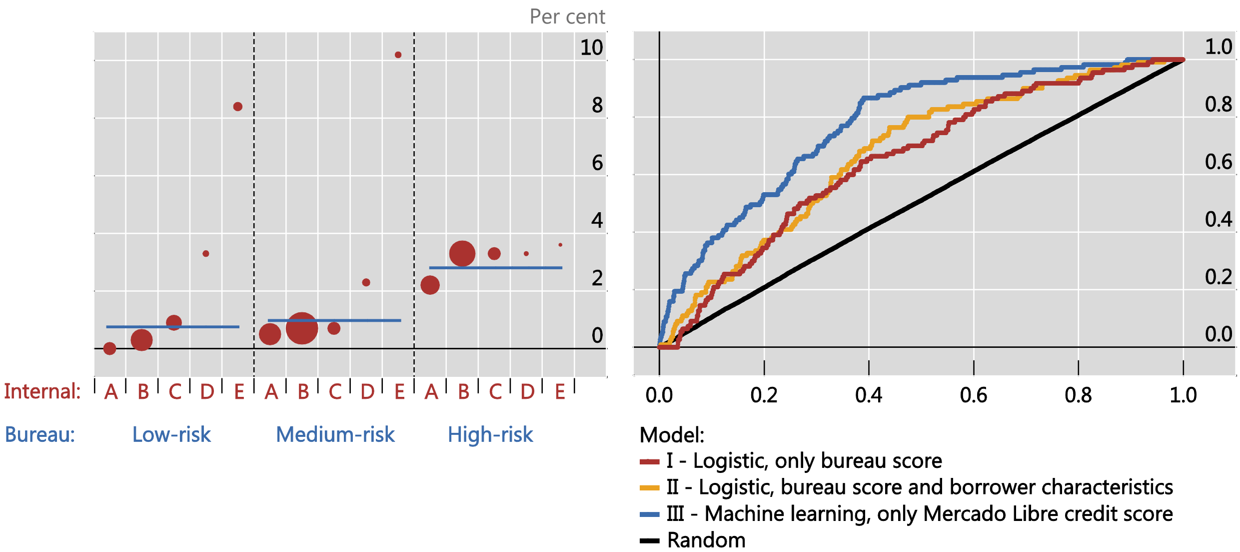
Frost et al (2019) also present empirical evidence that the use of more granular data with machine learning can help to improve the predictive power of prepayment prospects, especially for small merchants who are typically not served by banks. In the case of Mercado Libre, internal ratings are more granular (A to E) than those of the credit bureaus in Argentina (low-risk to high-risk). Banks rely on information from credit bureaus but augment it with other borrower characteristics and soft information (Graph 3, left-hand panel). However, as most of Mercado Libre's clients are unbanked, the analysis below is more specific to cases in which traditional soft information collected by banks is not available.

For a given bureau rating (eg low-risk), the expected loss rate is strictly monotonic with the internal rating (ie the patterns of the dots show that the internal rating increases with expected loss). Conversely, for a given internal rating (eg C, D or E), the loss rate is not strictly monotonic with the credit bureau risk. For example, the dot associated with internal rating D in the low-risk bureau category indicates a higher risk than the internal rating D in the medium-risk bureau category. Moreover, the internal rating has a broader range, covering losses from 0.0% to 10.2%; the bureau rating ranges from 0.7% to 2.8%.

Most importantly, by using the internal scoring model, Mercado Libre is able to provide credit to the profiles assessed as high-risk by the bureau. The size of the dots in the left-hand panel of Graph 3 is proportional to the share of the firms in the rating distribution; a substantial number of clients are in the credit bureau high-risk category. Because banks use a mix of credit bureau information, hard information from financial statements and soft information from loan officers, this segment may have much less access to traditional banking services. With its more granular scoring model, Mercado Libre offers 30% of its credit to this category.

Credit assessment and big data analytics | Graph 3

Loss rate[1] | Receiver operating characteristics (ROC) curve[2]



[1] The loss rate is the volume of loans more than 30 days past due relative to the origination volume. To date, Mercado Libre's internal rating system has proved better able to predict such losses. It segments loan originations into five different risk groups as compared with the three clusters identified by the bank bureau. The size of the dots is proportional to the share of the firms in the rating distribution. [2] True positive rates versus false positive rates for borrowers at different thresholds for a logistic model with only the credit bureau score (I), a logistic model with the bureau score and borrowers' characteristics (II), and a machine learning model with the Mercado Libre credit score (III). A random model is included for comparison purposes. The ROC curve shows that the machine learning model has superior predictive power to both the credit bureau score only and the credit bureau score with borrower characteristics.

Source: BIS (2019); Frost el al (2019).

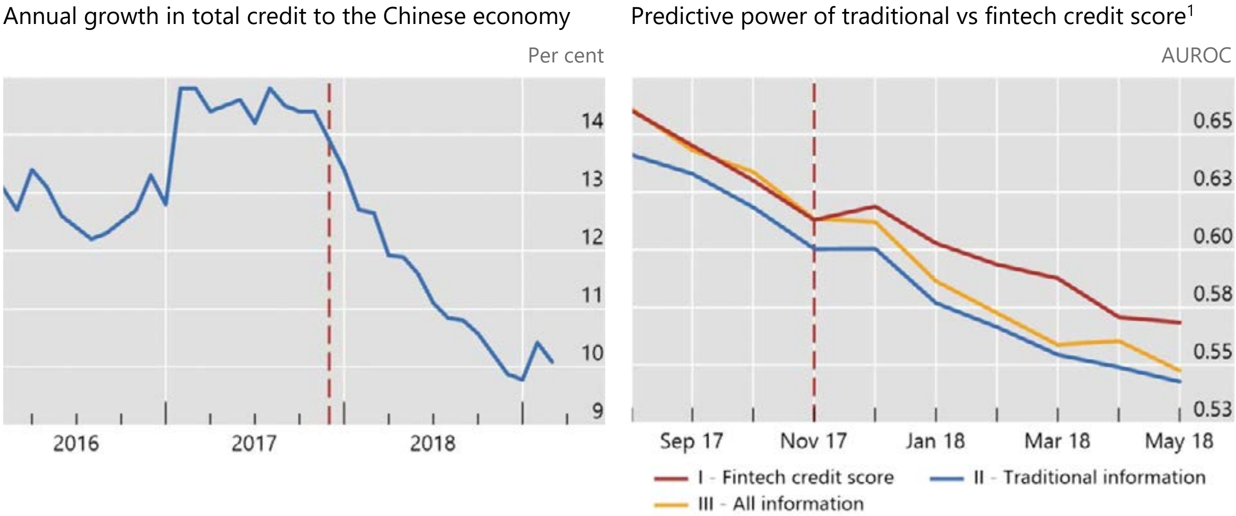
Further, the internal rating system based on machine learning techniques and data obtained from the e-commerce platform can outperform simple models based on bureau score and borrower characteristics in predicting defaults (Graph 3, right-hand panel). Performance is measured in this case on the y-axis directly by the area under the receiver operating characteristics (AUROC) curve. The higher the AUROC, the higher the predictive power of the model. The right-hand panel of Graph 3 reports the respective performance of the three models using this measure. The AUROC ranges from 50% (purely random prediction) to 100% (perfect prediction). From Graph 3 it is evident that the predictive power rises the most for the model that uses a machine learning technique applied to the data from the e-commerce platform.

While encouraging, these results prompt the questions (i) whether the performance of machine learning or fintech-based credit scoring models is superior to bank models that also use soft information, and (ii) if their performance can be sustained over full business and financial cycles.

To answer these questions, Gambacorta et al (2019) use a unique dataset on loan transactions from a leading Chinese fintech company to compare the predictive power of credit scoring obtained with machine learning and big data with more traditional linear models in the case of a shock. The predictive performance of the models in terms of credit losses and defaults is analysed both in "normal times" and following an (exogenous) change in regulatory policy on shadow banking in

November 2017 (dashed line in the two panels of Graph 4). The new rules led traditional and new financial institutions to tighten their lending requirements, causing a significant drop in the growth of total credit (the left-hand panel of Graph 4) and an increase in the number of defaults. As in Frost et al (2019), performance is measured directly by the area under AUROC curve (y-axis).

Fintech credit scores show greater predictive power after a change in regulation — Graph 4



The vertical dashed line indicates when the People's Bank of China (PBoC) issued specific draft guidelines to tighten regulations on shadow banking. In particular, from 17 November 2017, financial institutions have not been allowed to use asset management products to invest in commercial banks' credit assets or provide "funding services" for other institutions (such as fintech companies) to bypass regulations. The new rule has had a huge impact on fintech companies' funding sources. The PBoC set also a limit on the interest rates charged by P2P lending companies. All annualised interest rates, which include the upfront fees charged for loans, were capped at 36%. The effects of these new rules were also reinforced by the strict measures concerning online micro-lending that were imposed on December 1, 2017 by China's Internet Financial Risk Special Rectification Work Leadership Team Office.

[1] The vertical axis reports the Area Under the ROC curve (AUROC) for every model. The AUROC is a widely used metric for judging the discriminatory power of credit scores .The AUROC ranges from 50% (purely random prediction) to 100% (perfect prediction).

Source: Gambacorta et al (2019).

In normal times the fintech credit score outperforms the linear model with traditional information, but performs in line with the linear model that uses all information, ie both traditional and non-traditional data (right-hand panel of Graph 4). However, the fintech credit score outperforms the other models in the aftermath of the modification in credit conditions that followed the change in regulation. One possible interpretation of this finding is that credit scoring models based on machine learning better capture the non-linear relationship between variables in a period of stress.

Monitoring and collateral

The cost of enforcing loan repayments is an important component of total financial intermediation cost. To reduce enforcement problems banks usually require borrowers to pledge as collateral tangible assets, such as real estate, with the aim of increasing recovery rates in case of default. Another precaution is monitoring. Banks spend time and resources monitoring their clients' projects to limit the risk that borrowers implement them differently from what was initially agreed. Through the monitoring process, firms and financial intermediaries also develop long-term relationships and build mutual trust, which makes default a less attractive option for borrowers.

Big techs can address these issues differently. When a borrower is closely integrated into an e-commerce platform, for example, it may be relatively easy for a big tech to deduct the (monthly) payments on a credit line from any of the borrower's revenues that flow through its payment account. In contrast, banks may not be in the position to do so as the borrower may have accounts with other banks. Given network effects and high switching costs, big techs could also enforce loan repayments by the simple threat of a downgrade or an exclusion from their ecosystem if borrowers default on their payments. Anecdotal evidence from Argentina and China suggests that the combination of massive amounts of data and network effects may allow big techs to mitigate the information and incentive problems that are traditionally addressed through the posting of collateral. This could explain why, unlike the supply of corporate loans from banks, that of big techs does not closely correlate with asset prices (Graph 2, right-hand panel).

Potential costs of big techs' use of personal data

Big techs' entry into finance brings efficiency gains and lowers barriers to the provision of financial services. But the very features that bring benefits also have the potential to generate new risks and costs associated with market power. Once a captive ecosystem is established, platforms can exploit their market power and network externalities to increase user switching costs, exclude potential competitors, and consolidate their position by raising barriers to entry.

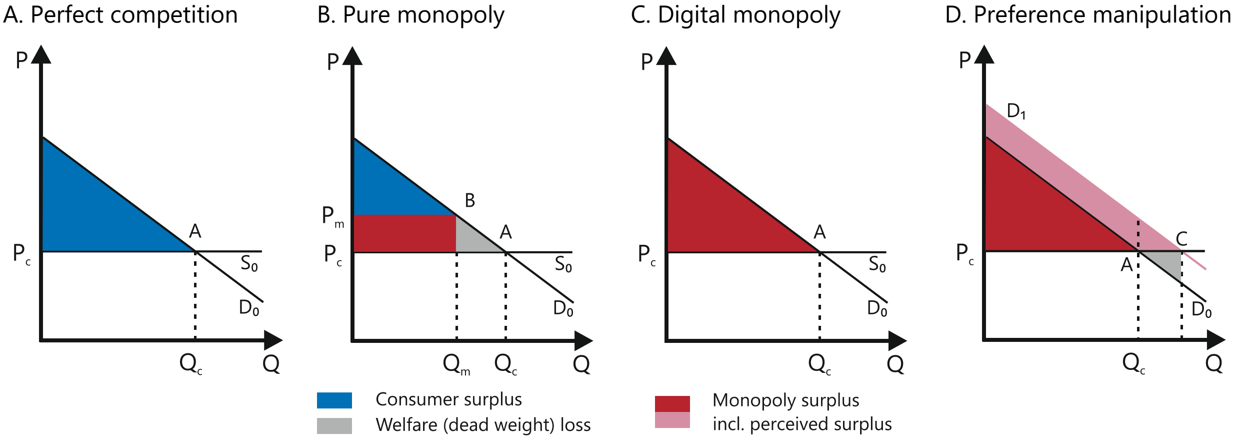
Monopolistic use of data and price discrimination

A major risk is the monopolistic use of data. One special aspect of data as an input of production is non-rivalry: data can be used many times over, and by any number of firms simultaneously, without being depleted. Thanks to non-rivalry, data generate increasing returns in both scale and scope (Farboodi et al. (2019)). Any single additional piece of data (eg a credit score) has additional value when it is combined with an existing large stock of data or across sources. For this reason, data are more valuable to big tech firms, giving rise to feedback loops and so-called digital monopolies.

Data can be used for price discrimination. Once their dominant position in data is established, digital monopolies may use the data not only to assess a potential borrower's creditworthiness or riskiness, but also to identify the highest rate a given borrower would be willing to pay for a loan or the highest premium a client would pay for insurance (ie, their individual "reservation price"). Fuelled by big data, algorithmic price discrimination is able to parse the population of potential customers into finer and finer subcategories – each matched with a different price. In some cases, sellers are even able to set personalized pricing, marching down the demand curve and setting a different price for each consumer.[3]

A significant source of profits for big techs comes from this extraction of the consumer's surplus. Bar-Gill (2019) illustrates this mechanism by comparing the equilibrium outcomes under perfect competition (Graph 5, panel A), pure monopoly (ie, without the use of big data; panel B), and digital monopoly (panel C). Under perfect competition, financial services are priced at their marginal cost (Pc), and the entire surplus accrues to consumers (blue area). The pure monopoly also sets one single price but, as this price is typically higher than the marginal cost, the supply of services is reduced. The overall surplus shrinks (by an amount that corresponds to the grey area), and the monopoly corners some of it (red area). In this case, consumers are worse off. Panel C presents the case of a *digital* monopoly. Using big data and sophisticated algorithms, a monopolistic big tech identifies each consumer's reservation price, and sets a personalized price just below it. Price discrimination allows the big tech to increase the quantity sold, back up to the competitive outcome (ie from point B to point A). The deadweight loss is eliminated, which increases the overall social welfare back to the perfect competition level. However, the entire surplus now goes to the big tech. In that case, efficiency is restored at the cost of the redistribution of the surplus in favour of the digital monopoly. But, ultimately, consumers are worse off than under a pure monopoly.

From market competition to market manipulation | Graph 5



Source: Adapted from Bar-Gill (2019).

Algorithmic biases and abuse of market power

The algorithms used to process data may also develop biases, leading to unethical discrimination (eg based on race, religion, etc) and greater inequality (O'Neil (2016)). For instance, one recent study of the US mortgage market found that black and Hispanic borrowers were less likely to benefit from lower interest rates from machine

3 There is empirical evidence of price discrimination based on information collected online about consumers. The price difference for identical products may vary up to 30%, depending on the location and the characteristics (for instance, browser configurations) of different online visitors. See Mikians et al. (2013) and Bar-Gill (2019).

learning-based credit scoring models than non-Hispanic white and Asian borrowers (Fuster et al. (2019)).

Even more worrying is the potential for intentional manipulation. Evidence suggests that big tech firms can exploit consumers' behavioural biases. For instance, one experiment based on about 670,000 unaware Facebook users found that people's emotional state can be projected onto others through contagion. This ability to make people experience the same emotions without being aware of the cause clearly raises economic, not to mention ethical, concerns (Kamer et al (2014)).[4]

When consumer preferences can be manipulated, the loss of surplus for consumers can be large. Panel D in Graph 5 illustrates this point. It represents what happens when a digital monopoly persuades its consumers to overestimate the benefit from a service or product. Graphically, the consumer demand curve shifts eastward (from D_0 to D_1). The overestimation causes some consumers to purchase the product, even though its actual value to them is lower than the price. As the additional surplus is only *perceived* (light red area), these consumers, who purchase the product only because of their misperception of the benefit, suffer an even greater welfare loss (light red areas) than under price discrimination. This outcome too is inefficient (grey area).

Anti-competitive behaviour

Another potential market failure could arise from big techs' *control over key digital platforms*. Once a captive ecosystem is established, potential competitors face steep costs and high risks in setting up rival platforms.

On the one hand, the fixed cost of setting up a new network of users, for instance, can be prohibitive.[5] This could allow big techs to engage in traditional anti-competitive practices such as "tying-in sales", cross-subsidising activities, and purchasing competitors.[6] On the other hand, big techs' search, mobile network, social network, or e-commerce platforms have become essential facilities for an ever wider range of business activities.

Platforms now often serve as essential selling infrastructures for financial services providers, while at the same time big techs compete with these same providers. When a network operator owns a smartphone-based payment system, for example, it can undermine competitors' access to its own digital platform by charging competitors (ie banks or rival big techs) high fees to connect with its (payment) system. Similarly, the owner of a search engine may redirect users away from competitors toward their

4 The notion that firms may actively change preferences and create wants, eg through advertising and salesmanship, is already present in Galbraith (1958). But the scope for such actions may be greater in the case of big techs, due to their command over much richer customer information and their integration into their customers' everyday life.

5 To date there is no evidence of big techs hindering their competitors' provision of financial services on their platforms. But examples of anticompetitive practices can be found in other sectors of activity, such as advertising. For example, in March 2019 the European Commission fined a big tech for imposing a number of restrictive clauses in contracts with third-party websites which prevented its rivals from placing their search adverts on those websites.

6 For example, Facebook acquired Instagram – a photo app – in 2012, Onavo – a data-security app that tracks users' smartphone activity – in 2013, WhatsApp – a messaging service – 2014; and Tbh – a social-polling app – in 2017. When Snapchat rebuffed its purchase offer in 2013, Facebook responded by cloning the app's most successful features.

own brands. With a captive user base, dominant platforms can raise the price of their services, and extract a larger share of customer surplus.

Privacy

When information is gathered without the informed consent of the consumer, it often infringes on personal privacy. Popular health websites have been found to share people's sensitive data (eg medical symptoms, diagnoses, drug names, etc) with dozens of companies around the world, including big tech firms such as Google, Amazon, and Facebook (Financial Times (2019a)). These risks are still greater when firms underinvest in data security, leading to data breaches (Carrière-Swallow and Haksar (2019)).

Furthermore, armed with this knowledge, companies may be in a better position to sell specific treatments, services or financial products that may not be in the users' interests. Consumer welfare will not necessarily benefit from the collection and use of personal data by profit-oriented firms, if these are left to their own devices.

The new policy trade-off between efficiency and privacy

The benefits and costs of the use of personal data in finance raises important policy questions. These go beyond the traditional ones of financial stability and competition, extending also to a new trade-off between data efficiency and privacy.[7]

Challenges for public policy

A first challenge is related to assigning the control and ownership of personal data. Control and ownership of data are rarely clearly defined. In some cases, users volunteer to provide their data for free (eg posts on social networks, online registrations). In other cases, companies infer personal information indirectly, eg through users' social network or internet searches. Users may also unintendedly and inadvertently surrender data, eg through their digital footprints or geo-localisation. Ownership and control of data is also difficult to re-assign, eg to users, due to the wide variety of data and the ways data are gathered, as well as their intangibility and non-rivalry. In many countries, the default outcome is that financial institutions or big techs acquire customer data at very low cost and keep *de facto* control.

A second challenge is addressing the heterogeneity in personal data. The efficiency gain from personal data-sharing crucially depends on their type. Some data are purely private or only meant to be shared with a restricted number of users – eg medical records. At the other extreme are data that people may want to share freely, and which can be shared without causing any harm. In between, there may be data that can be lent out (temporarily) and combined with other data, eg for credit assessments or insurance pricing. Some data are not valuable to users (eg browsing histories), but may be valuable to private sector companies, as they may help both general and customer-specific services to be better targeted. Users may want to sell such data to the highest bidder. But an efficient and complete market for personal

[7] See BIS (2019) and Petralia et al. (2019).

data has not yet emerged. To establish such a market, one must first determine who controls the data.

A third challenge relates to the value of privacy, and whether privacy should be traded off against other goals in the first place. In assessing the cost of the widespread sharing and use of data, one needs to consider how much people value their privacy.[8] Some argue that data privacy has the attributes of a fundamental right, which cannot be traded off against economic benefits.[9]

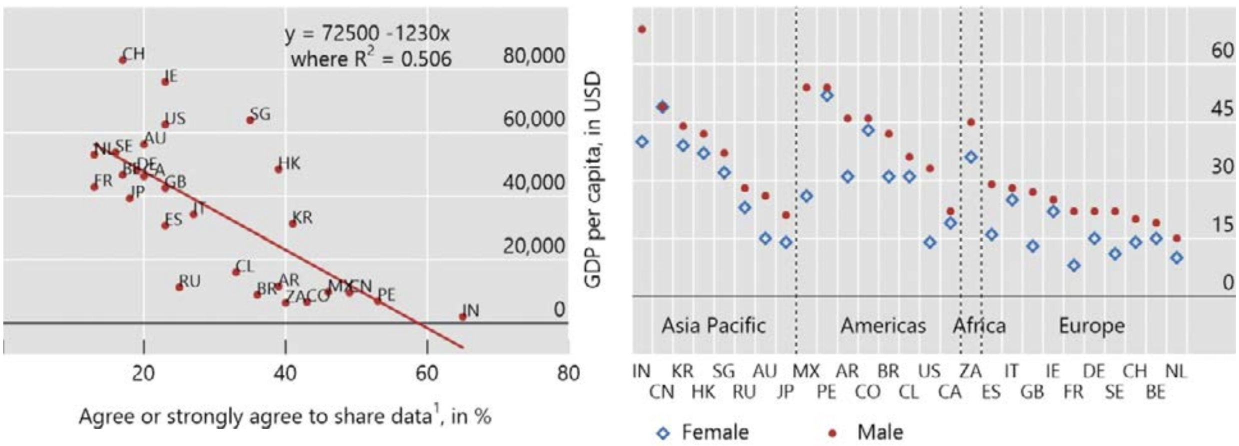
However, the evidence suggests that cultural preferences towards data privacy differ across jurisdictions, and even between different social segments (Chen et al (2020).). For example, in one recent survey, respondents were asked if they would be open to their bank securely sharing their data with other organisations in exchange for better offers on financial services.[10] In India, 65% of respondents said yes. In the Netherlands, this was only 13%.

Overall, it appears that willingness to share data correlates with per capita income, declining as incomes increase (see Graph 6, left-hand panel). Within jurisdictions, there are large differences by age and gender. For instance, 38% of 25- to 34-year-olds globally were willing to share their data, but only 16% of those over 65 were. At the global level, 34% of men were willing to share data, but only 27% of women, with this gap even larger in some countries (see Graph 6, right-hand panel).

Willingness of users to share data for better offers in financial services varies — Graph 6

Willingness to share data falls with GDP per capita | Willingness to share data is higher for men than women[1]

Per cent of respondents



[1] The question in the survey reads "I would be comfortable with my main bank securely sharing my financial data with other organisations if it meant that I received better offers from other financial intermediaries"; for Belgium the figure is calculated over Belgium and Luxembourg.

Sources: EY (2019); IMF, *World Economic Outlook*; Chen et al (2020).

Another important aspect in valuing privacy is with whom people are willing to share their private data. A recent survey on the anonymous sharing of genetic and medical information points, again, to wide cross-country differences. For example, more than 65% of the Chinese and Indian respondents indicate that they would share

[8] For a further discussion, see Acquisti et al. (2016).

[9] For a discussion of data rights in Europe, and of the grounding of such rights in eg the EU Charter of Fundamental Rights, see BEUC (2019).

[10] See EY (2019). The survey covered 27,000 consumers in 27 markets.

their health data with governmental authorities. In contrast, the proportion is less than 20% in European countries.[11] Such varied preferences make it difficult to reach universally acceptable solutions when designing controls and pricing data.

Policy options

There are several, potentially complementary, approaches to address the efficiency-privacy trade-off raised by the widespread use of personal data. One approach consists in restricting the processing of user data. For example, recent data protection laws (eg in the European Union, Brazil, California, Japan, Singapore) have clarified data collection and use to protect personally identifiable information. The challenge with these laws is how to balance differences in privacy concerns and use of data. For instance, India has instituted storage rules for payment system data motivated by privacy, access by regulators to necessary information and obligations to the judiciary in cases of legal disputes.[12] Some other jurisdictions have taken measures with a wider ambit that may restrict data flows across borders.[13] The rationale for such measures relates to law enforcement and monitoring and supervision purposes. Yet, such frictions in the use of data could lead to cost inefficiencies, limiting their potential benefits.[14]

A second approach is to give consumers greater control over their personal data. This could allow customers to grant competitive firms access to relevant information, and can thus foster competition. Recent open banking initiatives (eg in the European Union, Australia, and Mexico) are examples of concrete policy actions in this direction. These policies require financial firms to make their customers' transactions (or equivalent) data portable, ie directly transferable to third parties or competitors, typically through open application programming interfaces (APIs). Open banking rules selectively restrict the range of data that can be transmitted (eg financial transaction data), as well as the type of institutions among which such data can be shared (eg accredited deposit-taking institutions). In this sense, they help resolve inefficiencies through the allocation of property rights and the creation of a competitive market for data – the decentralised or "Coasian" solution.[15] As access is asymmetrical, though, open banking rules do not fully level the playing field between big tech firms and incumbent service providers.

A third approach is a set of public infrastructures on which a layer of services can be built. This includes important foundations for digital services such as digital identity, like Aadhaar in India and MyInfo in Singapore, and the development of data management protocols. Once these infrastructures are in place, payments, digital government services and a host of other solutions become possible. For example, the launch of India's Unified Payment Interface (UPI) facilitated entry by new firms and

[11] The Welcome Sander Institute, "Your DNA, Your Say" global online survey is on ongoing survey that gathers public attitudes towards genomic data sharing for over 37,000 individuals in 22 countries.

[12] India does not restrict the transfer of payment system data overseas if one of or more of the counterparties is a non-resident, or for processing purposes even when both counterparties are residents, but in the latter case local data storage is mandatory as the data belong to Indian citizens.

[13] One example is China (see Cyberspace Administration of China (2019)).

[14] According to Aaronson (2019), 58% of the countries in the world have now adopted or are adopting data protection laws. Many such laws contain provisions affecting cross-border data flows. It is still too early to assess whether such laws are effective in addressing risks (see also Mitchell and Mishra (2019)).

[15] This is named after Ronald Coase; see Coase (1960).

spurred competition, which drastically reduced prices for consumers.[16] By making consumers data-rich and giving them greater control over their data, important benefits for users can be reaped (Nilekani (2018)).

Domestic and international policy coordination

How to define and regulate the use of data are issues that need to be coordinated at both the domestic and international level.

At the domestic level, central banks and financial regulators may need to upgrade their understanding of personal data issues. And they need to coordinate with competition and data protection authorities. Yet the mandates and practices of these bodies may not always be compatible. For example, the specificity of the financial sector may not accord with the general standards that competition and data privacy laws often apply to a wide range of industries. Moreover, financial regulation is often based on international standards, while data protection and competition policy are mostly national – to the extent that countries may have a unique competition or data protection authority.[17]

At the international level, regulations on the use of personal data diverge widely. In the European Union, the General Data Protection Regulation (GDPR) assigns data rights to individuals. In India, the India Stack generates large volumes of new data, and users have control over them, but data privacy regulations have to yet be adopted. In China and several other countries, data localisation rules prevent data from being shared across borders. In the United States, a patchwork of sector-specific legislation means that, in practice, companies have relatively free access to data, and some companies, most famously Apple, have resisted calls to share data with public authorities (Apple (2016)). Meanwhile, only a few countries have a national data or artificial intelligence strategy.[18] As the digital economy expands across borders, there is a need for international cooperation on rules and standards.

[16] See Financial Times (2019b).

[17] For example, competition policy in the United States falls under the remit of both the Department of Justice and the Federal Trade Commission.

[18] See https://datagovhub.org/.

References

Aaronson, S A, "What are we talking about when we talk about digital protectionism?", World Trade Review, vol 18, no 4, October 2019.

Acquisti, A, C Taylor and L Wagman, "The economics of privacy", *Journal of Economic Literature*, vol 54, no 2, 2016.

Apple, "A message to our customers", February 2016.

Bar-Gill, O, "Algorithmic price discrimination: when demand is a function of both preferences and (mis)perceptions", *University of Chicago Law Review*¸ vol 86, 2019.

Bureau Européen des Unions de Consommateurs (BEUC), "Access to consumers' data in the digital economy", position paper, November 2019.

BIS, "Big tech in finance: opportunities and risks", *Annual Economic Report 2019*, Chapter III, June 2019.

Carrière-Swallow, Y and V Haksar, "The economics and implications of data: an integrated perspective", *IMF Departmental Papers*, vol 19, no 16, 2019.

Coase, R, "The Problem of Social Cost", *Journal of Law and Economics*, Vol 3, 1960.

Cyberspace Administration of China, "Security Assessment of Cross-border Transfer of Personal Information", June 2019.

Chen S, S Doerr, J Frost, L Gambacorta and H S Shin, "The fintech gender gap", mimeo, 2020.

EY, *Global FinTech Adoption Index*, 2019.

Farboodi, M, R Mihet, T Philippon and L Veldkamp, "Big data and firm dynamics", *NBER Working Papers*, no 25515, January 2019.

Financial Times, "How top health websites are sharing sensitive data with advertisers", 13 November 2019a, www.ft.com/content/0fbf4d8e-022b-11ea-be59-e49b2a136b8d.

Financial Times, "The Indian payments conundrum", 18 November 2019b.

Frost, J, L Gambacorta, Y Huang, H S Shin and P Zbinden, "BigTech and the changing structure of financial intermediation", *BIS Working Papers*, no 779, April 2019.

Fuster, A, P Goldsmith-Pinkham, T Ramadorai and A Walther, "The effect of machine learning on credit markets", *VoxEU*, 11 January 2019.

Galbraith, J, *The affluent society*, 1958.

Gambacorta, L, Y Huang, H Qiu and J Wang, "How do machine learning and non-traditional data affect credit scoring? Evidence from a Chinese fintech firm", *BIS Working Papers*, no 834, 2019..

Hau, H, Y Huang, H Shan and Z Sheng, "Fintech credit, financial inclusion and entrepreneurial growth", mimeo, 2018.

Huang, Y, C Lin, Z Sheng and L Wei, "Fintech credit and service quality", mimeo, 2018.

Jullien, B, "Two-sided markets and electronic intermediaries", *CESifo Working Papers*, no 1345, November 2004.

Kamer, A, J Guillory and J Hancock, "Experimental evidence of massive-scale emotional contagion through social networks", *Proceedings of the National Academy of Sciences of the United States of America*, vol 111, no 24, March 2014.

Mikians, J, L Gyarmati, V Erramilli and N Laoutaris, "Crowd-assisted search for price discrimination in e-commerce: First results", in *Proceedings of the Ninth ACM Conference on Emerging Networking Experiments and Technologies*, pp. 1–6, 2013.

Miller, D, and P Friesen, "A longitudinal study of corporate life cycle", *Management Science*, vol 30, no 10, 1984.

Mitchell D and N Mishra, "Regulating cross-border data flows in a data-driven world: how WTO law can contribute", Journal of International Economic Law, vol 22, no 3, September 2019.

Nguyen Trieu, H, "Why finance is becoming hyperscalable", Disruptive Finance, 24 July 2017.

Nilekani, N, "Data to the people: India's inclusive internet", *Foreign Affairs*, September/October 2018.

O'Neil, C, *Weapons of math destruction: how big data increases inequality and threatens democracy*, Broadway Books, 2016.

Petralia, K, T Philippon, T Rice and N Veron, "Banking Disrupted? Financial Intermediation in an Era of Transformational Technology", *Geneva Reports*, no 22, 2019.

Previous volumes in this series

969 October 2021	Inter-agency coordination bodies and the speed of prudential policy responses to the Covid-19 pandemic	Michael Brei and Blaise Gadanecz
968 October 2021	Indebted demand	Atif Mian, Ludwig Straub and Amir Sufi
967 October 2021	Joined at the hip: monetary and fiscal policy in a liquidity-driven world	Guillermo A Calvo and Andrés Velasco
966 October 2021	The Treasury market in spring 2020 and the response of the Federal Reserve	Annette Vissing-Jørgensen
965 October 2021	Technological capacity and firms' recovery from Covid-19	Sebastian Doerr, Magdalena Erdem, Guido Franco, Leonardo Gambacorta and Anamaria Illes
964 September 2021	Macroeconomic policy under a managed float: a simple integrated framework	Pierre-Richard Agénor and Luiz A Pereira da Silva
963 September 2021	The Fed takes on corporate credit risk: an analysis of the efficacy of the SMCCF	Simon Gilchrist, Bin Wei, Vivian Z Yue and Egon Zakrajšek
962 August 2021	Global lending conditions and international coordination of financial regulation policies	Enisse Kharroubi
961 August 2021	Private equity buyouts and firm exports: evidence from UK firms	Paul Lavery, Jose-Maria Serena, Marina-Eliza Spaliara and Serafeim Tsoukas
960 August 2021	Is window dressing by banks systemically important?	Luis Garcia, Ulf Lewrick,and Taja Sečnik
959 August 2021	Macroeconomic effects of Covid-19: a mid-term review	Phurichai Rungcharoenkitkul
958 August 2021	Sharing asymmetric tail risk: Smoothing, asset pricing and terms of trade	Giancarlo Corsetti, Anna Lipińska and Giovanni Lombardo
957 August 2021	Ripple effects of monetary policy	Frederic Boissay, Emilia Garcia-Appendini, and Steven Ongena
956 August 2021	Are households indifferent to monetary policy announcements?	Fiorella De Fiore, Marco Lombardi and Johannes Schuffels

All volumes are available on our website www.bis.org.