Customer Data Access and Fintech Entry: Early Evidence from Open Banking[†]

Tania Babina[‡] Greg Buchak[§] Will Gornall [¶]

Current Version: April 21, 2022
First Version: September 18, 2021

Abstract

Open banking is the trend of empowering customers to share their banking data with fintechs and other banks. We compile a novel dataset documenting that governments in 49 countries have implemented open banking policies and 31 more are in active discussions. Following adoption, fintech venture capital investment increases by 50%, with more comprehensive policies showing larger effects. We examine the policy tradeoffs with a quantitative model of consumer data production and usage. Our calibrations show that customer-directed data sharing increases entry by improving entrant screening ability and product offerings, but harms some customers and can reduce ex-ante information production.

Keywords: Open banking, entrepreneurship, fintech, financial innovation, data access, data rights, data portability, Big Data, financial regulation, financial sector, banks

[†]Babina is the corresponding author and is located at Columbia University, 665 West 130th Street, New York, NY 10027. We are grateful to Mark Boyd for sharing Platformable's data, and Zhaochen Jiang for outstanding research support. For comments, we are grateful to Isha Agarwal, Jonathan Berk, Bo Bian, Charlie Calomiris, Olivier Darmouni, Douglas Diamond, Maryam Farboodi, Sean Higgins, Steve Kaplan, Arvind Krishnamurthy, Luc Laeven, Stefan Nagel, Tomek Piskorski, Ragu Rajan, Amit Seru, Amir Sufi, Luigi Zingales, and seminar and conference audiences at Columbia; Stanford; NBER Economics of Privacy Conference; UBC; UBC Winter Finance Conference; Universidad Carlos III de Madrid; USC. Gornall thanks the SSHRC for their support.

[‡]Columbia. Email: tania.babina@gsb.columbia.edu.
[§]Stanford. Email: buchak@stanford.edu.
[¶]University of British Columbia. Email: will.gornall@sauder.ubc.ca.

The increasing ease with which data are collected, stored, and analyzed has made data a critical input in economic decision making. Data's growing economic importance has led to an active discussion around who should control the data generated through private economic activity: a firm or its customers. This issue is particularly salient in the financial services sector, where banks' provision of financial products inherently generates useful customer data. Periodic direct deposits, overdrafts, and late payments help predict a potential borrower's riskiness. Transactions are informative about price sensitivity and consumption preferences. Account balances and spending patterns are useful for customized financial advice. Importantly, these data have historically been under the bank's exclusive control. This control led to ex-post market power in providing additional products and provided additional incentives to form customer relationships ex-ante.

Data give banks a comparative advantage in pricing, marketing, and customizing financial services. As a motivating example, Figure 1 Panel (a) shows that non-banks and fintech lenders, which lack such customer data, overwhelmingly use standardized underwriting models such as FICO when originating US residential mortgages. Banks are much more likely to use non-standard credit models, allowing them to exploit their data. These non-standard models lead to more individualized pricing: Panel (b) shows that non-standard models lead to more dispersed interest rate residuals than standard models.

Banks' exclusive customer data access is being upended by a movement known as open banking (OB). OB is the trend of empowering customers to share their banking data with other financial service providers. For example, a recent immigrant may have a bank account and a job but a limited credit history. OB allows her to use a phone application or website to easily share her bank account history with potential lenders. Access to these data lets lenders confirm her employment and income and helps her get credit.

While some banks have implemented OB of their own accord, many governments have taken an active role in promoting or even mandating it. As of October 2021, the regulators of 80 countries have taken steps—some major and others still tentative—to implement policies to promote the adoption of OB. Many consumers and businesses seem eager to take ownership of their data—for example, South Korea reports 30 million users and 100 million accounts just two years after implementing OB.[1] Policymakers hoping for increased competition and innovation reason that allowing customers to share their bank data will allow new entrants and other banks to better compete for business. This could lead to innovative entry, lower prices, and greater access to financial products and services. However, and largely absent from policy discussions, mandated data sharing raises concerns around the distributional consequences of such policies in the short run and their longer-term impact on ex-ante information production.

In this paper, we explore the causes and consequences of government policies to promote

[1] See here.

OB. We first test whether OB policies have achieved their stated objectives of promoting data sharing and innovative entry. We then contextualize our reduced-form findings in a calibrated quantitative model measuring the observed benefits against potential tradeoffs. In doing so, we make four key contributions to the literature on banking, entrepreneurship, and the economics of data. First, we assemble the first comprehensive, standardized dataset of government-led OB policies. Our hand-collected data are instrumental to our empirical findings and pave the way for future research as the longer-term effects of OB unfold. Second, we document the rich heterogeneity in policy approaches and examine the political and economic forces that might explain the introduction of these policies. Third, we provide preliminary evidence on the effects of OB policies on incumbent banks and the creation of fintech startups. Finally, we provide and calibrate a quantitative modeling framework for data production and use applicable to both OB and more general contexts.

We begin with careful data collection on worldwide OB approaches, with a particular emphasis on government policies to promote OB. We uncover vast heterogeneity in OB policy details, with regulators facing many consequential choices. For example, countries in the European Union (EU) have tended to adopt OB regimes with mandatory data sharing by banks but without regulator-supplied technical standards. In contrast, East Asian countries have favored voluntary participation but spelled out detailed technical standards. We classify key dimensions of OB government policies and assemble a comprehensive dataset covering the 168 largest countries—covering more than 99% of world GDP and 98% of world population. We summarize our granular policy data by constructing a country-level OB Strength Index, which captures the comprehensiveness of government-led OB policies and can be used as a measure of OB policy treatment intensity. Interestingly, we find that OB policy adoption is not well-predicted on the margin by ex-ante country characteristics related to economic or financial development, levels of innovation, or the quality of local institutions.

We begin our assessment of the effectiveness of OB policies by examining whether OB policies have indeed led to increased access to bank customer data. We look directly at the creation of open Application Programming Interfaces (APIs) by banks—the technology used to access and share customer bank data—and show that their prevalence is twice as high in countries with OB policies. Moreover, more comprehensive OB policies, as measured by our OB Strength Index, are associated with greater use of APIs by banks. This suggests OB policies are achieving their proximate objective of increasing data access.

We next provide preliminary evidence on whether OB policies successfully promote innovative entry, which we measure using venture capital (VC) investment into fintech startups. We use the staggered implementation of OB policies and a standard difference-in-difference analysis to show that the number of VC-backed fintech financings increases by half and the amount of money invested doubles following OB policy adoption. Larger increases in fintech VC activity are associated with more comprehensive policies, as measured by our OB

Strength Index.

Because these policies are fundamentally endogenous, we make several arguments for causality. First, and perhaps most importantly, we provide event studies that show a discontinuous increase in startup numbers and fundraising volumes after the introduction of OB policies, with no clear pre-trends. The absence of pre-trends supports the identification assumption and ameliorates the reverse causality concern that growing fintech VC investment leads governments to adopt OB policies. Second, our results are robust to including region-by-year fixed effects, ruling out that fintech VC investment and OB adoption are jointly the result of region-specific economic or political trends. Third, we address the potential concern that countries implementing OB simultaneously enacted broader innovation-promoting policies by showing that our results are robust to controlling for contemporaneous non-fintech VC investment and that OB does not predict more non-fintech VC. Fourth, we conduct a falsification test using VC funding of cryptocurrency startups—an industry we posit that, so far, has not greatly benefited from OB, but might be moved by the same potential confounders. Finally, we show that our results are present on both the extensive and intensive margins of OB adoption, with relatively weak policies not having measurable effects.

We also look for evidence of increased financial inclusion and competition. Our results here are inconclusive: we find little effect on remittance costs, bank account uptake, or measures of bank competition. We caveat that these so-far null results are limited for three reasons. First, outcome data are limited by the recent adoption of these policies. Second, even when data are available, substantial real effects may take years to materialize. Third, our model has ambiguous predictions for the effect of OB on outcomes other than entry, as the next paragraphs discuss.

Motivated by the reduced-form evidence, we provide a general-purpose quantitative model of data production and usage and calibrate it to two financial products: mortgages and financial planning advice. The model is based on standard IO models of consumer choice with heterogeneous consumers. We incorporate data use into the model by allowing some firms—incumbent banks—to observe consumer preferences and marginal costs, while potential new entrants only observe distributions. Our model nests many notions of data use in the literature, including screening, price discrimination, and the creation of better products. Unequal data access gives incumbents monopoly power to provide more targeted products and exposes potential entrants to adverse selection, which together discourage new entry. OB makes these data more widely available, thus mitigating these frictions and encouraging the endogenous entry of new firms.

This model allows us to frame the economics of our reduced-form results. Using structural parameters calibrated on pre-OB data, our model predicts increased fintech entry after OB adoption that is quantitatively in line with our reduced-form findings. Additionally, the model's comparative statics are consistent with the heterogeneity analysis in the data. Beyond rationalizing the findings on entry, our model highlights two key drawbacks of mandated data access. First, depending on how data is used, some consumers—in particular customers who are costlier or have a higher willingness to pay—are hurt by wider data availability, even if they can choose to opt out. This potential cost is most relevant in OB use cases that use data to screen or price discriminate against potential customers. In contrast, these distributional effects are not present in OB use cases where the data is used to provide higher quality or more targeted products, where customer welfare increases broadly. Our reduced-form results point to increased entry in both types of applications, suggesting that both are quantitatively important.

Beyond these distributional effects, the model highlights a second potential drawback of OB. The ex-post data monopolies enjoyed by incumbents give them extra incentive to produce data ex-ante. Eliminating these ex-post rents reduces their ex-ante incentives to gain customers and can reduce financial inclusion in the long run. Though a full estimation of the model is beyond the scope of this paper, our calibrations suggest that there are quantitatively important tradeoffs for policymakers to consider when adopting OB and that OB's overall welfare effects are ambiguous even when it leads to observably more fintech entry.

To summarize, we find that the adoption of OB government policies leads incumbent banks to invest in technology to share customer data and spurs VC investment in fintechs. Weaker OB implementations are measurably less effective. The potential implications of OB for academics, policymakers, and industry are large. By giving the customers the ability to share their financial data, OB promises to upend the organization of the financial sector while increasing competition and financial innovation. The welfare effect of this, however, is far from obvious, as our model highlights, which calls for additional research on specific use cases and OB implementations.

Our paper proceeds as follows. In Section 1, we situate our contribution in the literature. In Section 2, we detail our data collection, summarize OB approaches around the world, and explore factors that give rise to OB policies. In Section 3, we examine the effects that OB policies have had so far. In Section 4, we provide an economic framework for evaluating our results, and in Section 5 we conclude.

1 Related Literature

Our paper contributes to several strands of literature. First, our research question and methodology connect to the broader literature on cross-country bank regulation. In the wake of the financial crisis, much of this literature focuses on regulation and bank risk, for example, Laeven and Levine (2009), Beck et al. (2013), and Ongena et al. (2013). Our paper is closer to research on regulation and competition, such as that by Claessens and Laeven (2004) who argue contestability and regulation are key drivers of bank competition or Barth et al. (2004)

who argue for the role of disclosure and private incentives. We contribute by showing that government policies to promote bank customer data sharing foster entry into the financial sector.

Second, we engage with the fundamental question, originating with Diamond and Dybvig (1983) and Diamond (1984), over what, if anything, makes banks special relative to other financial intermediaries. While fintechs and other non-depository institutions have gained significant market share in transaction-oriented functions like origination and servicing, as Gopal and Schnabl (2020) and Buchak et al. (2018b) show, they have been slower to replace banks in deeper intermediation roles like underwriting, monitoring, and balance sheet lending. Importantly, banks appear to derive significant value from engaging in multiple intermediation activities simultaneously, as in Egan et al. (2017), Aguirregabiria et al. (2019), or Benetton et al. (2021), which suggests there may exist significant barriers that limit the growth of new single-product competitors in these roles.

Information lies at the heart of relationship banking (Ramakrishnan and Thakor, 1984; Boot and Thakor, 1997) and our paper directly addresses the idea that aggregating data across multiple business lines leads to significant informational advantages. This explanation dates to Petersen and Rajan (1994), Petersen and Rajan (1995), and more recently Granja et al. (2018). Recent empirical work by Ghosh et al. (2021) shows, for example, a direct effect of transaction data on screening quality for Indian commercial loans. Berg et al. (2020) and Di Maggio et al. (2021) show the value of alternative data more generally. OB provides an empirical setting in which banks potentially lose the informational advantage that their wide scope provides to them, and paves the way for an analysis of how important these informational advantages are to banks, consumers, and entrants.

Third, we add to the nascent literature examining the implications of data ownership rights. The growing theoretical work on data use typically views data as either an input to production that improves product quality or a way to improve screening or monitoring in settings with information asymmetries. Mandated data sharing generates complex competitive interactions that depend on how data are used. Taking the production-input view, Jones and Tonetti (2020) show that a firm may hoard product-improving data to prevent entry, and giving data property rights to consumers can generate allocations that are close to optimal. Farboodi et al. (2019) model customer-generated data as valuable in forecasting business conditions and suggest that large firms benefit more from data, a fact confirmed empirically by Babina et al. (2021) who show that larger firms benefit more from their AI investments. Emphasizing the information economics view, theoretical finance literature like He et al. (2020) and Parlour et al. (2020) highlight how data sharing and portability can increase the quality of lending while having ambiguous effects on consumer welfare and bank profits.

We build on this largely theoretical literature in two ways. We provide, to the best of our knowledge, the first empirical study on the impact of government policies that open access to rich customer-level financial and transaction data. While conceptually related to credit registries, e.g., Djankov et al. (2007) and Hertzberg et al. (2011), OB policies differ in important respects. They typically cover consumers regardless if they use credit products and cover much richer data types (including transactions, income, and savings data); they give consumers the option of opting in while typically requiring banks' participation; and they are designed from the outset to facilitate ease-of-data-access by potential bank competitors. As we show, these aspects of OB are important in driving its effects. Thus, our paper speaks to these important differences while also providing evidence of the effects of adopting data-sharing policies more generally. Beyond that, we provide a general-purpose quantitative framework for studying the production and use of consumer data in the context of OB. Building on common tools in the IO/finance literature, (e.g., Egan et al. (2017), Di Maggio et al. (2021), Buchak et al. (2018a), Benetton et al. (2021)) we connect data to knowledge of consumer heterogeneity around marginal costs, willingness to pay, and desired customization. Through these channels, we synthesize both the input-to-production and information economics views of data, and highlight their quantitative importance across particular applications. In contrast to the theoretical models of, e.g., He et al. (2020) and Parlour et al. (2020), our model emphasizes new firm entry and innovation, which is a key policy goal of OB. Because it is quantitative, the model can be easily and credibly calibrated with standard techniques and estimates already in the literature.

Fourth, our structural model allows us to connect to and broaden the literature around the industrial organization of the financial sector. This literature has studied the role of banks and the increased competition they face from non-depository institutions, e.g., Buchak et al. (2018a), Buchak et al. (2018b), Fuster et al. (2019), Jiang et al. (2020) (mortgages), Erel and Liebersohn (2020), Gopal and Schnabl (2020) (small business lending in the US), Di Maggio and Yao (2021), De Roure et al. (2021) (personal loans), and Buchak et al. (2021) (deposits). These papers typically highlight the complex interplay between technology and regulation and how they interact with the comparative advantages of depository and non-depository institutions.[2] Our results also connect to the growing literature on financial system structure and financial inclusion (e.g., Claessens and Rojas-Suarez (2016), Bartlett et al. (2022), or Philippon (2019)).

Finally, our paper is connected to the literature on the drivers of entrepreneurship and innovation.[3] We show a large effect of OB policies on innovative entrepreneurship, which adds

[2] Literature reviews on the impact of technology in finance can be found in Stulz (2019), Vives (2019), Allen et al. (2020), Thakor (2020), Berg et al. (2021), and Boot et al. (2021).

[3] Entrepreneurs play a crucial role in prominent theoretical explanations for economic growth, including Schumpeter (1911), Lucas (1978), and Baumol (1990). Relative to incumbent firms, new firms have faster productivity and employment growth. This literature includes Kortum and Lerner (2000), Foster et al. (2008), Gennaioli et al. (2012), Haltiwanger et al. (2013), Decker et al. (2014), Glaeser et al. (2015), and Akcigit and

to a literature that has shown mixed results on whether policymakers are able to promote high-growth entrepreneurship. Acs et al. (2016) question the general effectiveness of public policies to encourage entrepreneurship, with subsidies of angel investing found to be ineffective (Denes et al., 2020), while Bai et al. (2021) argue government funding of early-stage companies increases local innovation. Other work shows the positive impact of less entry regulation (Klapper et al., 2006; Mullainathan and Schnabl, 2010), more optimistic beliefs (Puri and Robinson, 2007), venture capital availability (Kaplan and Lerner, 2010), weaker competition laws (Phillips and Zhdanov, 2017), lower investor eligibility requirements (Lindsey and Stein, 2019), R&D subsidies (Babina and Howell, 2018) and academic funding (Babina et al., 2020).

Beyond policy impacts, we join the relatively sparse literature connecting data access to innovation. Recent work by Ahnert and Doerr (2021) shows that bank use of information technologies increases employment in new firms. We contribute by showing that government efforts to promote data sharing in the financial sector have fostered investments in VC-backed fintechs.

2 Institutional Background, Data, and Descriptive Analysis

This section describes the institutional background of OB, describes the data collection process, and provides high-level summary statistics.

2.1 Institutional Background on Open Banking

OB describes a broad trend where upon customer request, financial intermediaries share—willingly or by regulatory fiat—access to their customers' data with other financial service providers. There are two primary non-mutually exclusive ways in which OB is spreading around the world: industry-led, where banks and fintechs adopt OB without government intervention, and government-led, where regulators institute policies to promote the adoption of OB by the financial sector. This paper primarily focuses on government-led OB.

While the specifics of government OB efforts vary dramatically, the United Kingdom's (UK) Open Banking Initiative provides an instructive introduction: in 2017, the UK's Competition and Markets Authority introduced one of the first OB initiatives, with the aim of increasing innovation and competition in the retail banking sector. The initiative required that by 2018, the nine largest banks "*give their personal and business customers the ability to access and share their account data on an ongoing basis with authorized [by the government] third parties.*"[4] Here, third parties refers to both fintechs and other banks. Additionally, these nine banks were required to allow those third parties to make payments authorized

Kerr (2018).

[4]Page 11 of "Open Banking, Preparing for Lift off" document. Link to the official policy document.

by banks' customers from their accounts—a practice called payment initiation. OB differs from the UK's existing private sector credit registries in important ways: it gave consumers more control over the data, it covered more data, it was free to the requester (with the customer's permission) and, perhaps most importantly, banks were forced to participate. These differences are typical and mean that OB goes well beyond traditional credit registries.

Data access and payment initiation typically occur through a bank-provided API. APIs are a technology that allows two computer systems (e.g., a bank's and a fintech's) to speak to each other over a network. OB APIs are published by the data provider and are a set of standardized, programmatic commands that allow data users to interact with the provider's customer database and to perform financial services on customers' behalf. The particulars are regime-specific, but API functionality in OB typically allows read access (e.g., querying account data) and sometimes allows write access (e.g., payment initiation).

By opening bank data, regulators aim to create an environment where financial intermediaries—both incumbents and fintech entrants—can create new or improved financial services for bank customers and better compete with existing services. The prototypical use case is customer financial account aggregation. A typical person has financial accounts scattered across several financial intermediaries: her bank account; several credit cards; a mortgage; an investment account; and so on. Rather than separately monitor each of her accounts, she may find it helpful to have this information collected and displayed in a single place. This also facilitates budgeting, customized financial planning, and other innovative applications. What are her spending habits? Does she have recurring payments or subscriptions she may have forgotten about? Which credit product should she pay down first? How much should she contribute monthly to her retirement account if she wishes to retire by a certain age? With OB, fintech startups can access, aggregate, and analyze these separate accounts to provide customized financial advice.

Other use cases of OB include consumer lending, where potential lenders can access the myriad, and otherwise private, information that a consumer's home bank has about her. For example, with a customer's permission, a fintech lender could use the customer's bank's API to query her bank account transactions and payroll information to help price a loan to that customer. In this way, OB can reduce search costs and level the information playing field between a consumer's home bank and potential competitors. Beyond financial advice and consumer lending, many other use cases have emerged, including automatic overdraft borrowing, product suggestions using customer data, SME lending, accounting, and identity verification.

While API-enabled OB is currently mainstream, fintechs have historically achieved similar functionality through what is known as "screen scraping" where a customer gives her login credentials for each of her financial institutions to the fintech (e.g., Mint.com). The fintech's software then uses the customer's credentials to log in to each financial institution and extract account data from the financial institution's webpage. Although screen scraping accomplishes similar results to accessing an OB API, screen scraping has numerous weaknesses, including security risks, privacy issues, inefficiency, and unreliability.[5] The API-enabled OB approach allows for a better controlled, more secure, and targeted access to the data that the customer intends to share.

2.2 Data Collection Methodology for Open Banking Around the World

We create a comprehensive and detailed database of OB government policies around the world. Our hand-collected dataset details the OB government policies (or the lack thereof) of the largest 168 countries. This section describes our methodology broadly; Appendix A provides further detail. We base our sample on countries with at least one million people according to the IMF 2018 data or at least 10 VC-backed companies.[6] We aim to be as comprehensive as possible while focusing on a sample of countries for which there is reliable data on OB initiatives, if they exist. In total, we collect data on OB for 168 countries, representing more than 98% of global population and more than 99% of global GDP.

For each country, we manually search for official OB policy documents using Google, and when those are not available, for descriptions of government-led OB initiatives from law firms, research papers, journalists, and industry participants.[7] We classify these policies on multiple dimensions, giving preference to official policy documents (laws, regulations, policy papers, and official statements) to classify the various dimensions of OB policies into standardized categorical variables. Where official policy documents are unavailable, we use other sources.

We ensure accuracy by performing multiple cross-checks. First, two authors independently classify each country's regime and jointly reconcile any discrepancies. Second, we use automated news topic searches to uncover any material potentially missed in our manual searches.[8] Third, we reconcile our results against a database of OB regulations maintained

[5]First, screen scraping creates security risks because it requires the fintech to store the user's full account credentials, trains users in the bad security practice of handing over credentials, violates bank fraud protection rules, and is incompatible with two-factor user authentication. Second, screen scraping creates privacy issues because it requires the user to give the scraper access to all the user's information with a financial institution, rather than the specific information they want to share. Third, screen scraping is inefficient both for the fintech (who must implement separate code for each bank website) and for the bank (whose web servers must generate large amounts of webpage content irrelevant for the fintech's software). Finally, screen scraping is unreliable and is frequently broken by simple website changes.

[6]The IMF data are from here. The VC data are from PitchBook and are described later in this section.

[7]We use Google as our primary search engine because it has the lion's share of the world search market (88% in June of 2021; see Statista.com). To ensure that using Google does not bias our findings for countries that rely more on other search engines, we also tried using local search engines (e.g., "Yandex" in Russia, "Baidu" in China). We generally found that these alternative search engines did not provide additional relevant articles.

[8]For a given country, a program searches Google for all news articles mentioning ("country name" and "open banking" and ["government" OR "central bank" OR "law" OR "regulation" OR "regulatory framework" OR "supervision")]. This search provides a list of sources of potentially relevant information on government OB interventions to compare to our manual collection. A research assistant then reads the top 10 resulting

by Platformable,[9] an OB advocacy group.

2.3 Summary Statistics on Open Banking Regimes

As of October 2021, 87—or 52%—of the 168 countries in our sample have at least a nascent OB effort. Some of these are market-led and have no government involvement. Of the 80 government-led pushes, many have not left the early-discussion phase, while some regulators have fully implemented their policies or even moved on to follow-on regulations. Table 1 provides summary statistics on our hand-collected OB data both overall and by region, which we now discuss.

Market- versus Government-led Approaches Two main non-mutually exclusive ways in which OB is spreading around the world are government-led, where regulators institute government policies to promote the adoption of OB by the financial sector, and industry-led, where the financial sector participants coordinate on adopting OB. While relatively few countries have purely market-driven approaches, government-driven approaches are common. We find that 48% of countries—nearly all countries with any OB whatsoever—have a government-led approach of some kind.

While this paper focuses on government-led OB policies, 29% of countries have market-driven OB initiatives. The US and Switzerland are prototypical examples: industry consortiums coordinated to create standards for OB APIs with little direct government intervention.[10]

There is significant heterogeneity by region,[11] with Europe & Central Asia having the highest degree of both market-led and government-led OB regimes at 66% and 80% respectively. East Asia's approach tilts more towards government-led, with only 21% of countries having a market-led approach compared to 63% with a government-led approach. Other regions have relatively less OB, although OB is present in all regions, including Sub-Saharan Africa and Latin America.

For government-led approaches, regulators frequently cite one or more policy justifications or policy mandates for implementing OB regimes in their official policy documents and interviews. The three most common are to promote innovation, competition, and financial inclusion. Table 1 shows that 97% of regulators cite innovation as a policy mandate; 82% cite competition, and 29% cite financial inclusion.[12] There is significant regional heterogeneity in

articles for any discussion on government OB–however small—and flags articles for the review by the authors.

[9] Platformable's data are described here.

[10] See here and here. In counties that do not mandate banks to share data or in market-led regimes, customer data sharing is based on bilateral negotiations and contracts between banks that decide to share customer data and fintechs that use those data to provide financial services. For example, in the US these contracts can include language stipulating that the fintech may not provide competing services (see here).

[11] Regions are based on the World Bank classification.

[12] This variable is missing for countries with no regulatory OB approach and for countries in the early stages

financial inclusion being an OB policy goal: only 10% of countries in Europe & Central Asia cite financial inclusion, whereas other regions are much more likely to do so. Additionally, OB laws commonly exist beside a related regulatory effort on either data privacy rights[13] or more rarely, general data sharing rights.[14]

Finally, we note that the EU adopted and implemented a common OB framework known as the Revised Payment Services Directive (PSD2).[15] PSD2 obligated participating countries to implement its provisions in their respective banking regulations. In the country-level summary statistics later in this section, we keep the participating countries separate. For the purposes of our analyses in Sections 2.5 and 3, we weight all countries covered by the PSD2 as a single pooled observation.

Implementation Status and Key Dates of Government-led Policies Government-led OB approaches vary both in when they were adopted and how far the implementation has progressed. For countries with some government interest in OB policies, we categorize a country's OB implementation status on a 0 to 7 scale, where 0 denotes no effort toward OB, 1–2 correspond to ongoing policy discussions, 3–5 correspond to being in the process of implementation, and 6–7 correspond to full implementation.[16]

Panel (a) of Figure 2 shows the geographical distribution of government-led OB initiatives based on their maturity. As of October 2021, among countries with a government-led approach to OB, 31 (38%) are at the discussion stage, 14 (18%) are in the process of implementation, and 35 (44%) are fully implemented or already seeing follow-on policies. We refer to the 49 countries in the latter two groups as having implemented OB. To provide three examples along the implementation timeline, OB discussion is underway in the US,[17] Brazil is in the process of implementing OB,[18] and the UK has fully implemented its Open Banking Initiative and is considering a follow-on "open finance" regulation.[19] Figure 2 Panel

of implementation that have not clearly indicated a policy mandate or objective.

[13]E.g., the General Data Protection Regulation (GDPR) in the EU.

[14]E.g., open data, which extends beyond the financial sector.

[15]PSD2, Directive (EU) 2015/236.

[16]More specifically, the stages are (1) pre-discussion (some government interest is announced but no actual law or policy implementation is taking place); (2) discussion (the actual law has been discussed or rulemaking is taking place); (3) pre-implementation (the major policy-making has concluded, but nothing is yet binding/implemented); (4) early implementation (some data sharing requirements are binding, e.g., bank-level product information, but not personal account/transactions); (5) mid-implementation (personal account/transaction data sharing is binding or OB infrastructure/technical standards have been put in place, but not all planned elements are in place); (6) fully implemented (full implementation as described in the law/rulemaking/policy documents); (7) follow-on regulation or policies (OB is implemented, and regulators are actively working on related policies, such as open finance or open data, or on implementing additional pieces of infrastructure for OB).

[17]The Consumer Financial Protection Bureau (CFPB) is looking into whether to create regulation based on Dodd-Frank's Section 1033 that gives consumers the right to their financial data, but which was never codified into rulemaking and, hence, not legally binding. See here.

[18]See here.

[19]This policy would broaden data access beyond transaction accounts. See here.

(b) shows the passage year of countries' major OB government policies.

Requirements Set by the Regulator OB government policies differ in what they require of market participants, and indeed, whether they require anything at all. The UK, for example, places explicit de jure legal requirements on banks to participate. Other examples with binding regulatory approaches are Australia, Bahrain, Brazil, the EU, and Israel. In contrast, regulators in Singapore, Malaysia, and Russia do not explicitly mandate data sharing and instead facilitate the adoption of OB by mediating industry discussion, providing technical standards for APIs, or providing infrastructure for data sharing.

As shown in Table 1, among the countries whose OB initiatives have advanced sufficiently for these issues to be decided, we find that 88% require banks to share data, while the other 12% do not. Additionally—and often—39% of countries' regulators lay out technical specifications for APIs while the remainder do not. There is significant regional variation in government-led approaches regarding mandatory data sharing and technical specifications. In particular, OB regimes in Europe & Central Asia tend to have mandatory data sharing (97%) but do not set technical specifications (15%). Conversely, OB regimes in East Asia are less likely to be mandatory (60%) but more likely to set technical specifications (82%). Figure 3 Panels (a) and (b) show these differences graphically for mandatory data sharing and regulator-set technical specifications, respectively.

Finally, in addition to requiring incumbent banks to share data, some OB regimes also require sharing by data users—non-bank financial intermediaries (e.g., fintechs). In other words, some regimes require sharing reciprocity while others do not. Our data show that only 18% of regimes have data sharing reciprocity, where fintechs that use the data must share. There is no data sharing reciprocity in Europe & Central Asia, while other regions, particularly East Asia, tend to require it.

Open Banking Scope: Covered Services and Functions OB government-led regimes differ dramatically in what financial products and services are covered. OB in its narrowest incarnation covers only transaction accounts: checking accounts, and occasionally credit cards. Some regimes include a broader set of core consumer finance products: savings accounts, investments, and loans. Still broader regimes, bordering on "open finance" as opposed to merely "open banking," cover financial services such as insurance or small business lending.

By definition, all OB regimes cover at least transaction accounts. Fewer—34%—additionally cover non-transaction accounts, while fewer still cover a broader set of products such as insurance or small business lending. Regarding regional heterogeneity, Europe & Central Asia OB laws tend to be very narrow in scope, with only 3% covering non-transaction accounts. In contrast, OB policies in other regions are much broader, with 90% going beyond transaction accounts.

Regarding functionality, OB APIs can, in theory, be used both to read data (e.g., pull customer account information) and to write data (e.g., initiate payments). Some OB regimes focus on data sharing only, and some on both. Our data show that among those countries where this issue has been decided, only 5% focus on data sharing only, none on payments only, and 95% on both.

Open Banking Strength Index Using our hand-collected data on OB policies, we construct an OB Strength Index, which averages the four key OB policy dimensions. These four dimensions reflect whether the regulators have set policies that (i) mandate banks to share data, (ii) require financial service providers (such as fintechs) who use data to share data in return, (iii) set an API standard, and (iv) cover a wide range of financial products. This index ranges from 0 (all four dimensions no or not yet decided) to 1 (yes on all four dimensions). We use this index to examine whether more comprehensive OB policies result in higher levels of OB adoption and more financial innovation.

2.4 Non-Open Banking Policy Data

Venture Capital Data Spurring innovation is often a key objective of OB policies; however, innovative output is notoriously hard to measure. We use data on VC investment into startups as a proxy for innovative entry, as past research has shown that VC-backed startups are generally innovative, fast-growing entrants (Puri and Zarutskie, 2012; Gornall and Strebulaev, 2015). This proxy is a forward-looking measure of profit-motivated investors' expectations, which helps us analyze the effects of still-recent policy interventions.

Using PitchBook data, widely acknowledged as one of the best VC data sources for more recent years, we construct a country-by-year panel of VC deals for the past twenty years, from 2000 to the first half of 2021. We measure VC activity using two standard variables: the number of deals and the investment amount in millions of US dollars. Our interest lies in financial innovation, so we split the deals in each country-year into fintech deals and non-fintech deals, with fintech deals being the deals PitchBook places in the "Financial Software" sub-industry or the "Fintech" vertical. We are interested in measuring the impact of OB on innovation around specific use cases, but Pitchbook lacks more granular industry classifications. We overcome this by using PitchBook's keywords feature to define seven sub-industries of fintech: alternative lending, consumer finance, financial IT, payments, regtech, wealth management, and digital assets. Details of our classification are in Appendix B. Because of the recent cryptocurrency boom and bust cycles and the fact that digital assets are not related to OB, we reclassify digital assets startups as non-fintech for our main analysis, although this has only a small impact on our results.

Bank API Data Bank API data are from Platformable, which is a global leader in data on OB APIs.[20] To our knowledge, these data provide the best global coverage of banks' APIs, and allow us to examine whether government efforts to encourage OB adoption are actually effective at opening up banks by leading them to introduce APIs.

Explanatory Variables We compile a variety of other country-by-year variables, which are summarized in Table C1. We start with basic country-level data, including per capita GDP in thousands of US dollars and population in millions from the World Bank. From the World Bank, we also add standard measures of country-level financial sector development, including the quantity of private sector credit to GDP, the number of bank branches per 100k people, and the financial sector's Lerner Index. The Lerner index measures markups over marginal costs, ranges between 0 and 1, and captures the market power of banks, with higher values denoting less competition. In addition to those measures, we take the percentage of banks that are foreign owned from Claessens and Van Horen (2013).

To capture the quality of institutions, we use several indexes. The Rule of Law and Business Regulation Indexes from the Cato Institute are on a 0 to 10 scale, with higher numbers denoting more favorable conditions. The Corruption Perception Index is from Transparency International and is on a 0 to 100 scale, with higher numbers denoting more favorable conditions.

2.5 Drivers of Open Banking Government Policies

An important preliminary question is what drives countries to adopt OB policies. In the spirit of Kroszner and Strahan (1999), we examine what ex-ante country characteristics predict OB policy adoption. We run the following cross-sectional, country-level regression:

$$OB_i = X_i'\beta + Region_r + \epsilon_i, \tag{1}$$

where OB_i is one of two types of OB outcomes in country i described in Section 2.3 above. First, we use a 0/1 indicator for whether the government has implemented OB policies in a country as of October 2021. Second, we use a number ranging from 0 to 7, indicating how far the implementation of government OB policy has progressed, with 0 denoting none and 7 denoting fully implemented with follow-on regulation. X_i' is a vector of ex-ante country-level characteristics as of 2013.[21] Data availability causes the number of observations to fluctuate across specifications. $Region_r$ is a region fixed effect, which allows us to exploit within-region

[20] Platformable collects industry data on OB and open finance by systematically identifying API providers and consumers using bank and fintech website sources, fintech registers such as EUCLID (EU) and FCA (UK), assessing API consumers and providers from fintech association membership lists, and by surfacing new initiatives from newsletters and industry alerts. Data are collected on a rolling basis, with each entity assessed at least once every three months.

[21] We choose 2013 both because it predates the earliest OB regimes and because it is the final year that comprehensive Lerner Index data are available from the World Bank.

variation.

Table 2 shows our explanatory variables do not robustly predict either our binary (columns 1–5) or continuous (column 6) OB policy measures. Column 1 considers three measures of financial development, none of which robustly predict government-led efforts to promote OB within a country. Column 2 shows that OB policies are somewhat more likely to be adopted in countries with more non-fintech VC deals, but that fintech VC deals, in particular, are not predictive of adoption. Fintech innovation not driving OB policies is comforting because it speaks against preexisting financial innovation driving both OB rules and fintech VC deals. In column 3, we find a weak and statistically insignificant association between the fraction of foreign-owned banks and the adoption of OB policies. In column 4 we consider several rankings of local institution quality and see no strong patterns. In columns 5 and 6, we include the full set of predictor variables: none of them significantly predict our OB implementation indicator (column 5), while the financial sector's Lerner Index and the number of non-fintech VC deals have a weakly significant correlation with our continuous variable for the implementation progress of OB policies (column 6).[22] Since low overall levels of economic development could be associated with the introduction of OB policies, in addition to our region fixed effects, in all columns we control for both GDP per capita (and its square) and log population to prevent this association from driving the results. However, across columns 1–6, neither a country's GDP nor its population robustly predicts the introduction of OB government policies. Taken together, these results suggest that there are not particularly strong political economy issues around OB adoption.

3 The Economic Effects of Open Banking

In this section, we examine the effects of OB. We set the stage by showing that OB policies are associated with much more bank data sharing (Section 3.1). We then present our main empirical result that OB has a casual impact on fintech VC investment (Section 3.2). Finally, we discuss other real outcomes (Section 3.3).

3.1 Government-led Open Banking and Incumbent Banks' Data Sharing

Government OB policies either force or encourage banks to allow other financial service providers to access their customer data upon customer request. The basic threshold question is whether banks indeed share their data following OB mandates to do so. Since APIs are the main technology for data sharing, we look at whether the presence of OB government policies

[22] As shown in Appendix Table C2, our 0–7 measure of OB implementation gives similar results for columns 1 through 4. Not reported, we also get similar results when fitting a Cox proportional hazards model or using the year of OB adoption as an outcome variable.

is associated with bank API offerings. Table 3 shows the results of a cross-sectional, cross-country regression of the prevalence of APIs in banks of each country against our measures of that country's OB implementation:

$$BankAPIs_i = \beta \times OB_i + X_i'\gamma + Region_r + \epsilon_i, \tag{2}$$

where $BankAPIs_i$ is the log-transformed number of banks with APIs (columns 1 to 3) or the percentage of the top 10 banks in each country that offer APIs (columns 4 to 6). OB_i is one of three types of OB outcomes. First, we use a 0/1 indicator for whether the government has already implemented OB policies in a country as of October 2021 (columns 1 and 4). Second, we use a continuous measure of how far the implementation of government OB policy has progressed, with 0 denoting none and 7 denoting fully implemented with follow-on regulation (columns 2 and 5). Third, we use the interaction between our 0/1 OB policy indicator and our 0 to 1 OB Strength Index (columns 3 and 6). $Region_r$ is a region fixed effect and X_i' is a vector of ex-ante basic economic country characteristics (GDP per capita and population).

There is a strong association between OB policies and bank API offerings. Column 1 shows that countries with OB policies have about twice as many banks offering APIs, with columns 2, 4, and 5 yielding qualitatively similar numbers. Columns 3 and 6 show that these effects are driven by more comprehensive OB policies. These results provide the first systematic evidence that government policies to promote OB might have already had a significant effect on data sharing in the financial service industry, and that counties that have more comprehensive OB policies (as measured by our OB Strength Index) are likely to see more data sharing. These results also suggest that banks are not voluntarily sharing data, a result consistent with the model we later present.

3.2 Open Banking Policies and Fintech Venture Capital Investment

Financial innovation is the most common goal of OB policies. Regulators hope that giving customers the right to share their financial data with new entrants will spark the creation of new firms that offer innovative financial products and services. The previous result shows that after OB adoption, banks provide the necessary technology for new entrants to access data. We now test whether this data access spurs innovative entry by using data on VC investments in fintechs and a standard panel event-study design:

$$FintechVC_{i,t} = \sum_{k \neq 0} \beta_k \times OBLag(k)_{i,k,t} + Country_i + Region_r \times Year_t + \epsilon_{i,t}, \tag{3}$$

where $FintechVC_{i,t}$ is a measure of fintech VC activity in country i and year t, measured as either the number of deals and the millions of US dollars invested.[23] $OBLag(k)_{i,k,t}$ is an event time indicator, equal to 1 if country i's adoption of OB government policy occurred k years from time t and zero otherwise.[24] We normalize the year of the policy's passage to zero so that the coefficient β_k measures changes in fintech VC activity k years before or after OB policy passage relative to the year of its passage. $Country_i$ and $Region_r \times Year_t$ are country and region-by-year fixed effects.

VC data pose two key challenges. First, VC activity is very skewed, with the US having far more VC investments than any other country. We correct for this using a $\log(1+x)$ transformation of our VC activity measures, which means our tests measure relative increases or decreases in VC activity occurring, rather than absolute changes. Second, the lack of central VC investment registries in most countries makes VC data challenging to collect. Appendix Table C3 summarizes our data and shows that PitchBook, despite being one of the best VC databases, has significant gaps in its international coverage. Due to a combination of data collection and low VC activity, only one-quarter of our post-2000 country-years have any fintech VC deals and more than half have no VC deals at all. To reduce the biases created by using log-transformed variables in the presence of zeros and VC data coverage issues, we restrict our consideration to countries with active PitchBook coverage. As our law passages occur in 2016 or later and PitchBook coverage improves over time, we restrict our analysis of VC activity to the 2011–2021 period. In addition, we consider only countries that PitchBook already covered before our regression sample period by focusing on countries with five or more fintech deals in the 2000-2010 pre-period, which we refer to as high-coverage countries.[25] Our focus on high coverage countries and our tests using VC dollars, which load on large and hard-to-miss deals, help attenuate concerns that PitchBook coverage improvements are correlated with the passage of OB government policies. Although only 13% of countries are high-coverage, they include 91% of the VC deals and 94% of the investment value. Thus, our analysis of OB policies on fintech VC activity uses the sample of high-coverage countries in the 2011-2021 period. 99% of these high-coverage country-years have at least one fintech deal, dramatically reducing the econometric issues associated with log-transforming zeros. Because we condition on pre-period deals, our results can best be thought of as speaking to countries that already have developed VC markets.[26] Because our filter drops a large

[23]The staged nature of VC investments means that deal counts tend to measure earlier stage investment and dollar amounts tend to measure later stage investment.

[24]For countries in the sample that never adopt OB, $OBLag(k)_{i,k,t}$ is zero everywhere; these countries help identify region-by-year fixed effects.

[25]Specifically, we consider Australia, Belgium, Brazil, Canada, China, Germany, Denmark, Finland, France, India, Ireland, Israel, Japan, the Netherlands, Norway, Poland, Russia, Spain, Sweden, the United Kingdom, and the United States of America.

[26]The results in Table 4 continue to hold with similar coefficients for the entire sample of countries; however, the large number of zeros makes it hard to interpret the results.

number of country-years that *never* had OB, identification in this specification comes chiefly (though not entirely) through the staggered adoption of OB within countries. Intuitively, our regression is comparing VC activity in countries at time t to other countries in the region that will adopt OB but have not adopted it yet. The key identifying assumption is that absent the treatment, countries within a region would have been on parallel trends.

Figure 4 presents the results from the event-study specification in Equation 3 and shows a relative absence of pre-trends in fintech VC activity, followed by a sharp increase after treatment with an OB policy. This pattern holds for the number of deals (Panel (a)) and the amount invested (Panel (b)). In both panels, there is a clear inflection point around the year of the OB policy passage and a change of large economic magnitude: log deals increases by about half and log dollars about doubles. The absence of pre-trends is consistent with the parallel trends assumption and OB having a causal impact on country-level fintech VC activity. As a robustness check, in Appendix Figure C1 we control for contemporaneous non-fintech VC deals as a proxy for innovation more generally and see the same pattern. This addresses a potential concern that OB adopters enacted broader innovation promoting policies.[27]

Table 4 uses a difference-in-difference design to examine the relationship between OB policies and fintech VC activity:

$$FintechVC_{i,t} = \beta \times OB_{i,t} + Country_i + Region_r \times Year_t + \epsilon_{i,t}, \tag{4}$$

where $OB_{i,t}$ is a dummy variable equal to one if OB was adopted in country i before year t and other variables are as in Equation (3). We are interested in the coefficient β which measures log change in fintech VC activity following the introduction of government OB efforts.

Fintech companies receive significantly more VC investment following the adoption of OB policies, whether measured by the number of deals or the dollars invested. Our coefficients are both statistically significant and large in economic magnitude. Using our preferred specification from Equation 4, we find a 0.53 increase in log fintech VC deals (column 2 of Table 4) and a 1.3 increase in log fintech VC dollars (column 5). We rerun these tests using contemporaneous non-fintech VC deals as an additional control and find that the coefficients remain statistically significant and economically large (columns 3 and 6). Our coefficients are also significant if we use year fixed effects rather than region-by-year fixed effects (columns 1 and 2).

As an additional robustness test, in Appendix Table C4 we show that no single country

[27] As a caveat to this test, it is debatable whether it is appropriate to include the contemporaneous control for non-fintech VC deals since OB innovation may spill over into other, non-financial sectors of the economy. For example, anecdotally, small and medium-sized enterprises (SMEs) are somewhat unexpected beneficiaries of OB, benefitting from OB applications such as accounting management and financial planning.

drives our effect by running a leave-one-out test of our core specification (column 2) that excludes each country in turn, and Germany and France together (the two countries powerful enough to have an impact on the passage of OB government policy in the EU). Our coefficients are generally stable across these varying samples, suggesting our results are a general phenomenon. In Appendix Table C5 we show that the results in Table 4 are robust to using an inverse hyperbolic sine transformation of the dependent variable (an alternative to our main log-transformation), alleviating concerns associated with log-transformation in the presence of zeros. Finally, in Appendix Table C6 we show that OB has no effect on non-fintech VC deals when measured using an analogous specification to Table 4; a test that helps address the concern that country-level trends in innovation drive both OB and VC.

Fintech Investment by Product Area OB might have a larger impact on some areas of finance than others. We test this in Table 5 by looking at the impact of OB on specific product areas within fintech. Our empirical design follows Equation 4, but with a dependent variable based on VC investment only in companies targeting a specific fintech use case. As discussed in Appendix B, we define these fintech categories based on Pitchbook's fintech industry map. Across product areas and specifications, we see economically large increases in log deal counts, ranging from 0.62 to 0.76. The notable and reassuring exception to this trend is digital assets, where we see insignificant and economically small effects. This is intuitive and serves as a placebo test: digital assets, such as cryptocurrency, are largely unrelated to OB functionality.

Specifically in Panel (a) of Table 5, we see similar magnitudes and statistically significant results across the categories that are relevant for OB: alternative lending, which would presumably lever newly available data to make lending decisions; consumer finance, which is the focus of most OB reforms; financial IT, the backend infrastructure necessary to utilize OB functionality; regtech, which could follow from either novel OB applications or increased compliance demands; and wealth management, which is greatly streamlined through OB account access. Payments, which is one of the key focuses of OB, shows a similar increase in deals, although it is not statistically significant due to a less precisely estimated parameter. Controlling for the non-fintech VC deal volume in Panel (b) has little impact on these results. Although coefficients decrease slightly across the board, all but wealth management retain statistical significance, while the increase in payments gains statistical significance. Overall, the most statistically robust result is on financial IT fintechs—this is intuitive because the infrastructure to enable bank data sharing between banks and other financial service providers is a necessary stepping stone for financial service providers utilizing OB functionality to build their products.

Open Banking Policy Choices and Fintech Investment Given the diversity of OB policy choices described in Section 2.3, it is natural to ask whether the specifics of OB policies matter. In particular, we examine whether each of the four main OB policy characteristics (discussed in Section 2.3) is associated with greater fintech VC activity. We examine these effects by rerunning Equation 4 with OB adoption interacted with policy choice as the variable of interest:

$$\begin{aligned} FintechVC_{i,t} = & \beta_1 \times OB_{i,t} + \beta_2 \times PolicyDimension_i \times OB_{i,t} \\ & + Country_i + Region_r \times Year_t + \epsilon_{i,t} \end{aligned} \tag{5}$$

where $PolicyDimension_i$ measures a dimension of policy heterogeneity and $PolicyDimension_i \times OB_{i,t}$ is that policy dimension interacted with OB being adopted in the country. The country fixed effects absorb level differences between countries with and without these OB policy characteristics. Countries that do not have OB policies or where the policy choice has yet to be decided are coded as zeros. Other variables are as defined above.

Table 6 shows that more comprehensive OB policies, as measured by our OB Strength Index (column 5), are associated with larger increases in fintech VC deals. In terms of specific policy choices, countries that require banks to share data see economically and statistically significantly increased fintech effects following the introduction of OB policies (column 1). A similar larger effect is seen for countries that require reciprocal data sharing, where non-bank data users such as fintechs must themselves share their customer data with, say, banks when customers request (column 2). Standardized APIs (column 3) and broader financial product coverage (column 4) have positive and economically large coefficients that fail to reach statistical significance.

Overall, more comprehensive OB government policies appear to have larger impacts on the entry of innovative financial firms. OB policies where banks or data users must share data seem to be a particularly important policy dimension for increased fintech entry—in fact, consistent with banks being reluctant to share data, policies that do not obligate banks to share have no significant effect. The model that we present in Section 4 will show that large increases in innovative entry are a natural implication of open data policies.

3.3 Direct Measures of Competition and Financial Inclusion

We show there is a significant increase in fintech investment following the adoption of OB policies, which suggests both innovation and future competition. We test for direct effects on product market competition and financial inclusion using a variety of measures. Specifically, we examine IMF data on transaction volumes (internet banking, mobile money, bank borrowing, total deposits), and accounts (deposit, credit card); World Bank data on remittance costs and bank competition (concentration and profitability); and CRSP data on bank share prices.[28] Unfortunately, we are not able to effectively test for OB effects using these outcome measures. The OB policies have passed relatively recently, which creates two hurdles. First, there are significant time lags involved with cross-country data, which makes it difficult to test the effect of policies that commonly took effect in the last few years. Second, many financial inclusion and competition outcomes, such as remittance costs or the fraction of the population with bank accounts, depend on slow-moving consumer tastes. VC deals are a forward-looking measure of competition and innovation: although we are unable to detect short-term changes in these additional outcomes, increased VC investments suggest that investors expect long-term changes. Further, as we show in the following section, our modeling suggests OB has ambiguous implications for outcomes other than entry.

4 An Economic Framework for Open Banking

In this section, we introduce a model to examine how wider access to bank customers' data—OB—affects entry, competition, and consumer welfare. The model provides a general-purpose theoretical and quantitative framework to study data use in a competitive context. We first frame our empirical results with a straightforward calibration in the context of two financial products using off-the-shelf estimates from the literature. Consistent with our empirical results, the model predicts that OB increases entry and fintech innovation. Next, because we are still in the early years of OB regimes and detailed outcome data are not yet available, we use the model to evaluate OB's longer-term welfare and distributional consequences. We find that while OB leads to increased entry and innovation, it has ambiguous effects on other outcomes, including financial inclusion and data production. Our model speaks to three main issues.

Innovation, entry, and unequal data access: Financial innovation and fintech entry are the most frequently cited motivations for OB policies. Therefore, departing from other models of OB, e.g., He et al. (2020) and Parlour et al. (2020), we explicitly model new entry on the extensive margin and product improvements on the intensive margin. This allows us to pinpoint how data access enables innovation. For example, we can ask whether entry is driven by innovative uses of data that directly increase consumer welfare, or simply by new entrants avoiding being adversely selected against by better-informed incumbents (e.g., Di Maggio and Yao (2021)).

Data use and data production: Our model provides a general framework that nests many uses of consumer heterogeneity data. Most models of data focus on their value at the aggregate level (e.g., improving output quality (Jones and Tonetti, 2020) or business practices (Farboodi et al., 2019)). We depart from these models by having data reveal consumer

[28] We forgo international patent data due to data lags.

heterogeneity. Such heterogeneity is especially important in the finance context because there are large variations in consumer financial situations and needs. Beyond providing a framework for analyzing data *use*, we endogenize data *production*. In the finance context, consumer data are typically generated through repeated interactions with service providers, e.g., a history of financial transactions with a bank. Our framework recognizes how providing a product that produces data can facilitate the provision of a different product. Modeling this lets us evaluate how allowing financial service providers to monopolize the customer data they generate impacts both data-producing and data-using products.

Distributional consequences: Finally, our focus on consumer heterogeneity lends itself to an analysis of distributional consequences. We quantify how open data policies help some consumers while hurting others. For instance, when data is used for screening or price discrimination, the effect of OB is that while borrowers with low default probabilities benefit from the widespread dissemination of their data, costlier borrowers are hurt. In contrast, when the data is used to provide better products, all consumers benefit. Thus, policymakers concerned with financial inclusion—particularly among higher-risk populations—face the risk of undermining redistributive goals depending on how the data is used.

In summary, our model highlights a fundamental tradeoff of wider access to consumer financial data: it increases entry, innovation, and competition, but reduces some consumers' access to *data-using* products like loans and *data-producing* products like transaction accounts. These negative effects arise through two channels. First, opening data removes ex-post data monopolies, which reduces firms' ex-ante incentives to generate data. This increases the price of the data-producing products, leading to fewer consumers using them. Second, depending on how it is used, opening data may lead to higher prices for customers whose data inform firms that they are costlier to serve or more willing to pay. These customers pay more even if they are given the choice to opt out of data sharing because opting out is taken as a negative signal. Charging some customers more is part of how opening data increases entry. Informed incumbents always use customer data to identify high-cost customers and charge them more. Without OB, those customers switch to entrants who lack information and thus undercharge them, which creates entry-deterring adverse selection. With OB, entrants also charge the high-cost customers more, hurting those customers but eliminating adverse selection and increasing entry and competition. These channels create an inherent conflict between OB's main goals of increased competition and innovation, and promoting financial inclusion.

4.1 Model

The model extends a standard discrete choice framework by explicitly considering data production and usage. Data allow firms to improve their products or pricing by learning about the characteristics of heterogeneous consumers. The model has two periods. In the first period, a fixed number of firms compete to provide a *data-producing* product to consumers. In the second period, the first-period incumbents and new entrants compete to provide a *data-using* product to the same consumers. Customer-level data are useful for providing data-using products. For example, the pricing of a loan (data-using product) is improved using data from a transaction account (data-producing product) that reveal the customer's credit risk, as shown by Ghosh et al. (2021). Alternatively, an automated financial planner (data-using product) uses balances from financial accounts (data-producing products) to offer customized financial advice.

The two periods connect through whether—and which—firms can access consumer data. Under relationship banking, only the firm that provided a customer with the data-producing product can use her data. Under OB, all firms marketing the data-using product to a customer can access her data, regardless of whether they produced that data. In order to focus on the key tradeoffs highlighted above, our main model does not allow customers to opt out of this data sharing. In Appendix D.1, we show that allowing consumers to opt out of OB does not change our main takeaways, as opting out of data sharing partially "unravels" (Grossman, 1981).[29]

Figure 5 outlines the modeling framework, with Panel (a) showing the relationship banking environment, and Panel (b) showing the OB environment. Relationship firms always set customer-specific prices and product offerings, while non-relationship firms can only offer those in the OB regime. We now present the model in reverse-chronological order, beginning with the data-using period.

4.1.1 Period Two: Data-using Period

Consumer data and market structure: A mass m of heterogeneous consumers, indexed by i, can consume a data-using product. Each consumer is endowed with a vector of characteristics for this period, χ_i, whose distribution $dF(\chi_i)$ is known to the customer and all firms.

A customer produced data if she used a product in the (previous) data-producing period. Access to these data allows firms to learn the customer's specific realization of χ_i. The policy regime determines which firms can access her data. Under a relationship banking regime, the firm that provided her the data-producing product learns χ_i and all other firms know only the distribution. This represents the status quo for banked customers, where the firm providing

[29] Allowing customers to choose whether to share data only partially mitigates the welfare costs for the worst customers (i.e., those with higher marginal costs or lower price sensitivity). Customers whose information allows them to get a better price choose to share their data, which leads firms to charge high prices to consumers who do not. Beyond that, a large empirical literature has documented near-universal ineffectiveness in electronic privacy or contract disclosures, e.g., Ben-Shahar and Schneider (2011).

the data-producing product is informed but does not share data with competing firms. Under an OB regime, all firms have access to her data and learn χ_i. This could arise either due to voluntary sharing by the data-producing bank or by regulatory fiat. If the customer did not use the product in the first period, no data was produced and so no firm observes χ_i. To fix ideas, we assume a consumer's data provide information on her willingness to pay (α_i), product customization needs (f_i), and consumer-specific marginal cost (mc_i):

$$\chi_i \equiv (\alpha_i, f_i, mc_i)\,. \tag{6}$$

These proxy for key uses of data: willingness to pay covers pricing (what interest rate?) and marketing (do we offer a credit card or a mortgage?), customization needs covers product tailoring (how can we set up a financial plan?), and marginal costs covers both usage (will they exploit credit card bonuses?) and risk (will they default?).

Products in the data-using period are offered by I incumbents and an endogenous number, N, of new entrants. All firms offer products to all customers, who choose a single product out of the available offerings.

Consumer demand: Customer i makes a discrete choice of firm j's product from among the $I+N$ competing firms.[30] Product ij is characterized by $\nu_{ij} \equiv (p_{ij}, g_{ij})$, where p_{ij} is price and g_{ij} are non-price characteristics. These non-price characteristics could be beyond the firm's control (e.g., a customer's preference for traditional banks over fintechs) or under the firm's control (e.g., whether the firm had a relationship with customer i in the prior period). Consumer i receives the following indirect utility from product ij:

$$u(\nu_{ij}, \chi_i) \equiv -\alpha_i p_{ij} + \delta(g_{ij}, f_i) + \epsilon_{ij}. \tag{7}$$

Here, α_i is the consumer's price sensitivity and p_{ij} is the price. $\delta(g_{ij}, f_i)$ is the value the customer gets from the product as a function of non-price product characteristics. ϵ_{ij} is a horizontal taste shock whose iid realization is known to the consumer but unknown to firms, which creates differentiation and gives firms market power.

Among the offerings and an outside option, u_0, the consumer chooses the product which offers the highest indirect utility. Let $s_j(\boldsymbol{\nu}_i, \chi_i)$ denote the probability that a customer with characteristics χ_i chooses firm j's product given all product offerings, $\boldsymbol{\nu}_i$. This quantity is obtained by integrating across the taste shock, $\boldsymbol{\epsilon_i}$:

$$s_j(\boldsymbol{\nu}_i, \chi_i) = \int \mathbb{I}\left\{u(\nu_{ij}, \chi_i) > u(\nu_{ik}, \chi_i),\ \forall k \neq j\right\} dF(\boldsymbol{\epsilon_i}). \tag{8}$$

Firms: Conditional on entry, firms compete in a differentiated Bertrand structure. Firm

[30] We assume each firm can only offer one product to each consumer in each period; however, the model could be extended to multi-product firms.

j's marginal cost for customer i is equal to the sum of mc_j, a firm-specific cost common to all of j's potential customers, and mc_i, a customer-specific cost that is common to all firms selling to customer i:

$$mc_{ij} \equiv mc_j + mc_i. \tag{9}$$

Informed firms observe the customer's characteristics, χ_i. Based on those, they set consumer-specific prices and product characteristics, ν_{ij}. Uninformed firms only observe whether the customer used the data-producing product in the prior period. Because of that, they offer a single price and product to all customers that used the data-producing product and another price and product to all customers that did not. Products and prices are set to maximize period two profit from that customer, Π_{ij}:

$$\Pi_{ij} = \begin{cases} \max_{\nu_{ij}} s_j(\boldsymbol{\nu}_i, \chi_i)(p_{ij} - mc_{ij}) & \text{if firm } j \text{ has access to customer } i\text{'s data} \\ \max_{\nu_j} \int s_j(\boldsymbol{\nu}_i, \chi_i)(p_j - mc_{ij}) dF(\chi_i) & \text{otherwise.} \end{cases} \tag{10}$$

Each firm is in one of four data environments for each customer. First, if the customer did not consume the product in the first period, there is no data. Second, if there is OB and the customer used a product in the first period, all firms have access to her data. Finally, if there is no OB and the customer consumed a product in the first period, there are two data environments: one for the firm that supplied her the product and now enjoys a data monopoly, and another for other firms that lack access to her data and compete with that firm. Firm profit varies across these data environments.

Each firm's period two profit is equal to its profit across all customers and their associated data environments, minus an entry cost c for new entrants.

$$\Pi_j = \int_i \Pi_{ij} di - c. \tag{11}$$

Entry at cost c in the second period implies that $\Pi_j = c$ for second-period entrants.

Consumer utility and product offerings: A consumer's ex-ante expected utility in the second period (before the realization of her ϵ_{ij} taste shocks) depends on the product offerings she faces:

$$Eu(\boldsymbol{\nu}_i) = \int \max_j \{u(\nu_{ij}, \chi_i)\} \, dF(\boldsymbol{\epsilon_i}). \tag{12}$$

These product offerings in turn depend on the information environment. As above, firms adjust their prices and product offerings based on both their own data and whether other firms have those data. We use Δu to denote the change in expected period-two consumer utility caused by generating data.

4.1.2 Period One: Data-producing Period

The data-producing period has a similar market structure to the data-using period. For notational convenience, we superscript variables in the data-producing period with p. I firms indexed by j compete to offer a financial product to consumers indexed by i. We assume that customers are homogeneous at this time, apart from horizontal taste shocks, so that consumer i's indirect utility from choosing product j is:

$$u_{ij}^p = -\alpha^p p_j^p + \beta \Delta u + \delta_j^p + \epsilon_{ij}^p. \tag{13}$$

Here, α^p is price sensitivity and p^p is the price. Δu is the extent to which producing data changes the customer's expected indirect utility in the next period (which does not depend on the as-yet-unrealized second-period characteristics χ_i) and β is the extent to which the consumer weighs this future utility gain, with values below 1 reflecting myopia or impatience. δ_j^p is the non-price product characteristics associated with the firm and ϵ_{ij}^p is an iid horizontal taste shock, known to the consumer but not to the firm.

Let $s_j^p(\mathbf{p})$ denote the expected market share of firm j, obtained by integrating across taste shocks ϵ_{ij}^p:

$$s_j^p(\mathbf{p}) = \int \mathbb{I}\left\{ u_{ij}^p > u_{ik}^p,\ \forall k \neq j \right\} dF(\boldsymbol{\epsilon_i^p}). \tag{14}$$

Further, let $\Delta\Pi$ denote the extent that possessing consumer data increases a firm's profit in the second period, based on the change in profit across different information environments (Equation (10)). Firm j internalizes this benefit and sets a price p_j^p to maximize total expected profit across both periods, taking other firms' prices as given:

$$\Pi_j^p = \max_{p_j^p} s_j^p(\boldsymbol{p^p}, \boldsymbol{\delta^p})(p_j^p - mc^p + \Delta\Pi), \tag{15}$$

where mc^p is the marginal cost of providing the data-producing product.

4.1.3 Equilibria

We focus on two types of equilibria: short-run and long-run. In both cases, we restrict our attention to symmetric equilibria where all informed firms charge the same consumer-specific price and all uninformed firms charge the same price to customers in a given data environment.

Short-run equilibrium: The short-run analysis holds ex-ante data production fixed and focuses only on the data-using period. Intuitively, customers have already formed data-producing relationships whose terms will not adjust immediately, and we examine outcomes in the data-using market when the data produced through these relationships are made available through OB. Conditional on the mass of consumers who formed data-producing relationships in the previous period, a short-run equilibrium consists of (i) the number of new entrants N, (ii) data-using period prices and product characteristics $\boldsymbol{\nu_i}$, and (iii) data-using period consumer demand $s_j(\boldsymbol{\nu}_i, \chi_i)$. Firm entry follows from the zero-profit condition applied to Equation (11), market shares as a function of prices and entry follow from Equation (8), and prices follow from the optimal pricing conditions arising from Equations (10).

Long-run equilibrium: Our long-run equilibrium allows the data-producing period prices and quantities to adjust, representing long-run changes in the banking market. This allows OB policies to impact ex-ante data acquisition. A long-run equilibrium consists of the same components (i) to (iii) as a short-run equilibrium and additionally (iv) data-producing period prices $\boldsymbol{p}^p$ and (v) data-producing period consumer demand $s_j(\boldsymbol{\nu}_i, \chi_i)$. The data-using period market share (Equation (14)) and pricing (Equation (15)) equations determine these. Consumer choices in the data-producing period impact the data-generating period. A measure m^{banked} of the mass m of consumers use a data-producing product and thus face a data-using product market where a single firm is informed (relationship banking) or all firms are informed (OB). A measure $m - m^{banked}$ choose the outside option of not using a product in the data-producing period and thus generate no data.

4.2 Calibration and Discussion

We breathe life into the model using simple calibrations based on two data-using products: non-GSE residential mortgages and financial advice. As detailed in Appendix D.2, the key objects for calibration are the distributions governing consumer heterogeneity, which we take from Buchak et al. (2018a) for mortgages and Di Maggio et al. (2021) for financial advice. The calibrations focus on heterogeneity in marginal costs and willingness to pay for mortgages, and heterogeneity in product customization needs for financial advice.

4.2.1 Data-Using Product Markets and the Information Environment

We first focus on the short-run effects of OB. OB introduction, in the context of the model, corresponds to shifting those customers with relationship banks from the relationship banking regime, where only the relationship bank observes their characteristics, to the OB regime, where all potential entrants observe their characteristics. This short-run analysis, which holds the amount of data in the economy fixed, corresponds most closely to our empirical results.

Figure 6 compares short-run equilibrium outcomes under OB to those under relationship banking for mortgages (magenta) and financial advice (cyan). Our calibration shows that OB dramatically increases firm entry (N) in both product markets, consistent with the large fintech VC increases we find in our empirical analysis (Table 4). Data-using product volumes also increase. Overall consumer welfare increases, while the profit of incumbent banks decreases. On net, total welfare (consumer welfare plus firm profits) increases. The welfare gains are much larger in the financial advice context because, as will be described in detail below, the data in the financial advice context is used to improve product quality for all consumer types. In contrast, in the lending context, the data is used for price discrimination, which benefits some consumers while harming others and offsetting these welfare gains.

4.2.2 Explaining the Impact of Information Using Consumer Heterogeneity

These aggregate effects are largely driven by how data access changes the nature of product pricing and customization across the spectrum of consumer heterogeneity. Although we focus on OB, these mechanisms apply to open data more generally. Figure 7 shows outcomes for consumers of different characteristics across the information environments in the short run. The lines correspond to the three short-run equilibria in which either all consumers are banked under relationship banking (green), all consumers are banked under OB (blue), or all consumers are unbanked (red). The first row of plots focuses on the cross-section of customer marginal costs in the mortgage case, the second row on consumer willingness to pay in the mortgage case, and the third row on consumer desire for customization in the financial advice case. The first column shows total product volumes, the second column shows prices, and the third column shows the portion of customers that switch away from their relationship bank to new entrants or non-relationship incumbents (outsider share).

Beginning with mortgages and marginal costs, Panels (a)–(c) show that with no information, quantities, prices, and outsider share are uniform across borrower types because borrowers are indistinguishable to lenders. Introducing relationship banking allows the relationship bank to set prices (i.e., mortgage interest rates) conditional on customer marginal cost (which reflects default probability), while uninformed competitors must offer a single pooling price. The relationship bank thus offers low-cost borrowers lower prices and high-cost borrowers higher prices. Low-cost borrowers borrow more at lower prices and the relationship bank also profits because its information monopoly gives it market power. This greater dispersion in bank interest rates is consistent with the reduced-form evidence from the US mortgage market presented in Figure 1. In the mortgage context, relationship banks, which possess more customer-specific data, are more likely to use non-standardized underwriting models to make credit decisions (Panel (a)). As a result of using more informative data in setting prices, residualized interest rate dispersion is greater (Panel (b)). These reduced-form facts align exactly with our model's prediction.

Further, this is consistent with Di Maggio and Yao (2021): In our model, informed relationship banks cream skim low-cost borrowers through lower prices and uninformed entrants face adverse selection from the remaining high-cost borrowers, who take advantage of their relatively favorable pooling offers. OB eliminates this adverse selection by allowing non-relationship lenders to access consumer data and condition prices on customer marginal cost. This leads relationship and non-relationship lenders to offer the same low prices to low marginal cost borrowers and the same high prices to high marginal cost borrowers. New entrants are no longer adversely selected, which increases entry and competition. Offsetting these benefits, incumbents' profits decline and high marginal cost consumers are harmed by broken pooling.

Panels (d)–(f) consider consumer willingness to pay in the mortgage setting, which we measure as $1/\alpha_i$. We see the same pattern of relationship banking reducing entrant profitability and thus entry. Across all data environments, consumers with a greater willingness to pay borrow in greater quantities. However, as more data enters the economy—first through relationship banking, then through OB—there is progressively more lending to low willingness to pay borrowers and less lending to high willingness to pay borrowers. This is mediated through prices: with no data, pricing is uniform, but with progressively more data, lenders charge high (low) willingness to pay borrowers more (less). Under relationship banking, new entrants charge a uniform price while informed relationship lenders charge low willingness to pay borrowers more and so entrants overwhelmingly serve high willingness to pay borrowers and lose low willingness to pay borrowers.[31]

The customer heterogeneity results for marginal costs and willingness to pay highlight two different consequences of OB. When consumer data are informative about marginal costs, OB increases entry by reducing the adverse selection faced by new entrants. This increases competition at a cost to high-cost borrowers. If high-cost borrowers mainly represent vulnerable sub-populations (e.g., low-income households), then OB can undermine regulators' goals of using OB policies to promote financial inclusion.[32] In contrast, when consumer data are informative about willingness to pay, OB primarily facilitates price discrimination. Lenders charge more to the borrowers with a high willingness to pay, which on the margin reduces the quantity of credit provided to these especially-eager-to-borrow individuals, reducing their utility. Again, if especially-eager-to-borrow individuals are mainly from vulnerable sub-populations, OB policies could actually undermine financial inclusion.

Third, we examine the financial advice product. Panels (g)–(i) plot outcomes versus how much customization the consumer needs, which we measure as $|f_i - \bar{f}|$. Customers with f_i farther from the mean require more customized advice. For product volumes, customers requiring more customization are worse served in the no information and relationship banking scenarios. Under OB, all customers are equally well served. For prices, when no financial firms have customer data, customers face a relatively low uniform price but get an inferior,

[31] This is consistent with results in Buchak et al. (2018b), who find that fintech mortgage lenders charge higher rates and appear to serve borrowers with a higher willingness to pay for convenience.

[32] Alternatively, if better data enable firms to exclude borrowers who are high cost because they are fraudulent, this may be beneficial.

uncustomized product. In the relationship banking case, customers who require the most customization pay the highest prices, because the relationship bank is the only one that can offer a customized product and so it benefits from significant market power. In the OB case when customer information is more widely available to firms, pricing is uniform but higher (along with product quality), as firms now provide optimally customized advice for which they charge more.

Observe that unlike the mortgage case, where data being used for price discrimination leave some consumers better off and others worse off, data in the financial advice case are used in such a way to make the product better for all consumers. Because these alternate uses of data lead to dramatically different distributional outcomes in terms of welfare, it is useful to examine which is likely to be quantitatively relevant in practice. Returning to the reduced-form results in Table 5, there appears to be evidence that both price discrimination and product improvement uses are important: On the price discrimination side, there are significant increases in the alternative lending and consumer finance product areas, while on the improved product side, there are significant increases in wealth management.

4.2.3 Quantity and Cost of Information Revelation on Firm Entry

In Figure 8, we examine how the amount of customer information revealed by OB changes firm entry in the short-run equilibrium. The model captures this comparative static through the unconditional dispersion in customer characteristics, σ_{mc}, σ_{α}, or σ_f, which is the amount of unobserved heterogeneity that consumer data access eliminates. For example, low dispersion may correspond to a lax OB regime that mandates little data sharing,[33] while high dispersion may correspond to a strong OB regime that mandates exhaustive data sharing across multiple products. We vary the dispersion of customer marginal cost, σ_{mc}, in the mortgage context (Panel (a)), and product customization needs, σ_f, in the financial advice context (Panel (b)). More informative data cause OB to have larger effects on entry, consistent with our reduced-form results in Table 6 showing that OB implementations mandating greater data sharing (e.g., when banks must share, and when OB covers products beyond transaction accounts) lead to greater entry.

In Panel (c), we examine how fintech entry changes following OB, depending on the magnitude of entry costs. Varying this model parameter allows us to speak to policy choices that make it less costly for fintechs to use banks' customer data. For example, a low entry cost might correspond to OB policies that set technical standards for bank APIs: instead of each bank using its specific technology for data sharing and each fintech needing to customize its technology to each bank, making all banks use the same data-sharing technology lowers fixed

[33] Or to a pre-OB information environment where a credit registry already revealed most of the useful information for the particular application.

startup cost to access these data. A medium cost might correspond to no technical standards and fintechs integrating with each bank's APIs separately. A high cost might correspond to the relationship banking case, where fintechs must use screen scraping or write one-off bilateral contracts with banks to access customer data. Unambiguously, and unsurprisingly, higher entry costs in the OB regime depress its effect on entry.

4.2.4 Ex-ante Incentives for Data Production and Consumer Myopia

The previous sections focused on the short-run equilibrium and ignored the fact that open data could change the supply and demand of the data-producing product (e.g., transaction accounts) in the long run. We now focus on that choice by considering the long-run equilibrium. An important consideration is whether consumers anticipate their transactions data having value for data-using products (e.g., mortgages or financial advice), which is captured by parameter β in Equation (13). We consider both fully rational customers (cyan), who consider their data-using period utility when making data-producing product decisions, and myopic consumers (magenta), who ignore it. In Figure 9, we show how moving from relationship banking to OB in our mortgage calibration case impacts the use of the data-producing product, incumbents' profit over the two periods, consumer welfare, and total welfare.

Eliminating banks' data monopolies allows customers to capture more of the value of their data in the data-use period. However, by increasing ex-post competition it reduces the value of customer relationships to banks (Figure 6) and banks' ex-ante desire to form relationships, similar to Petersen and Rajan (1995) or Boot and Thakor (2000). Myopic customers do not anticipate the value that their data will bring them, so OB reduces both data-producing product volumes and incumbent bank profit. Rational customers anticipate the value that their data will bring, so OB makes them more willing to pay for data-producing products. In fact, OB increases the extent to which customers value their data so much that incumbent banks could theoretically benefit from voluntarily adopting OB. The large effect of OB policies on bank API offerings (Table 3) and the fact that forced data sharing by banks drives our fintech VC results (Table 6) suggest that not all banks have been sharing data voluntarily, consistent with consumer irrationality or other frictions that prevent banks from capturing the consumer value created by opening data.

5 Conclusion

Our paper examines the dramatic rise of OB, which is now present in some form in roughly 80 countries. Using a hand-collected dataset of OB government policies around the world, we show that OB policies lead incumbent banks to set up technologies to share their customer data. Innovative fintechs appear to value these consumer data and raise significantly more money from venture capitalists following OB policies. We document significant heterogeneity in these policies' timing, purpose, and implementation, and that this heterogeneity is important, with large increases in fintech activity following strong OB implementations and weak regimes that do not force banks to share data having no effect.

We interpret these results by creating a general framework to model data use and sharing, which we calibrate to two financial product markets. Our model shows that OB increases entry because data access reduces adverse selection against entrants and increases entrant's product quality. Although our results suggest OB is achieving its innovation-promotion goals, our framework highlights two areas where policymakers should be concerned. First, OB reduces the value banks' capture from their data which reduces their ex-ante incentive to produce that data. Second, information sharing hurts customers whose data suggest that they are costly or not price sensitive. Even consumers who opt out of sharing are potentially harmed, as opting out sends a negative signal to banks and fintechs. These effects can be widespread and unpredictable. For example, OB data is increasingly used to screen potential renters via the screening service Tink, and customers who are unwilling to share their data risk being cut out of basic housing markets.

As policymakers set the path of future banking regulation, our paper helps put these tradeoffs in context. Data lie at the heart of relationship banking and large financial institutions benefit from their special ability to aggregate huge amounts of consumer data. Because of that, removing banks' monopoly on customer data has the potential to transform the very nature of relationship banking. If opening data pares back banks' economies of scope, the entire banking ecosystem could reorganize around more specialized and interconnected firms. The large reaction of fintech investment to OB shows the potential for disruption and just how valuable innovators perceive these data to be.

More generally, the role that data ownership and access play in endogenously creating and maintaining market power is a first-order question in an increasingly data-driven economy, sectors of which are dominated by a small number of data-intensive firms. Opening data to potential competitors and innovators in order to spur innovation, increase competition, and ultimately, raise welfare is a natural policy response, and our paper is the first to provide a global comparative analysis of such policy initiatives.

References

Acs, Z., T. Åstebro, D. Audretsch, and D. T. Robinson (2016). Public policy to promote entrepreneurship: A call to arms. *Small Business Economics 47*(1), 35–51.

Aguirregabiria, V., R. Clark, and H. Wang (2019). The geographic flow of bank funding and access to credit: Branch networks, local synergies, and competition.

Ahnert, T. and S. Doerr (2021). Does IT help? Information technology in banking and entrepreneurship. *IMF Working Papers 2021*(214).

Akcigit, U. and W. R. Kerr (2018). Growth through heterogeneous innovations. *Journal of Political Economy 126*(4), 1374–1443.

Allen, F., X. Gu, and J. Jagtiani (2020). A survey of fintech research and policy discussion.

Babina, T., A. Fedyk, A. He, and J. Hodson (2021). Artificial intelligence, firm growth, and product innovation. *Firm Growth, and Product Innovation (November 9, 2021)*.

Babina, T., A. X. He, S. T. Howell, E. R. Perlman, and J. Staudt (2020). The color of money: Federal vs. industry funding of university research.

Babina, T. and S. T. Howell (2018). Entrepreneurial spillovers from corporate R&D.

Bai, J., S. Bernstein, A. Dev, and J. Lerner (2021). Public entrepreneurial finance around the globe.

Barth, J. R., G. Caprio Jr, and R. Levine (2004). Bank regulation and supervision: What works best? *Journal of Financial Intermediation 13*(2), 205–248.

Bartlett, R., A. Morse, R. Stanton, and N. Wallace (2022). Consumer-lending discrimination in the fintech era. *Journal of Financial Economics 143*(1), 30–56.

Baumol, W. J. (1990). Entrepreneurship: Productive, unproductive, and destructive. *Journal of Political Economy 98*(5, Part 1), 893–921.

Beck, T., O. De Jonghe, and G. Schepens (2013). Bank competition and stability: Cross-country heterogeneity. *Journal of Financial Intermediation 22*(2), 218–244.

Ben-Shahar, O. and C. E. Schneider (2011). The failure of mandated discourse. *University of Pennsylvania Law Review 159*, 647.

Benetton, M., G. Buchak, and C. Robles-Garcia (2021). Competition and scope in banking: Synergies and product distortions in business lending.

Berg, T., V. Burg, A. Gombović, and M. Puri (2020). On the rise of fintechs: Credit scoring using digital footprints. *The Review of Financial Studies 33*(7), 2845–2897.

Berg, T., A. Fuster, and M. Puri (2021). Fintech lending.

Boot, A., P. Hoffmann, L. Laeven, and L. Ratnovski (2021). Fintech: What's old, what's new? *Journal of Financial Stability 53*, 100836.

Boot, A. W. and A. V. Thakor (1997). Financial system architecture. *The Review of Financial Studies 10*(3), 693–733.

Boot, A. W. and A. V. Thakor (2000). Can relationship banking survive competition? *The Journal of Finance 55*(2), 679–713.

Buchak, G., J. Hu, and S.-J. Wei (2021). Fintech as a financial liberator.

Buchak, G., G. Matvos, T. Piskorski, and A. Seru (2018a). Beyond the balance sheet model of banking: Implications for bank regulation and monetary policy.

Buchak, G., G. Matvos, T. Piskorski, and A. Seru (2018b). Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of Financial Economics 130*(3), 453–483.

Claessens, S. and L. Laeven (2004). What drives bank competition? Some international evidence. *Journal of Money, Credit and Banking*, 563–583.

Claessens, S. and L. Rojas-Suarez (2016). Financial regulations for improving financial inclusion. *Center for Global Development*.

Claessens, S. and N. Van Horen (2013). Impact of foreign banks. *Journal of Financial Perspectives 1*(1).

De Roure, C., L. Pelizzon, and A. V. Thakor (2021). P2P lenders versus banks: Cream skimming or bottom fishing?

Decker, R., J. Haltiwanger, R. Jarmin, and J. Miranda (2014). The role of entrepreneurship in US job creation and economic dynamism. *The Journal of Economic Perspectives 28*(3), 3–24.

Denes, M. R., S. T. Howell, F. Mezzanotti, X. Wang, and T. Xu (2020). Investor tax credits and entrepreneurship: Evidence from US states.

Di Maggio, M., M. Egan, and F. Franzoni (2021). The value of intermediation in the stock market. *Journal of Financial Economics*.

Di Maggio, M., D. Ratnadiwakara, and D. Carmichael (2021). Invisible primes: Fintech lending with alternative data.

Di Maggio, M. and V. Yao (2021). Fintech borrowers: Lax screening or cream-skimming? *The Review of Financial Studies 34*(10), 4565–4618.

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *The Review of Economic Studies 51*(3), 393–414.

Diamond, D. W. and P. H. Dybvig (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy 91*(3), 401–419.

Djankov, S., C. McLiesh, and A. Shleifer (2007). Private credit in 129 countries. *Journal of Financial Economics 84*(2), 299–329.

Egan, M., A. Hortaçsu, and G. Matvos (2017). Deposit competition and financial fragility: Evidence from the US banking sector. *American Economic Review 107*(1), 169–216.

Egan, M., S. Lewellen, and A. Sunderam (2017). The cross section of bank value.

Erel, I. and J. Liebersohn (2020). Does fintech substitute for banks? Evidence from the paycheck protection program.

Farboodi, M., R. Mihet, T. Philippon, and L. Veldkamp (2019). Big data and firm dynamics. *AEA Papers and Proceedings 109*, 38–42.

Foster, L., J. Haltiwanger, and C. Syverson (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review 98*(1), 394–425.

Fuster, A., M. Plosser, P. Schnabl, and J. Vickery (2019). The role of technology in mortgage lending. *The Review of Financial Studies 32*(5), 1854–1899.

Gennaioli, N., R. La Porta, F. Lopez-de Silanes, and A. Shleifer (2012). Human capital and regional development. *The Quarterly Journal of Economics 128*(1), 105–164.

Ghosh, P., B. Vallee, and Y. Zeng (2021). Fintech lending and cashless payments.

Glaeser, E. L., S. P. Kerr, and W. R. Kerr (2015). Entrepreneurship and urban growth: An empirical assessment with historical mines. *Review of Economics and Statistics 97*(2), 498–520.

Gopal, M. and P. Schnabl (2020). The rise of finance companies and fintech lenders in small business lending.

Gornall, W. and I. A. Strebulaev (2015). The economic impact of venture capital: Evidence from public companies.

Granja, J., C. Leuz, and R. Rajan (2018). Going the extra mile: Distant lending and credit cycles.

Grossman, S. J. (1981). The informational role of warranties and private disclosure about product quality. *The Journal of Law and Economics 24*(3), 461–483.

Haltiwanger, J., R. S. Jarmin, and J. Miranda (2013). Who creates jobs? Small versus large versus young. *Review of Economics and Statistics 95*(2), 347–361.

He, Z., J. Huang, and J. Zhou (2020). Open banking: Credit market competition when borrowers own the data.

Hertzberg, A., J. M. Liberti, and D. Paravisini (2011). Public information and coordination: evidence from a credit registry expansion. *The Journal of Finance 66*(2), 379–412.

Jiang, E., G. Matvos, T. Piskorski, and A. Seru (2020). Banking without deposits: Evidence from shadow bank call reports.

Jones, C. I. and C. Tonetti (2020, September). Nonrivalry and the economics of data. *American Economic Review 110*(9), 2819–58.

Kaplan, S. N. and J. Lerner (2010). It ain't broke: The past, present, and future of venture capital. *Journal of Applied Corporate Finance 22*(2), 36–47.

Klapper, L., L. Laeven, and R. Rajan (2006). Entry regulation as a barrier to entrepreneurship. *Journal of Financial Economics 82*(3), 591–629.

Kortum, S. and J. Lerner (2000). Assessing the contribution of venture capital to innovation. *The Rand Journal of Economics 31*(4), 674.

Kroszner, R. S. and P. E. Strahan (1999). What drives deregulation? Economics and politics of the relaxation of bank branching restrictions. *The Quarterly Journal of Economics 114*(4), 1437–1467.

Laeven, L. and R. Levine (2009). Bank governance, regulation and risk taking. *Journal of Financial Economics 93*(2), 259–275.

Lindsey, L. A. and L. C. Stein (2019). Angels, entrepreneurship, and employment dynamics: Evidence from investor accreditation rules. In *Sixth Annual Conference on Financial Market Regulation*.

Lucas, R. E. (1978). On the size distribution of business firms. *The Bell Journal of Economics*, 508–523.

Mullainathan, S. and P. Schnabl (2010). Does less market entry regulation generate more entrepreneurs? Evidence from a regulatory reform in Peru. In *International Differences in Entrepreneurship*, pp. 159–177. University of Chicago Press.

Ongena, S., A. Popov, and G. F. Udell (2013). "When the cat's away the mice will play": Does regulation at home affect bank risk-taking abroad? *Journal of Financial Economics 108*(3), 727–750.

Parlour, C. A., U. Rajan, and H. Zhu (2020). When fintech competes for payment flows.

Petersen, M. A. and R. G. Rajan (1994). The benefits of lending relationships: Evidence from small business data. *The Journal of Finance 49*(1), 3–37.

Petersen, M. A. and R. G. Rajan (1995). The effect of credit market competition on lending relationships. *The Quarterly Journal of Economics 110*(2), 407–443.

Philippon, T. (2019). On fintech and financial inclusion.

Phillips, G. M. and A. Zhdanov (2017). Venture capital investments and merger and acquisition activity around the world.

Puri, M. and D. T. Robinson (2007). Optimism and economic choice. *Journal of Financial Economics 86*(1), 71–99.

Puri, M. and R. Zarutskie (2012). On the life cycle dynamics of venture-capital-and non-venture-capital-financed firms. *The Journal of Finance 67*(6), 2247–2293.

Ramakrishnan, R. T. and A. V. Thakor (1984). Information reliability and a theory of financial intermediation. *The Review of Economic Studies 51*(3), 415–432.

Schumpeter, J. A. (1911). *Theorie Der Wirtschaftlichen Entwicklung.* Leipzig: Duncker & Humblot.

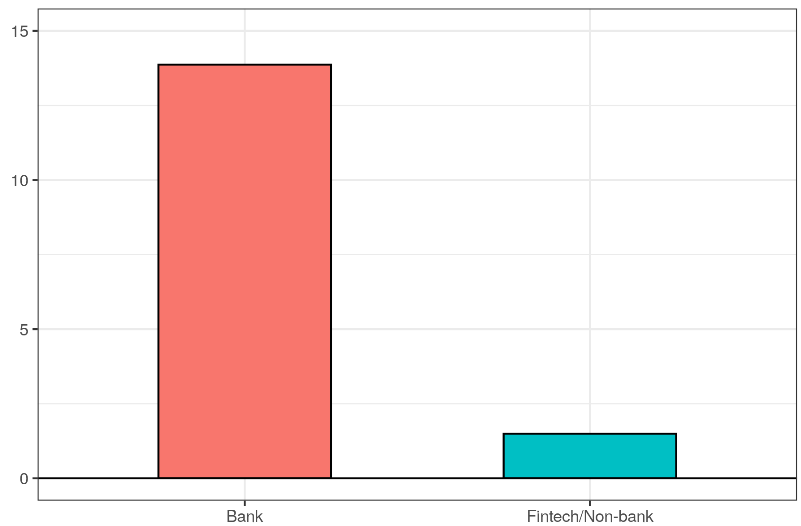
Stulz, R. M. (2019). Fintech, bigtech, and the future of banks. *Journal of Applied Corporate Finance 31*(4), 86–97.

Thakor, A. V. (2020). Fintech and banking: What do we know? *Journal of Financial Intermediation 41*, 100833.

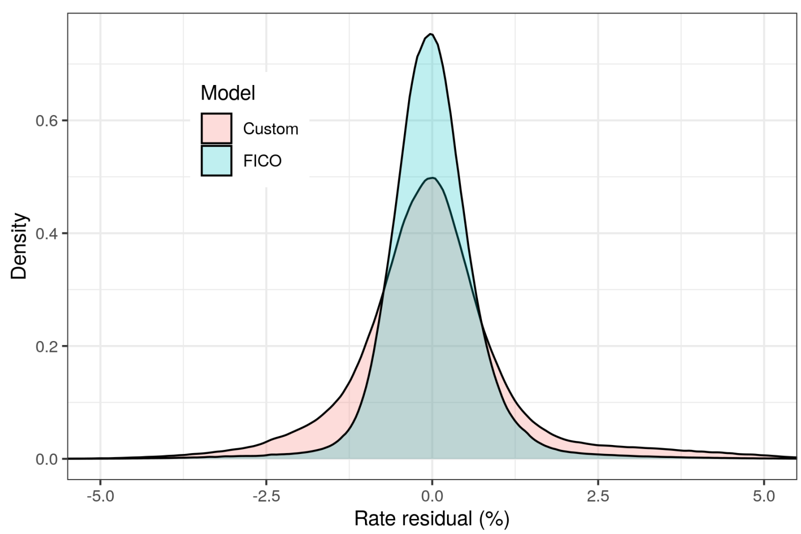
Vives, X. (2019). Digital disruption in banking. *Annual Review of Financial Economics 11*, 243–272.

Figure 1: Data Use by Banks and Non-banks/Fintechs in the US Mortgage Market

Note: This figure shows the use of credit-scoring models by banks and non-banks and interest rate residuals in the US residential mortgage market. Panel (a) shows the fraction of mortgages originated using a credit scoring model besides standardized Equifax, Experian, FICO, or Vantage Score for depository (red) and non-depository (blue) institutions. Panel (b) shows the distribution of interest rate residuals for custom (red) and standardized (blue) credit scoring models after controlling for interacted LTV, loan purpose, lien status, loan type, debt-to-income ratio, whether the loan is a reverse mortgage, open-end line of credit, made for a business purpose, HOEPA status, construction method, occupancy type, and conforming status fixed effects, plus year-MSA fixed effects. Data are from HMDA for 2018 and 2019, merged with the Avery file to identify lender type.



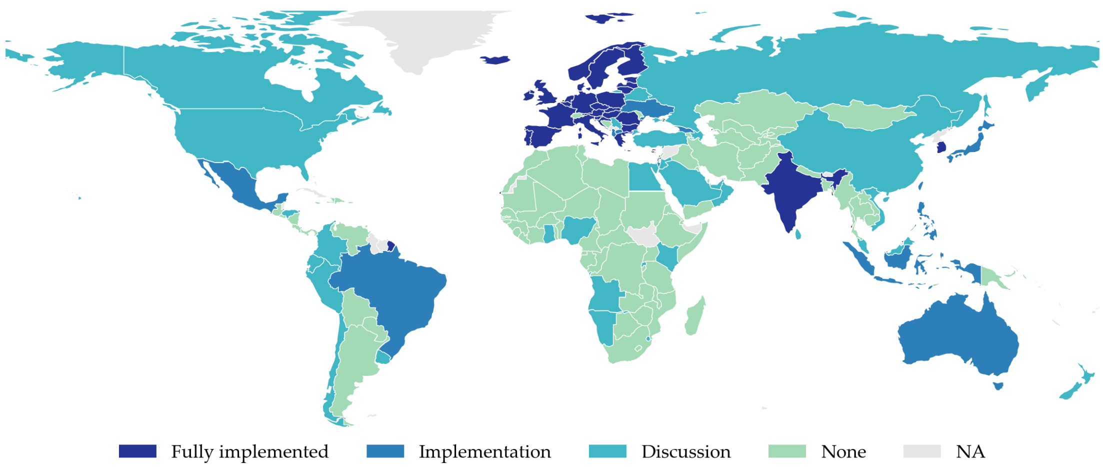
(a) Percentage of mortgages originated using alternate credit scoring methods



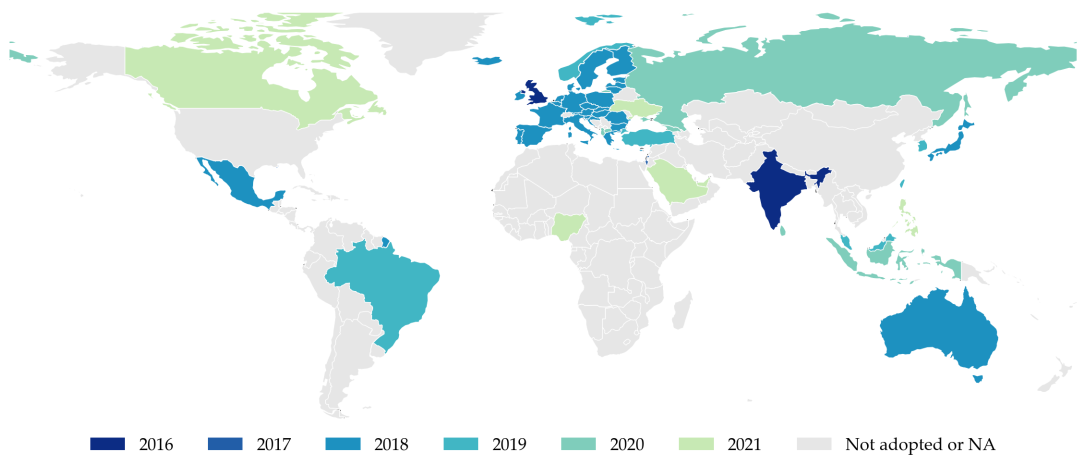
(b) Mortgage interest rate residuals by credit scoring method

Figure 2: Government-led Open Banking Regimes Around the World

Note: These maps show the current implementation status of government-led open banking policies and the year in which the major open banking policy was passed. Panel (a) shows the implementation status of their government open banking policies. Fully implemented corresponds to countries that have implemented open banking government policies; Implementation to those that have determined the specifics of the open banking approach and are currently implementing it; Discussion to those either considering implementing open banking policies or discussing that implementation; None to those with no government open banking approach; and NA to those where we have not collected data. Panel (b) shows the passage year of countries' major open banking policies. Data on government open banking policies are current as of October 2021.



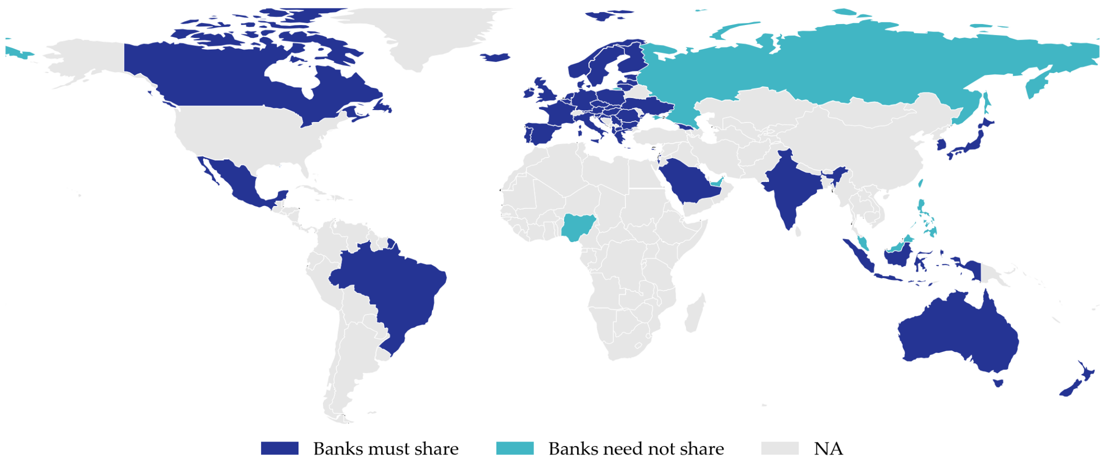
(a) Government open banking policy implementation status



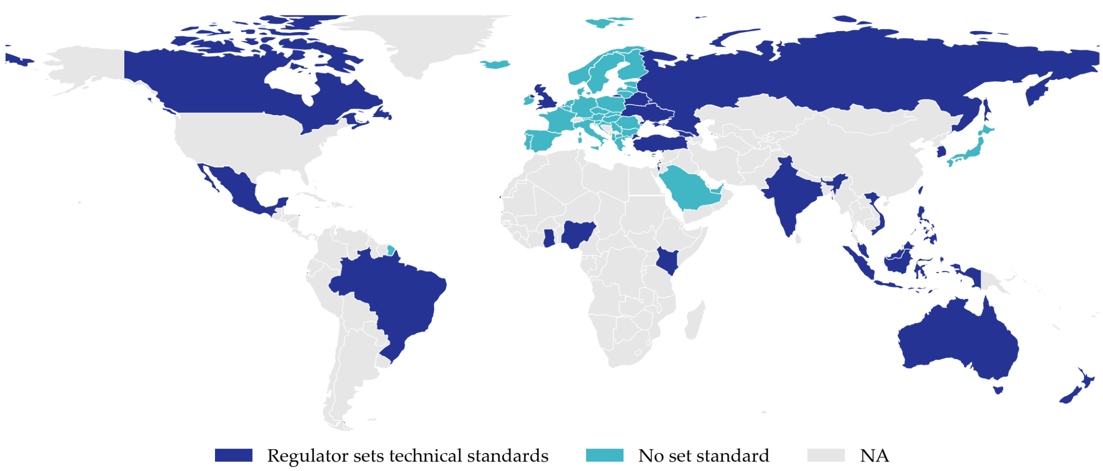
(b) Timeline of open banking adoption

Figure 3: Open Banking Government Policy Dimensions

Note: These maps show mandated data sharing and technical specifications among countries with government-led open banking efforts developed enough to specify those policy dimensions. Panel (a) shows whether the current or proposed policy requires banks to share data upon customer request. Panel (b) shows whether the regulator sets a technical standard for open banking application programming interfaces—the technology used to share bank customer data. Countries marked NA either have no government-led open banking regime, are too early in discussion for the issue to be decided, or were excluded from our data collection. Data on government open banking policies are current as of October 2021.



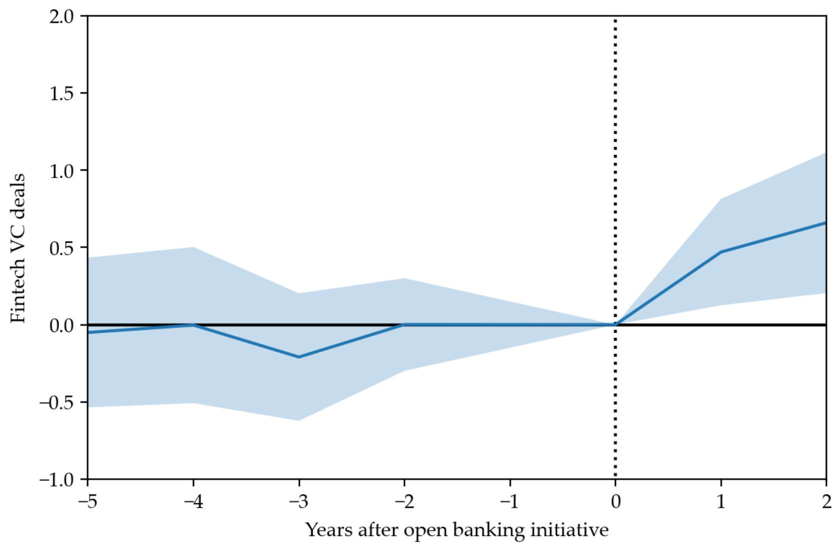
(a) Banks must share data on customer request



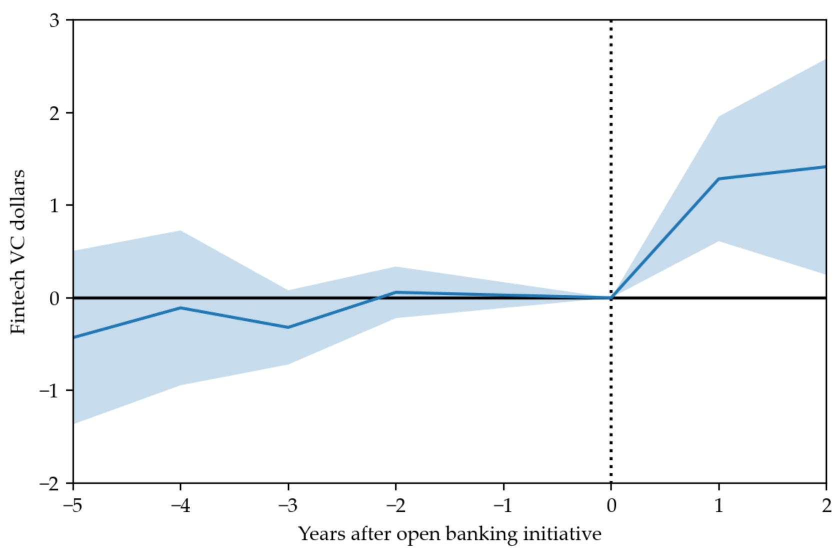
(b) Regulator sets technical standards

Figure 4: Event-study of Fintech Investment After Open Banking Government Policies

Note: This figure shows changes in fintech venture capital (VC) activity around the passage of open banking government policies using a panel event-study analysis. We perform this analysis on our high-coverage Pitchbook panel of 2011-2021 data for the 21 countries with at least five fintech VC deals in the 2000–2010 period. Panel (a) shows an event study on the log of one plus the number of fintech VC deals, and Panel (b) shows an event study on the log of one plus the millions of US dollars invested in fintech VC deals. Year 0 is the passage year of each country's major open banking initiative. The coefficient for year 0 is set to zero and other coefficients are presented net of country and region-by-year fixed effects based on the World Bank regions. European Union member states are weighted to count as a single country for estimates and standard errors. The shaded regions denote 95% confidence intervals calculated using standard errors clustered at the country level.



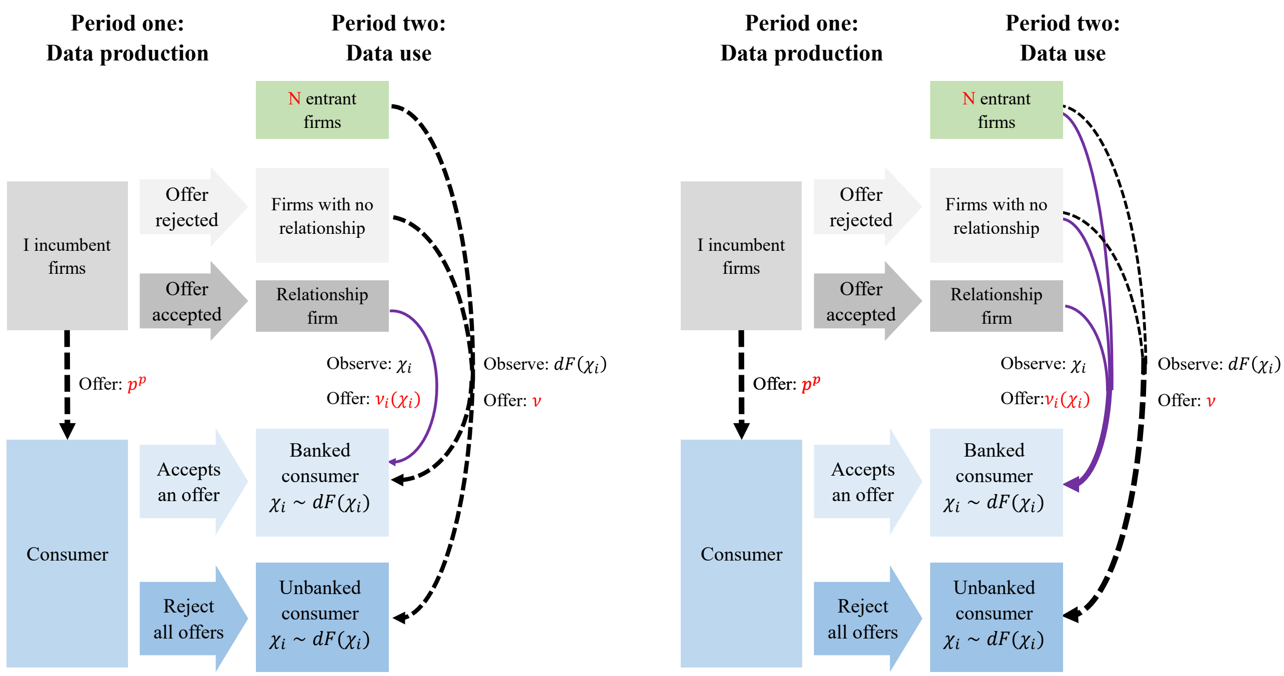
(a) Log of number of fintech VC deals



(b) Log of amount of fintech VC investment in millions of US dollars

Figure 5: Open Banking Modeling Framework

Note: This figure shows the modeling framework schematically. Panel (a) shows the relationship banking case and Panel (b) shows the open banking case. In both cases, incumbent firms offer financial products to a mass of customers in the data-producing first period. Some customers accept an offer and form a data-producing relationship, some do not and become unbanked. In the data-using second period, both types of customers receive offers from incumbent firms and new entrants. In the relationship banking case, a customer's characteristics are observed only by the firm that had a relationship with them in the data-producing period and only that firm can offer a product conditional on those characteristics (solid lines). Other firms can only make a pooling offer based on the distribution of characteristics (dashed lines). In the open banking case, all firms observe the characteristics of customers that used the data-producing product and can thus condition offers on those characteristics. In both cases, no firms observe the characteristics of the unbanked consumers. In both figures, quantities shown in red are equilibrium outcomes.



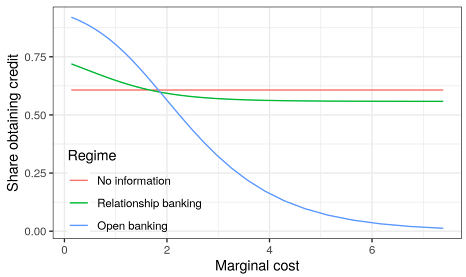
(a) Relationship banking (b) Open banking

Figure 6: Aggregate Outcomes of Open Banking

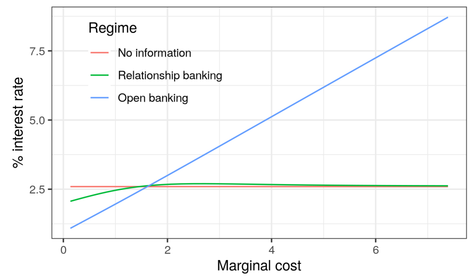
Note: This figure shows how open banking impacts outcomes in the data-using period. Magenta bars show outcomes for the non-GSE residential mortgage calibration, roughly following Buchak et al. (2018a). Cyan bars show outcomes for the financial advice calibration, roughly following Di Maggio et al. (2021). Each bar shows the percentage change in the relevant outcome caused by moving from a relationship banking regime (where only the incumbent relationship bank observes customer data) to an open banking regime (where all incumbent banks and new entrants observe customer data). Entrants is the change in the number of new entrants caused by open banking adoption. Volume is the change in consumer use of the respective data-using product. Incumbent profit is the change in total bank profit (as only incumbents are profitable in equilibrium) in the data-using period. Consumer welfare is the change in consumer utility in the data-using period. Total welfare is the change in incumbent profit plus consumer welfare in the data-using period.

Figure 7: Heterogeneous Effects of Open Banking

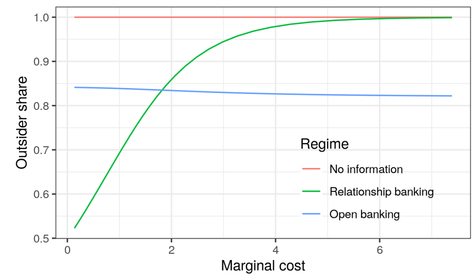
Note: This figure shows how open banking differentially impacts consumers. The red lines correspond to no financial firms having access to customer data, the green lines to only one financial firm having access (relationship banking), and the blue lines to all financial firms having access (open banking). The first row shows mortgage outcomes by borrower heterogeneity in marginal cost, mc_i. The second row shows mortgage outcomes by consumers' willingness to pay, $1/\alpha_i$. The third row shows financial advice outcomes by consumers' customization needs, $|f_i - \bar{f}|$. The three columns show product quantities, prices, and the share of consumers switching away from their relationship bank, respectively.



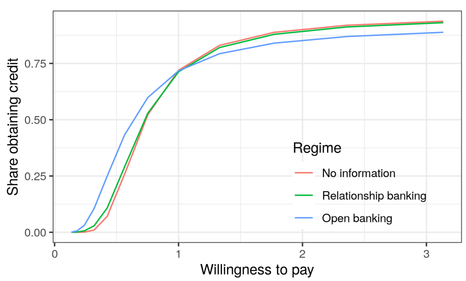
(a) Mortgage quantities



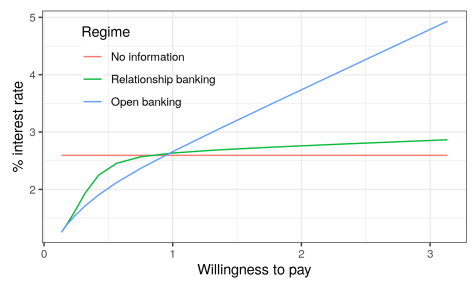
(b) Mortgage prices



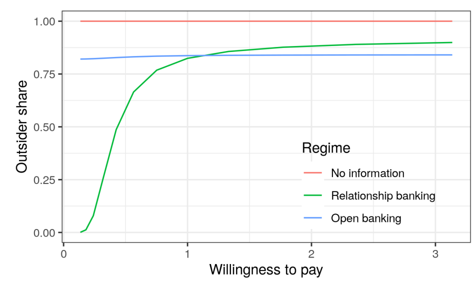
(c) Mortgage outsider share



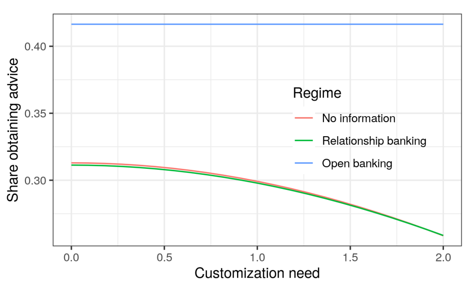
(d) Mortgage quantities



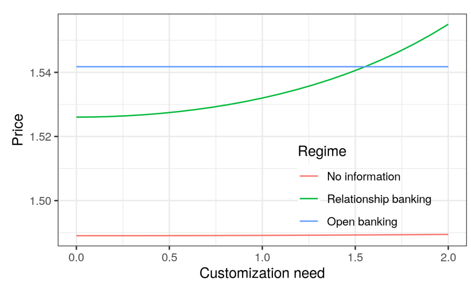
(e) Mortgage prices



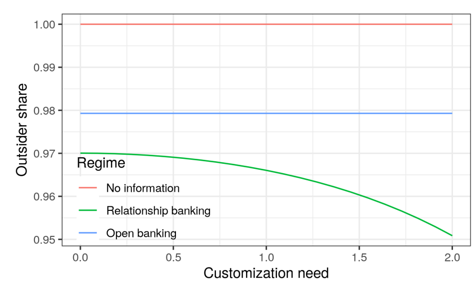
(f) Mortgage outsider share



(g) Advice quantities



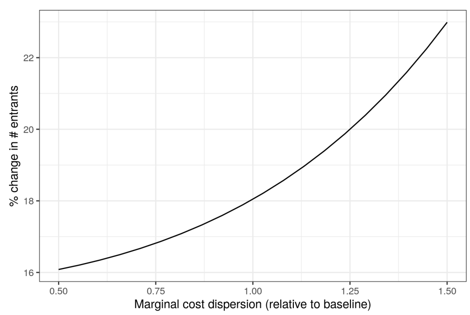
(h) Advice prices



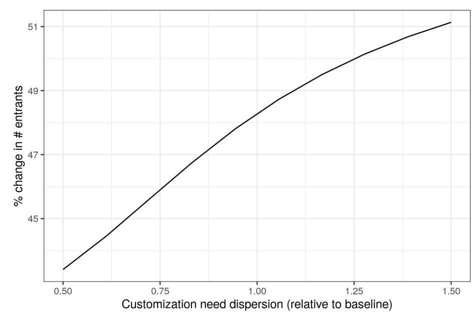
(i) Advice outsider share

Figure 8: Open Banking Comparative Statics

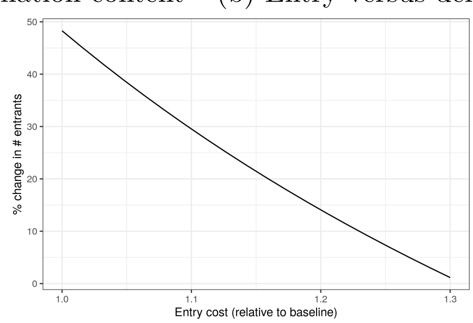
Note: This figure shows comparative statics for the open banking model. Panels (a) through (c) show the percentage change in firm entry caused by moving from relationship banking information regime to open banking as a function of varying dispersion of consumer marginal cost, σ_{mc}, in the mortgage case; varying dispersion in consumer customization needs, σ_f, in the financial advice case; and a varying entry cost paid by new entrants under open banking, c, in the financial advice case, respectively.



(a) Entry versus information content



(b) Entry versus demand for customization



(c) Entry versus fixed entry cost

Figure 9: Open banking and data-production

Note: This figure shows how open banking impacts outcomes across the two periods for the mortgage calibration, roughly following Buchak et al. (2018a). Magenta bars show outcomes if consumers are myopic and ignore the data-using period when choosing data-producing products. Cyan bars show outcomes if consumers are rational and fully incorporate period two utility in their data-producing product decisions. Each bar shows the percentage change in the relevant outcome caused by moving from a relationship banking regime (where only the incumbent relationship bank can observe customer data) to an open banking regime (where all incumbent banks and new entrants observe customer data). Transaction service volume is the change in the use of the data-producing product. Ex-ante incumbent profit is the change in total bank profit across the two periods. Consumer welfare is the change in total consumer utility across the two periods. Total welfare is the change in ex-ante incumbent profit plus consumer utility across the two periods.

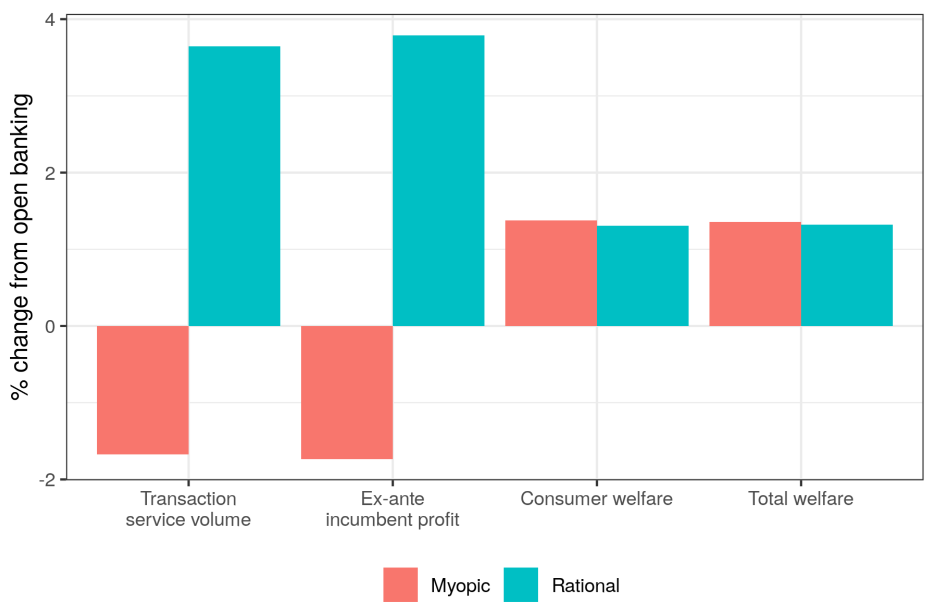


Table 1: Open Banking Regime Summary Statistics

Note: This table presents summary statistics on open banking regimes for 168 countries. The first number is the percentage of countries fitting the given criteria and the number in parentheses is the number of countries under consideration. Open banking presence considers all 168 countries for which data were collected while the other categories (policy mandates, status, participation, product scope, and functionality scope) consider only countries with a government-led open banking approach that has advanced far enough that the issue in question has been (at least preliminarily) decided. Columns split the sample into regions based on the World Bank classification.

Variable	Worldwide	Europe/Central Asia	East Asia	MENA	Sub-Sarahan Africa	South Asia	LATAM	North America
Open banking presence								
Any open banking	52% (168)	86% (50)	68% (19)	45% (20)	18% (45)	33% (6)	40% (25)	67% (3)
Market-led	29% (168)	66% (50)	21% (19)	25% (20)	2% (45)	0% (6)	20% (25)	33% (3)
Government-led	48% (168)	80% (50)	63% (19)	40% (20)	18% (45)	33% (6)	32% (25)	67% (3)
Policy justification								
Innovation	97% (65)	97% (39)	100% (11)	100% (5)	100% (4)	50% (2)	100% (3)	100% (1)
Competition	82% (65)	87% (39)	73% (11)	80% (5)	50% (4)	100% (2)	100% (3)	0% (1)
Inclusion	29% (66)	10% (39)	45% (11)	33% (6)	50% (4)	100% (2)	100% (3)	100% (1)
Status								
Discussion	38% (80)	12% (40)	33% (12)	62% (8)	88% (8)	50% (2)	75% (8)	100% (2)
Mid-implementation	18% (80)	12% (40)	50% (12)	12% (8)	0% (8)	0% (2)	25% (8)	0% (2)
Implemented	44% (80)	75% (40)	17% (12)	25% (8)	0% (8)	50% (2)	0% (8)	0% (2)
Policy strength								
Required data sharing	88% (57)	97% (37)	60% (10)	80% (5)	0% (1)	100% (1)	100% (2)	100% (1)
Data reciprocity	18% (56)	0% (36)	40% (10)	40% (5)	0% (1)	100% (1)	100% (2)	100% (1)
Regulator provides tech specs	39% (62)	15% (39)	82% (11)	40% (5)	100% (3)	100% (1)	100% (2)	100% (1)
Beyond transaction accts	34% (56)	3% (36)	90% (10)	75% (4)	100% (1)	100% (1)	100% (3)	100% (1)
Functionality scope								
Data sharing only	5% (58)	0% (38)	10% (10)	0% (5)	0% (1)	0% (1)	50% (2)	100% (1)
Payments only	0% (58)	0% (38)	0% (10)	0% (5)	0% (1)	0% (1)	0% (2)	0% (1)
Both	95% (58)	100% (38)	90% (10)	100% (5)	100% (1)	100% (1)	50% (2)	0% (1)

Table 2: Drivers of Open Banking Government Policies

Note: This table shows whether ex-ante country characteristics predict the implementation of open banking government policies. The dependent variable is an indicator variable equal to one if open banking was implemented in the country in question as of October 2021 in columns 1–5, and in column 6 the dependent variable is a score between 0 and 7 based on a country's open banking implementation progress as of October 2021, with 0 being no action, 1–2 being increasingly serious levels of discussion, and 3–7 being levels of implementation progress. The open banking implemented indicator corresponds to being in or after the pre-implementation stage or equivalently to a level of 3 or above. The independent variables are country characteristics measured as of 2013. Private sector credit to GDP, bank branches per 100k people, and the financial sector Lerner index are from the World Bank. Non-fintech and fintech VC deals are from PitchBook and are used after taking the log of one plus the number of deals. Foreign-owned banks are from the Claessens and Van Horen (2013) foreign bank ownership data. The Rule of Law and Business Regulation Indexes are from the Cato Institute and are on a 0 to 10 scale with higher numbers denoting more favorable conditions. The Corruption Perception Index is from Transparency International and is on a 0 to 100 scale with higher numbers denoting more favorable conditions. All specifications include GDP per capita in thousands of US dollars, the square of GDP per capita in hundreds of thousands of US dollars, the log of population, and region fixed effects as controls, all based on World Bank data. European Union member states are weighted to count as a single country for estimates and standard errors. The regressions are cross-sectional, where each country in the sample corresponds to a single data point. *** denotes p-value < 0.01, ** denotes <0.05, and * denotes <0.1.

	Open banking implemented (0/1)					OB implementation (0-7)
	(1)	(2)	(3)	(4)	(5)	(6)
Private sector credit to GDP	0.000				-0.002	-0.006
	(0.002)				(0.002)	(0.006)
Branches per 100k people	-0.003				-0.005	-0.029
	(0.006)				(0.007)	(0.021)
Financial sector Lerner index	0.149				-0.012	2.674*
	(0.463)				(0.401)	(1.366)
Non-fintech VC deals		0.080*			0.075	0.440*
		(0.042)			(0.072)	(0.254)
Fintech VC deals		0.070			0.135	0.523
		(0.073)			(0.099)	(0.348)
Foreign-owned banks			0.102		0.141	0.134
			(0.121)		(0.195)	(0.660)
Rule of Law Index				0.047	0.030	0.120
				(0.050)	(0.103)	(0.387)
Business Regulation Index				0.009	-0.011	-0.061
				(0.044)	(0.061)	(0.192)
Corruption Perception Index				0.001	0.002	0.038
				(0.005)	(0.008)	(0.029)
Per capita GDP ($k)	0.015	0.004	0.020***	0.013*	-0.005	-0.038
	(0.010)	(0.006)	(0.006)	(0.007)	(0.010)	(0.034)
Per capita GDP ($100k) squared	-1.275	-0.386	-1.560**	-1.196*	0.356	1.676
	(0.996)	(0.626)	(0.675)	(0.714)	(0.848)	(3.064)
Log population	0.044	-0.032*	0.049*	0.053**	-0.068	-0.145
	(0.049)	(0.019)	(0.027)	(0.025)	(0.059)	(0.173)
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	85	162	133	145	82	82
Adjusted R^2	0.308	0.386	0.320	0.337	0.422	0.617

Table 3: Open Banking Government Policy and Bank API Offerings

Note: This table shows the association between government open banking policies and banks' open application programming interfaces (APIs). The dependent variable in columns 1 to 3 is the log of one plus the number of banks offering APIs and in 4 to 6 it is the percentage of the top 10 banks in each country (as ranked by 2020 assets in Bureau van Dijk) that offer APIs. APIs are the technology used to share bank customer data under open banking. The independent variable of interest in columns 1 and 4 is an indicator variable equal to one if open banking was implemented in the country in question as of October 2021; in columns 2 and 5 is a 0–7 rating of the extent of open banking government policy implementation progress as of October 2021, with 0 being no action, 1–2 being increasingly serious levels of discussion, and 3–7 being levels of implementation progress; and in columns 3 and 6 it is the interaction of the open banking implemented indicator variable with our Open Banking Strength Index which is a measure of policy strength. The open banking implemented indicator corresponds to being in or after the pre-implementation stage or equivalently to a level of 3 or above. All specifications include GDP per capita in thousands of US dollars, the square of GDP per capita in hundreds of thousands of US dollars, the log of population, and region fixed effects as controls, all based on World Bank data as of 2013. European Union member states are weighted to count as a single country for estimates and standard errors. *** denotes p-value < 0.01, ** denotes <0.05, and * denotes <0.1.

	Banks with APIs			% of top 10 banks with APIs		
	(1)	(2)	(3)	(4)	(5)	(6)
Open banking implemented (0/1)	0.648***		-0.334	0.190***		-0.013
	(0.226)		(0.299)	(0.071)		(0.062)
Open banking implementation (0-7)		0.201***			0.052***	
		(0.043)			(0.013)	
OB Strength Index X OB implemented			1.131***			0.233***
			(0.368)			(0.089)
Per capita GDP ($k)	0.039***	0.025**	0.035***	0.005	0.002	0.004
	(0.013)	(0.012)	(0.012)	(0.003)	(0.003)	(0.004)
Per capita GDP ($100k) squared	-1.516	-0.273	-1.217	-0.138	0.149	-0.077
	(1.685)	(1.509)	(1.605)	(0.451)	(0.407)	(0.441)
Log population	0.291***	0.265***	0.287***	0.046***	0.040***	0.045***
	(0.049)	(0.047)	(0.049)	(0.014)	(0.013)	(0.014)
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	157	157	157	157	157	157
Adjusted R^2	0.629	0.672	0.645	0.488	0.531	0.500

Table 4: Effect of Open Banking Government Policy on Fintechs

Note: This table shows changes in fintech venture capital (VC) investment following the implementation of open banking government policies. The table uses a difference-in-difference design on our high-coverage Pitchbook panel of country-year data spanning 2011-2021 for the 21 countries with at least five fintech deals in the 2000–2010 period. The dependent variable in columns 1 to 3 is the log of one plus the number of fintech deals in a country-year, and in columns 4 to 6 it is the log of one plus the amount invested in millions of US dollars. The independent variable is an indicator variable equal to one if the year in question is after the year major open banking laws were passed in the country in question. Columns 3 and 6 include a control for non-fintech VC activity using Pitchbook data, transformed the same way as fintech VC activity. All specifications control for country fixed effects; columns 1 and 4 contain controls for year fixed effects; and columns 2, 3, 5, and 6 control for region-by-year fixed effects, where regions are based on the World Bank classification. European Union member states are weighted to count as a single country for estimates and standard errors. Standard errors are clustered at the country level. *** denotes p-value < 0.01, ** denotes <0.05, and * denotes <0.1.

	Fintech VC deals			Fintech VC dollars		
	(1)	(2)	(3)	(4)	(5)	(6)
After OB initiative	0.214*	0.532***	0.363**	0.746**	1.305**	1.104**
	(0.111)	(0.160)	(0.156)	(0.267)	(0.474)	(0.432)
Non-fintech VC deals			0.475***			
			(0.135)			
Non-fintech VC dollars						0.292**
						(0.095)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes			Yes		
Region-Year FE		Yes	Yes		Yes	Yes
Observations	231	231	231	231	231	231
Adjusted R^2	0.919	0.929	0.936	0.877	0.895	0.898

Table 5: Effect of Open Banking Government Policy on Fintechs by Product Area

Note: This table shows changes in fintech venture capital (VC) investment following the implementation of government open banking policies. The table uses a difference-in-difference design on our high-coverage Pitchbook panel of country-year data spanning 2011-2021 for the 21 countries with at least five fintech deals in the 2000–2010 period. The dependent variable in each specification is the log of one plus the number of VC deals in a country-year and given subsector of fintech, where subsectors are defined based on Pitchbook keywords as described in Appendix B. The independent variable is an indicator variable equal to one if the year in question is after the year major open banking laws were passed in the country in question. Panel (a) performs a panel regression with country and region-by-year fixed effects, where regions are based on the World Bank classification. Panel (b) controls for the log of one plus the number of non-fintech VC deals. European Union member states are weighted to count as a single country for estimates and standard errors. Standard errors are clustered at the country level. *** denotes p-value < 0.01, ** denotes <0.05, and * denotes <0.1.

Panel (a) Without non-fintech VC activity control

	Alternative lending	Consumer finance	Financial IT	Payments	Regtech	Wealth management	Digital assets
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
After OB initiative	0.737*	0.693**	0.760***	0.654	0.709***	0.624*	0.022
	(0.355)	(0.260)	(0.230)	(0.407)	(0.135)	(0.329)	(0.279)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	231	231	231	231	231	231	231
Adjusted R^2	0.866	0.835	0.877	0.863	0.876	0.875	0.828

Panel (b) With non-fintech VC activity control

	Alternative lending	Consumer finance	Financial IT	Payments	Regtech	Wealth management	Digital assets
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
After OB initiative	0.623*	0.589**	0.644***	0.508*	0.544*	0.424	-0.132
	(0.326)	(0.205)	(0.201)	(0.277)	(0.250)	(0.312)	(0.280)
Non-fintech VC deals	0.323	0.292	0.329	0.409	0.463*	0.562***	0.434
	(0.276)	(0.225)	(0.218)	(0.251)	(0.221)	(0.163)	(0.340)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	231	231	231	231	231	231	231
Adjusted R^2	0.868	0.838	0.880	0.868	0.883	0.883	0.834

Table 6: Effect of Open Banking Government Policy and Its Characteristics on Fintechs

Note: This table shows changes in fintech venture capital (VC) investment activity following the implementation of different types of open banking policies by governments around the world. The table uses a difference-in-difference design on our high-coverage Pitchbook data of country-year data spanning 2011-2021 for the 21 countries with at least five fintech deals in the 2000–2010 period. The dependent variable in each specification is the log of one plus the number of fintech VC deals in each country-year. The independent variables are different characteristics of open banking government policies interacted with an indicator variable equal to one if the year in question is after the year major open banking laws were passed in the country in question. In column 1 we indicate whether banks are mandated to share the data with other financial service providers upon consumer request; in column 2 whether there is data reciprocity between banks and other financial service providers (e.g., if fintechs have to share customer data with banks); in column 3 whether regulators set technical standards for open banking implementation; and in column 4 whether, in addition to bank payment accounts, open banking policies cover other financial products and services (e.g., mortgages, insurance). In column 5, we interact with the Open Banking Strength Index, which we define as the average of those four policy dimensions used in columns 1 to 4. All specifications have country and region-by-year fixed effects, where regions are based on the World Bank classification. European Union member states are weighted to count as a single country for estimates and standard errors. Standard errors are clustered at the country level. *** denotes p-value < 0.01, ** denotes <0.05, and * denotes <0.1.

	Fintech VC Deals				
	(1)	(2)	(3)	(4)	(5)
Banks must share X after OB	0.322**				
	(0.132)				
Users must share X after OB		0.394***			
		(0.120)			
Technical specification X after OB			0.165		
			(0.135)		
Beyond transactions X after OB				0.361	
				(0.202)	
OB Strength Index X after OB					0.751**
					(0.264)
After OB initiative	0.063	0.300*	0.288	0.257	0.018
	(0.133)	(0.146)	(0.195)	(0.170)	(0.268)
Non-fintech VC deals	0.446***	0.494***	0.500***	0.483***	0.500***
	(0.143)	(0.145)	(0.152)	(0.147)	(0.152)
Country FE	Yes	Yes	Yes	Yes	Yes
Region-Year FE	Yes	Yes	Yes	Yes	Yes
Observations	231	231	231	231	231
Adjusted R^2	0.935	0.936	0.935	0.936	0.936

Appendix

A Open Banking Data Collection and Variable Definitions

This appendix describes the construction of our OB government policies dataset and defines variables. Each observation in the dataset corresponds to a country's OB approach as of the collection date.[34]

A.1 What is an "approach" and what makes an approach "open banking"?

A government-led OB approach does not need to be a single law or policy; many countries' OB approaches in fact are composed of several separate policies. Rather, an approach encompasses the totality of the country's OB government efforts.

The line between OB policies and related but non-OB policies can be unclear, and a single simple definition cannot encompass all cases. For our purposes, there are two reasons for us to classify a regulatory approach as OB:

1. **Functional**: Does the regulator's approach have the key functional elements of OB? Specifically, does it facilitate programmatic access (e.g., through an API) to financial intermediaries' customers' data for the purposes of data sharing or payments?

2. **Nominal**: Do regulators, journalists, or industry groups refer to the regulation as "open banking"?

The functional approach is more objective and can be applied to countries that have progressed sufficiently far down the pathway of discussing and implementing OB policies. The nominal approach is useful in cases where regulators have only recently been discussing OB but none of the functional elements have yet been formalized. The following two regulations may be similar to OB but we do not consider to be OB policies and we list them as illustrative counterexamples:

1. General Data Protection Regulation (GDPR): This EU law grants consumers certain privacy rights over their data. However, GDPR is not an OB law because it does not mandate that commercial entities (specifically, banks) in possession of the data share it upon customer request. Note, however, that the EU does have an OB law, the PSD2.

2. Regulation related to central bank digital currencies (CBDC): Movements to create payment systems utilizing CBDC are payments regulations but are not open payments regulations, as they do not mandate open data sharing between market participants.

[34] Most recently, October 2021.

There have been many payments-related regulations (CBDC and other) that modernize payments but are not "open" in any sense, aside from, for example, reporting requirements to regulators.

Having defined what constitutes an "approach" and what makes an approach an "open banking" approach, we now define in detail the variables we collect and the classification decision rules. With each data category, we provide notes to clarify decision rules and address common questions.

A.2 Data categories and variable definitions

A.2.1 Open banking approach and regulatory mandate

- **market_led_initiative**: Is there a market-led initiative independent from government involvement, e.g., a consortium of banks or fintechs coordinating around OB standards?
  - **Yes.**
  - **No.**
- **government_led_initiative**: Is there a government-led initiative around OB?
  - **Yes.**
  - **No.**
- **regulatory_entity_type**: Which type of regulator is leading the OB effort?
  - **Monetary authority**: A financial regulator, e.g., a central bank.
  - **Competition authority**: A regulator tasked with anti-trust or other competition-related enforcement, e.g., the Competition and Markets Authority in the UK.
  - **Consumer protection authority**: A regulator tasked with consumer protection, e.g., the Consumer Financial Protection Bureau in the US or a data privacy authority.
- **innovation_mandate**: Is increasing innovation a proffered policy mandate?
  - **Yes**: Spurring the creation or adoption of new financial products or technologies is either discussed or explicitly stated as policy goals.
  - **No**: Otherwise.
- **competition_mandate**: Is increasing competition a proffered policy mandate?

- **Yes**: Increasing entry, increasing competition, decreasing markups, or related issues are either discussed or explicitly stated as policy goals.
- **No**: Otherwise.

- **inclusion_mandate**: Is increasing financial inclusion a proffered policy mandate?
  - **Yes**: Increasing access to the financial system, serving the unbanked, fighting inequality, or related issues are either discussed or explicitly stated as policy goals.
  - **No**: Otherwise.

How do we denote efforts coordinated between both regulators and market participants?

We define these as government-led efforts. The justification for this is that almost all major government policies involve some level of collaboration or input from industry. In the US, for example, there are open comment periods and meetings with industry and lobbyists. Fundamentally, however, these initiatives work through the government, and so to the extent that the government has any authoritative hand in leading the regulation, we consider it as government-led.

Which agency type do we select in cases where several are responsible?

We select the regulator most aligned with the proffered mandate or rationale for OB. For example, in the case of Australia, we select the Australian Competition and Consumer Commission because the country's OB policy mandate is most closely aligned with that of a "competition authority".

A.2.2 Timeline and initiative

- **initiative_name**: Name of the government-led policy initiative.
- **initiative_passed_date**: Date that the OB legislation is signed into law, or date when the first non-regulation government major effort to promote OB goes into effect (e.g., for Singapore we use November of 2016—the date when the Monetary Authority of Singapore (MAS) published a comprehensive roadmap: API Playbook—which, in effect, set the gold standard for regulatory advice on the topic in Asia: see here). For efforts that have not yet been signed into law or resulted in a major government policy, this field is TBD.
- **data_sharing_date**: First date at which the legal mandate on customers' data sharing begins to bind, or (in cases of non-legally-binding policies) when the government sets up the infrastructure that allows customer data sharing.
- **oct_21_status**: Implementation status as of October 2021.

- **Nothing**: No government-led OB.
- **Pre-discussion**: Some government interest but no actual law or implementation is taking place.
- **Discussion**: The actual law has been introduced or passed and rulemaking is taking place.
- **Pre-implementation**: The law is passed and rules have been set, but nothing is yet binding.
- **Early implementation**: Some data sharing requirements are binding (e.g., bank-level product information), but not personal account/transactions.
- **Mid implementation**: Personal account/transaction data sharing is binding, but not all planned elements are in place (e.g., not all planned API functionality exists.)
- **Fully implemented**: Full implementation as described in the law/rulemaking.
- **Follow-on regulation**: OB is implemented, and regulators are actively working on related regulation such as open finance or open data more broadly.

Which government effort do we focus on when there are several?

We focus on the first major government OB effort.[35] For example, in the United States, several regulatory bodies have expressed interest in OB (e.g., the Treasury/OCC and the Consumer Financial Protection Bureau (CFPB)). The CFPB's effort through Dodd-Frank Section 1033 is the most important US regulatory effort. In the UK, the 2016 CMA9 order was the first major open banking law, although it is subject to pending follow-up regulation to broaden its scope.

What if the precise date is unavailable?

In cases where the precise date cannot be found or is ambiguous for some reason, we use the most precise date that can be inferred from the data. For example, if the best information for a country that can be located says the law passed in "the second half of 2020," we will assign the date as July 1, 2020.

What event defines the data sharing date?

In cases where data sharing is mandated, this is the date. In cases where data sharing is not mandated but, for example, the regulator sets API standards, we use the date at which the API standards go into force. In cases where the regulation initially applies

[35] Given the recency of the OB trend, this is almost always also the latest OB approach with the exceptions being the United Kingdom and Sweden. These two countries had earlier, abortive OB attempts that we exclude due to their limited implementation.

only to a subset of later planned entities (e.g., the UK Open Banking Initiative applies to 9 large banks), we use the date at which the requirements first apply to any entity.

A.2.3 Standards

- **regulatory_technical_specifications**: Does the regulator set technical specifications for data sharing / payments?
  - **Yes.**
  - **No.**

  What happens when regulators and industry collaborate on technical specifications?

  This field is "Yes" if technical standards are either developed internally by the regulator, arrived at through collaboration of the regulator with industry participants, or mandated by the regulator to be developed by industry participants.

A.2.4 Related regulations

- **open_data_regulation**: Is there past, present, or ongoing regulation related to open data?
  - **Yes.**
  - **No.**
- **data_privacy_regulation**: Is there past, present, or ongoing regulation related to data privacy or rights?
  - **Yes.**
  - **No.**

  What is the difference between open data and data privacy?

  Open data refers to laws requiring data owners to make customer data public to other commercial entities. Data privacy laws (e.g., GDPR) regulate what data holders (e.g., banks) may do with the customer data they possess.

A.2.5 Open banking scope

- **financial_services_scope**: How wide is the set of financial products covered under OB?
  - **Narrow**: Transaction accounts only.

- **Broad**: Transaction accounts and other "core" financial products (e.g., loans).
- **Very broad**: Above products plus "non-core" financial products (e.g., insurance).

- **transaction_accounts_covered**: Does the regulation cover transaction accounts?
  - **Yes.**
  - **No.**
- **nontransaction_accounts_covered**: Does the regulator cover financial products aside from transaction accounts?
  - **Yes.**
  - **No.**
- **share_account_data**: Does the regulator either require or facilitate the sharing of customers' transaction account data?
  - **Yes.**
  - **No.**
- **payment_initiation**: Does the regulator require or facilitate technology to allow the initiation of customer payments by third parties?
  - **Yes.**
  - **No.**

What do we include in transaction accounts?

Any financial account that allows for cash-like transactions, e.g., checking accounts, debit cards, credit cards, and digital wallets.

What are core and non-core financial products?

Core products are consumer financial products that banks typically offer, including, e.g., loans or investment services. Non-core products are either consumer finance products that banks do not typically offer, e.g., insurance, or financial products that are not "consumer" finance products, such as small business loans.

Is a payment service like Venmo or Alipay an OB transaction service?

No, these services do not rely on open APIs interfacing with banks. See the definition of an OB approach above.

A.2.6 Sharing scope and reciprocity

- **data_holders_share**: Do data holders (e.g., banks) have to share their customers' data (upon customer request)?
  - **Yes.**
  - **No.**
- **data_users_share**: Do data users (e.g., fintechs) have to share their customers' data (upon customer request)?
  - **Yes.**
  - **No.**

A.2.7 Miscellaneous

- **PSD2**: Is this country a party to Europe's PSD2?
  - **Yes.**
  - **No.**

A.3 Miscellaneous notes

How do we define scope, sharing rules, and so on in cases where the regulators have not yet decided on an approach?

We denote these cases as "TBD" and exclude them from sections of the analysis where we split or condition on these variables.

Has Iceland adopted/implemented PSD2?

As of October 2021, Iceland has not implemented PSD2.

B Classification of Fintech Startups

PitchBook groups tens of thousands of startups into the "Financial Software" subindustry and the "Fintech" vertical, but does not offer a more granular industry definition. We overcome this using PitchBook's keywords feature with categories from PitchBook's 2021Q1 fintech market map and keywords derived from those startups. PitchBook's fintech market map divides recent fintech financing rounds into eight broad categories: alternative lending, capital markets, consumer finance, digital assets, financial services IT, payments, regtech, and wealthtech. Importantly, these categories were designed around use cases and without OB in mind.

Although innovative startups are by nature often hard to classify, these categories roughly span the current fintech market. Alternative lending includes retail and commercial lending. Capital markets includes institution-focused capital market applications, including trading, data, and capital management. Consumer finance encompasses digital banking, rewards programs, and credit cards. Digital assets covers cryptocurrency and related applications. Financial services IT includes both APIs and enterprise architecture. Regtech includes risk management and compliance startups. Wealth management includes investment advisory and brokerage services.

For each of those categories, we derive a list of keywords used by the startups in that category. These keywords were assigned by PitchBook analysts covering the company, with the typical company having four keywords. Keywords range from general to specific, for example, the most frequently used keywords for companies in the regtech segment of the market map are regtech vertical, fraud detection, fraud detection platform, regulatory compliance, fintech, artificial intelligence, and risk management.

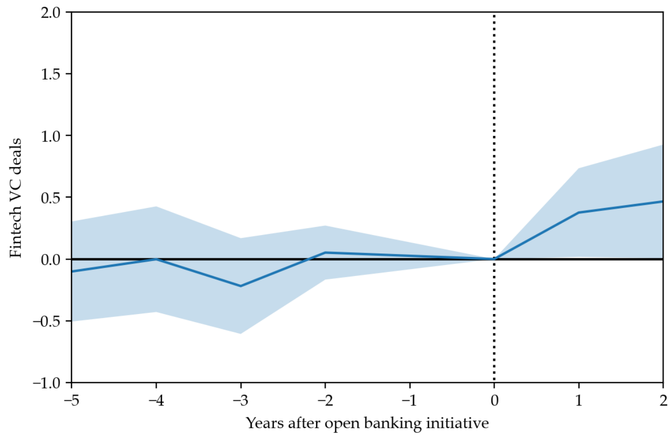
We find the relative frequency of each keyword within each category. For example, the regtech vertical keyword accounts for 3% of the keywords used by startups in the regtech category and less than 1% for all the other categories. A keyword is distinctive to a category if it is in the top 25 keywords for that category and its usage rate in that category is twice the sum of its usage rates in the other categories. Regtech vertical, fraud detection, fraud detection platform, and regulatory compliance are all distinctive keywords for the regtech category. Fintech, artificial intelligence, and risk management are not because they are commonly used across categories. The capital markets category focuses on institutional services and lacks distinctive keywords (its top keywords are financial technology, financial software, financial platform, and financial services) and so we drop it.

We assign fintech startups into categories using the distinctive keywords for each category. A startup is classified as a regtech startup if it is marked with regtech vertical, fraud detection, fraud detection platform, regulatory compliance, or other distinctive keywords for the regtech category. Fintech companies often offer a broad scope of services and can be hard to assign to a single category. Our keyword-based classification system accommodates this by allowing companies to be in multiple categories. For example, the US company SeedFi offers packages of borrowing and saving to lower-income customers placing it in both the alternative lending and consumer finance categories. The resulting categories are relatively balanced, with the largest categories (wealth management, financial IT) being about two-and-a-half times as large as the smallest category (consumer finance).

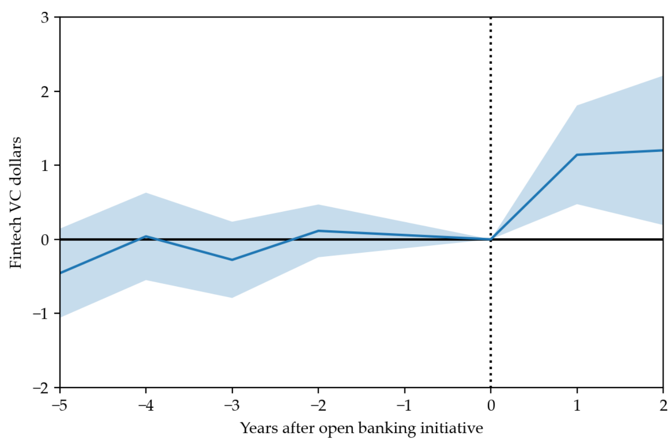
C Additional Tables and Figures

Figure C1: Event-study of Fintech Investment After OB Controlling for Non-Fintech VC

Note: This figure shows changes in fintech venture capital (VC) activity around the passage of open banking government policies using a panel event-study analysis that includes a control for non-fintech VC activity. We perform this analysis on our high-coverage Pitchbook panel of 2011-2021 data for the 21 countries with at least five fintech VC deals in the 2000–2010 period. Panel (a) shows an event study on the log of one plus the number of fintech VC deals, and Panel (b) shows an event study on the log of one plus the millions of US dollars invested in fintech VC deals. Year 0 is the passage year of each country's major open banking initiative. The coefficient for year 0 is set to zero and other coefficients are presented net of country and region-by-year fixed effects based on the World Bank regions. European Union member states are weighted to count as a country for estimates and standard errors. The shaded regions denote 95% confidence intervals calculated using standard errors clustered at the country level.



(a) Log of number of fintech VC deals



(b) Log of amount of fintech VC investment in millions of US dollars

Table C1: Country Data Summary Statistics

Note: This table presents summary statistics on country-year variables. Panel (a) reports values for all countries and the 2013 year, which we use for pre-open banking country characteristics for our cross-country regressions. Panel (b) reports values for 2011–2021 for our high-coverage PitchBook sample of countries that have at least five fintech venture capital (VC) deals in the 2000–2010 period, which we use for panel regressions of open banking's impact on fintech VC activity. For each variable, we present the number of observations, the average value, the standard deviation, and assorted percentiles. The first set of variables concern the status of open banking. After open banking (OB) initiative is equal to one in country-years after a major open banking policy was passed. The next three variables are set at the country level based on that country's OB policies as of October 2011: OB implemented is an indicator variable equal to one if the open banking policy was implemented or is in the pre-implementation stage, OB implementation is a 0-7 rating of the open banking policy progress where higher numbers denote more progress toward regulation, and the OB Strength Index is our 0-1 measure of open banking policy strength. VC deals, non-fintech VC deals, and fintech VC deals are presented next and are from PitchBook and used after taking the log of one plus the number (and are hence different from Table C3). Per capita GDP in thousands of US dollars, the square of per capita GDP in hundreds of thousands of US dollars, the log of population (in millions), private sector credit to GDP, bank branches per 100k people, and the financial sector Lerner index are from the World Bank. The Lerner index ranges between 0 and 1 and measures the market power of banks, with higher values denoting less competition. Foreign-owned banks are from the Claessens and Van Horen (2013) foreign bank ownership data. The Rule of Law and Business Regulation Indexes are from the Cato Institute and are on a 0 to 10 scale with higher numbers denoting more favorable conditions. The Corruption Perception Index is from Transparency International and is on a 0 to 100 scale with higher numbers denoting more favorable conditions.

Panel (a) 2013 for entire 168 country sample

	Count	Mean	Std. dev.	10th pct.	25th	50th.	75th	90th
Observations	168							
Open banking variables								
After open banking initiative	168	0.00	0.00	0.00	0.00	0.00	0.00	0.00
OB implemented (as of 2021)	168	0.29	0.45	0.00	0.00	0.00	1.00	1.00
OB implementation (as of 2021)	168	1.83	2.46	0.00	0.00	0.00	4.00	6.00
OB Strength Index (as of 2021)	168	0.15	0.26	0.00	0.00	0.00	0.25	0.50
Venture capital variables								
VC deals	168	1.58	1.97	0.00	0.00	0.69	2.79	4.79
Non-fintech VC deals	168	1.54	1.95	0.00	0.00	0.69	2.66	4.67
Fintech VC deals	168	0.58	1.06	0.00	0.00	0.00	1.10	2.11
Other explanatory variables								
Per capita GDP ($k)	162	14.74	20.69	0.76	1.41	5.62	18.75	45.63
Per capita GDP ($100k) squared	162	0.06	0.17	0.00	0.00	0.00	0.04	0.21
Log population	155	2.37	1.50	0.60	1.35	2.31	3.38	4.32
Private sector credit to GDP	149	54.19	47.37	11.08	18.70	39.79	70.53	124.82
Branches per 100k people	155	17.30	15.63	2.72	4.94	12.48	23.54	37.15
Financial sector Lerner index	94	0.31	0.13	0.17	0.23	0.29	0.37	0.47
Foreign-owned banks	134	0.43	0.28	0.03	0.20	0.42	0.66	0.80
Rule of Law Index	146	5.18	1.56	3.43	3.94	4.77	6.47	7.56
Business Regulation Index	146	6.39	1.29	4.72	5.72	6.33	7.35	8.09
Corruption Perception Index	160	42.27	19.97	20.90	28.00	37.00	53.25	75.00

Table C1: COUNTRY DATA SUMMARY STATISTICS (CONTINUED)

Panel (b) 2011–2021 for the 21 country high-coverage sample

	Count	Mean	Std. dev.	10th pct.	25th	50th.	75th	90th
Observations	231							
Open banking variables								
After open banking initiative	231	0.24	0.43	0.00	0.00	0.00	0.00	1.00
OB implemented (as of 2021)	231	0.81	0.39	0.00	1.00	1.00	1.00	1.00
OB implementation (as of 2021)	231	4.81	1.89	1.00	4.00	6.00	6.00	6.00
OB Strength Index	231	0.43	0.34	0.25	0.25	0.25	0.50	1.00
Venture capital variables								
VC deals	231	5.75	1.22	4.47	4.92	5.53	6.31	7.46
Non-fintech VC deals	231	5.67	1.21	4.36	4.82	5.42	6.21	7.29
Fintech VC deals	231	3.19	1.38	1.61	2.30	3.00	3.88	5.10
Other explanatory variables								
Per capita GDP ($k)	220	42.28	20.57	10.66	32.31	44.82	52.82	63.09
Per capita GDP ($100k) squared	220	0.22	0.18	0.01	0.10	0.20	0.28	0.40
Log population	220	3.57	1.52	1.68	2.24	3.64	4.52	5.39
Private sector credit to GDP	199	107.33	44.47	51.88	65.25	105.49	141.13	167.65
Branches per 100k people	197	24.86	13.72	8.92	14.79	21.86	32.93	38.36
Financial sector Lerner index	72	0.27	0.12	0.11	0.18	0.28	0.34	0.41
Foreign-owned banks	231	0.27	0.24	0.02	0.08	0.20	0.40	0.58
Rule of Law Index	126	7.08	1.44	4.35	6.78	7.45	8.18	8.61
Business Regulation Index	126	7.28	1.12	6.00	6.79	7.64	8.03	8.18
Corruption Perception Index	168	68.75	17.98	38.00	60.00	74.50	82.00	87.30

Table C2: Drivers of Implementation Progress of Open Banking Government Policies

Note: This table shows whether ex-ante country characteristics predict the extent of implementation of open banking government policies. The dependent variable is a score between 0 and 7 based on a country's open banking implementation progress as of October 2021, with 0 being no action, 1–2 being increasingly serious levels of discussion, and 3–7 being levels of implementation progress. The independent variables are country characteristics measured as of 2013. Private sector credit to GDP, bank branches per 100k people, and the financial sector Lerner index are from the World Bank. Non-fintech and fintech VC deals are from PitchBook and are used after taking the log of one plus the number of deals. Foreign-owned banks are from the Claessens and Van Horen (2013) foreign bank ownership data. The Rule of Law and Business Regulation Indexes are from the Cato Institute and are on a 0 to 10 scale with higher numbers denoting more favorable conditions. The Corruption Perception Index is from Transparency International and is on a 0 to 100 scale with higher numbers denoting more favorable conditions. All specifications include GDP per capita in thousands of US dollars, the square of GDP per capita in hundreds of thousands of US dollars, the log of population, and region fixed effects as controls, all based on World Bank data. European Union member states are weighted to count as a single country for estimates and standard errors. The regressions are cross-sectional, where each country in the sample corresponds to a single data point. *** denotes p-value < 0.01, ** denotes <0.05, and * denotes <0.1.

	OB implementation (0-7)				
	(1)	(2)	(3)	(4)	(5)
Private sector credit to GDP	0.005				-0.006
	(0.007)				(0.006)
Branches per 100k people	-0.021				-0.029
	(0.021)				(0.021)
Financial sector Lerner index	3.235				2.674*
	(1.969)				(1.366)
Non-fintech VC deals		0.587***			0.440*
		(0.192)			(0.254)
Fintech VC deals		0.088			0.523
		(0.315)			(0.348)
Foreign-owned banks			0.497		0.134
			(0.581)		(0.660)
Rule of Law Index				0.331	0.120
				(0.215)	(0.387)
Business Regulation Index				0.077	-0.061
				(0.185)	(0.192)
Corruption Perception Index				0.015	0.038
				(0.021)	(0.029)
Per capita GDP ($k)	0.093*	0.048	0.136***	0.074*	-0.038
	(0.047)	(0.031)	(0.030)	(0.038)	(0.034)
Per capita GDP ($100k) squared	-8.003*	-4.506	-11.431***	-7.632*	1.676
	(4.586)	(3.632)	(3.627)	(3.914)	(3.064)
Log population	0.389**	-0.141*	0.283**	0.323***	-0.145
	(0.186)	(0.077)	(0.118)	(0.110)	(0.173)
Region FE	Yes	Yes	Yes	Yes	Yes
Observations	85	162	133	145	82
Adjusted R^2	0.445	0.514	0.424	0.481	0.617

Table C3: PitchBook Data Summary Statistics

Note: This table presents summary statistics on our PitchBook venture capital (VC) deal data for 168 countries over 2000–2021. The first column presents statistics on the entire dataset, the next two columns present data for 2000–2010 and 2011–2021 for low coverage countries, and the final two columns present data for 2000-2010 and 2011-2021 for high-coverage countries. High-coverage countries are those with five or more fintech VC deals in the 2000–2010 period, while countries with fewer than five are low-coverage countries. The first set of rows presents the number of countries in each sample, both those with open banking implemented or in the pre-implementation stage as of October 2021 and those that have not reached that stage. The second set of rows presents the number of country-year observations in each sample, both those that are after an open banking policy was passed in that country and other observations. The third set of rows presents statistics on country-year VC investment: any VC deals indicates the percentage of country-years with a VC deal, mean and median raw VC deals present the average number of deals in country-years, and mean and median raw VC dollars ($m) presents the average value of VC deals in a country-year in millions of US dollars. The fourth set of rows presents similar statistics on the country-year fintech VC investment.

	All countries	Low-coverage countries		High-coverage countries	
	2000-2021	2000-2010	2011-2021	2000-2010	2011-2021
Countries					
Count of countries	168	147	147	21	21
Countries with open banking implemented	49	32	32	17	17
Countries without open banking implemented	119	115	115	4	4
Country-year observations					
Count of country-year observations	3,696	1,617	1,617	231	231
Country-years after open banking passed	139	0	84	0	55
Country-years before open banking passed	3,557	1,617	1,533	231	176
Country-year VC activity					
Any VC deals (%)	44.6	23.4	50.1	99.1	100.0
Mean raw VC deals	74.5	1.1	13.0	212.7	880.6
Median raw VC deals	0.0	0.0	1.0	38.0	251.0
Mean raw VC dollars ($m)	718.1	6.3	85.8	1,725.2	9,119.4
Median raw VC dollars ($m)	0.0	0.0	0.0	170.6	1,109.9
Country-year fintech VC activity					
Any fintech VC deals (%)	25.3	3.2	31.4	64.1	98.7
Mean raw fintech VC deals	6.0	0.0	1.8	8.4	74.8
Median raw fintech VC deals	0.0	0.0	0.0	1.0	19.0
Mean raw fintech VC dollars ($m)	81.2	0.1	20.9	66.8	1,085.9
Median raw fintech VC dollars ($m)	0.0	0.0	0.0	1.3	88.6

Table C4: Leave-one-out Country Effect of Open Banking Government Policy on Fintechs

Note: This table shows how our main estimate of the effect of open banking on fintech VC activity—the coefficient on "After OB initiative" in column 2 of Table 4—varies when we rerun this specification after excluding one country at a time from our sample. Each row corresponds to a different regression sample that is equal to our high-coverage Pitchbook panel data of country-year data spanning 2011-2021 for the 21 countries with at least five fintech deals in the 2000–2010 period, excluding the indicated country for that row for the first 21 rows and excluding both France and Germany for the final row. The Coefficient column presents the coefficient on post-open banking (parameter on After OB initiative) estimated using a difference-in-difference design on that sample, with the Standard error, t stat. and p-value columns similarly presenting their respective statistics. All specifications include country fixed effects and region-by-year fixed effects, where regions are based on the World Bank classification. European Union member states are weighted to count as a single country for estimates and standard errors. Standard errors are clustered at the country level. *** denotes p-value < 0.01, ** denotes <0.05, and * denotes <0.1.

	Coefficient	Standard error	t stat.	p-value
Excluding AUS	0.497**	0.162	3.065	0.012
Excluding BEL	0.535***	0.148	3.612	0.004
Excluding BRA	0.532***	0.163	3.259	0.009
Excluding CAN	0.532***	0.163	3.259	0.009
Excluding CHN	0.455**	0.203	2.244	0.049
Excluding DEU	0.539***	0.162	3.324	0.007
Excluding DNK	0.531***	0.163	3.261	0.008
Excluding ESP	0.524***	0.159	3.301	0.007
Excluding FIN	0.538***	0.159	3.394	0.006
Excluding FRA	0.533***	0.160	3.327	0.007
Excluding GBR	0.681***	0.049	13.859	0.000
Excluding IND	0.532***	0.163	3.259	0.009
Excluding IRL	0.550***	0.166	3.317	0.007
Excluding ISR	0.532***	0.163	3.259	0.009
Excluding JPN	0.537**	0.176	3.051	0.012
Excluding NLD	0.530**	0.170	3.127	0.010
Excluding NOR	0.535**	0.184	2.908	0.016
Excluding POL	0.527**	0.182	2.890	0.015
Excluding RUS	0.453**	0.195	2.318	0.043
Excluding SWE	0.519***	0.151	3.441	0.006
Excluding USA	0.532***	0.163	3.259	0.009
Excluding DEU and FRA	0.541***	0.162	3.329	0.007

Table C5: Effect of Open Banking Government Policy on Fintechs Using IHS Transform

Note: This table shows changes in fintech venture capital (VC) investment following the implementation of open banking government policies. The table uses a difference-in-difference design on our high-coverage Pitchbook panel of country-year data spanning 2011-2021 for the 21 countries with at least five fintech deals in the 2000–2010 period. The dependent variable in columns 1 to 3 is the inverse hyperbolic sine (IHS) of the number of fintech deals in a country-year, and in columns 4 to 6 it is the IHS of the amount invested in millions of US dollars. The independent variable is a dummy variable equal to one if the year in question is after the year major open banking laws were passed in the country in question. Columns 3 and 6 include a control for non-fintech VC activity using data from Pitchbook, transformed the same way as fintech VC activity. All specifications control for country fixed effects, columns 1 and 4 contain controls for year fixed effects, and columns 2, 3, 5, and 6 control for region-by-year fixed effects, where regions are based on the World Bank classification. European Union member states are weighted to count as a single country for estimates and standard errors. Standard errors are clustered at the country level. *** denotes p-value < 0.01, ** denotes <0.05, and * denotes <0.1.

	IHS fintech VC deals			IHS fintech VC dollars		
	(1)	(2)	(3)	(4)	(5)	(6)
After OB initiative	0.228*	0.538**	0.370*	0.771**	1.313**	1.110**
	(0.114)	(0.190)	(0.182)	(0.283)	(0.520)	(0.478)
IHS non-fintech VC deals			0.473***			
			(0.132)			
IHS non-fintech VC dollars						0.293***
						(0.087)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes			Yes		
Region-Year FE		Yes	Yes		Yes	Yes
Observations	231	231	231	231	231	231
Adjusted R^2	0.908	0.916	0.921	0.871	0.886	0.889

Table C6: Effect of Open Banking Government Policy on Non-Fintech VC

Note: This table shows changes in non-fintech venture capital (VC) investment following the implementation of open banking government policies as a placebo test counterpart to Table 4. The table uses a difference-in-difference design on a Pitchbook panel of country-year data spanning 2011-2021 for countries with at least five non-fintech deals in the 2000–2010 period. The dependent variable in columns 1 to 2 is the log of one plus the number of non-fintech deals in a country-year, and in columns 3 to 4 it is the log of one plus the amount invested in millions of US dollars. The independent variable is an indicator variable equal to one if the year in question is after the year major open banking laws were passed in the country in question. All specifications control for country fixed effects; columns 1 and 3 contain controls for year fixed effects; and columns 2 and 4 control for region-by-year fixed effects, where regions are based on the World Bank classification. European Union member states are weighted to count as a single country for estimates and standard errors. Standard errors are clustered at the country level. *** denotes p-value < 0.01, ** denotes <0.05, and * denotes <0.1.

	Non-fintech VC deals		Non-fintech VC dollars	
	(1)	(2)	(3)	(4)
After OB initiative	-0.042	0.045	0.136	0.104
	(0.079)	(0.107)	(0.174)	(0.194)
Country FE	Yes	Yes	Yes	Yes
Year FE	Yes		Yes	
Region-Year FE		Yes		Yes
Observations	748	748	748	748
Adjusted R^2	0.953	0.958	0.888	0.895

D Additional Model Discussion

D.1 Model Extension: Consumer Data Sharing Choices

In this appendix, we extend the model to accommodate consumers that strategically disclose or hide their financial information. Under an OB regime, we allow banked consumers to choose whether to reveal their data to other firms in the second period. The consumer makes this decision by comparing her expected utility if she shares her data and gets targeted pricing or if she does not share and gets targeted pricing only from her relationship bank. We enrich this decision along two dimensions. First, we allow the customer to value privacy per se. Second, we introduce noise into the customer's decision, reflecting well-documented facts regarding the ineffectiveness of consumer disclosures, particularly around financial products.[36] Following a similar discrete choice framework, we model the consumer's indirect utility of sharing or not sharing her data as

$$\begin{aligned} u_i^S &= -\phi + Eu^S(\boldsymbol{\nu}_i^S, \chi_i) + \epsilon_i^S \\ u_i^N &= Eu^N(\boldsymbol{\nu}_i^N, \chi_i) \end{aligned}$$

where ϕ represents her hedonic privacy valuation. Eu^S and Eu^N are her expected utilities in the data-using period if she shares or does not share, respectively. ϵ_i^S is an unmodeled shock around her attention to consumer disclosures. Choosing the greater of these utilities yields an endogenous probability of disclosing given by ψ_i.

Critically, the equilibrium prices $\boldsymbol{\nu}_i$ depend on the distribution of consumers in each information environment and as a result of the consumer's choice described above, this distribution is endogenously distorted away from the unconditional distribution $dF(\chi_i)$. Thus, all optimal prices in the sharing (S) and not sharing (N) regimes must be calculated under the equilibrium distributions conditional on the chosen information regime, $dF(\chi_i|S)$ and $dF(\chi|N)$. These are given by

$$\begin{aligned} dF(\chi_i|S) &\propto \phi_i \times dF(\chi_i) \\ dF(\chi_i|N) &\propto (1-\phi_i) \times dF(\chi_i). \end{aligned}$$

The equilibrium endogenous selection into data sharing is thus characterized by a set of probabilities ψ_i such that (1) optimal pricing $\boldsymbol{\nu}_i^S$ and $\boldsymbol{\nu}_i^N$ is consistent with the borrower type distributions induced by the selection probabilities, and (2) borrower choices, given these prices, are consistent with the probabilities.

Numerically, solving for this equilibrium can be accomplished as follows: First, conjecture for each borrower type a probability $\hat{\psi}_i$ that she will opt into data sharing. Second, calculate

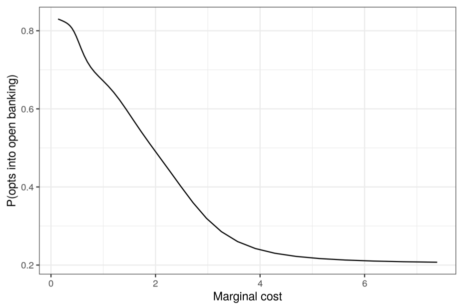
[36] See, for example, Ben-Shahar and Schneider (2011).

the equilibrium outcomes under the implied borrower type distributions in the sharing and not sharing regimes, including, importantly, borrower expected utility. Given these expected utilities under the conjectured opt-in probabilities, calculate the actual opt-in probabilities using the 'market share' equation given above, $\tilde{\psi}_i$. Finally, update the initial guess $\hat{\psi}'_i$ to be closer to the calculated $\tilde{\psi}_i$ and iterate until they converge.

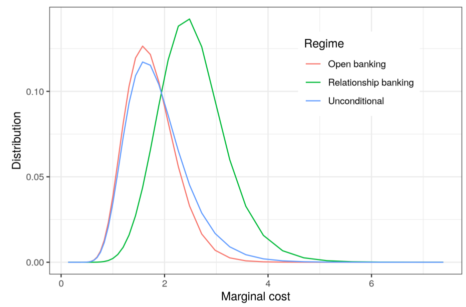
Figure D1 shows borrowers' optimal decisions to opt into data sharing, the resulting borrower distributions, and borrower expected utilities versus borrower marginal cost. Not surprisingly, lower marginal cost (i.e., lower default probability) borrowers endogenously opt in to data sharing because they benefit from lenders knowing their type, while higher marginal cost do not. This shifts the distribution of borrowers in the data-sharing regime to the left and the distribution of borrowers in the non-data-sharing regime to the right, meaning that the non-sharing pool becomes worse than the unconditional pool. In terms of utilities, the high-quality borrowers benefit from opting in to open banking and sharing their data, and the low-quality borrowers are hurt even though they refuse to share, because they can no longer pool with the high-quality borrowers and, therefore, receive higher interest rates. In sum, allowing consumers to opt out of data sharing does not change our main takeaways in Section 4, as opting out partially "unravels" (Grossman, 1981).

Figure D1: Consumer Data Sharing Decisions

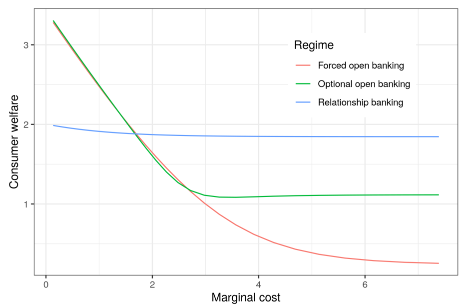
Note: This figure shows borrowers' optimal decisions to opt into data sharing (Panel (a)), the resulting borrower distributions (Panel (b)), and borrower expected utilities versus borrower marginal cost (Panel (c)). We consider a calibration based on the US mortgage market.



Panel (a) Opt-in probabilities



Panel (b) Marginal cost distributions



Panel (c) Borrower expected utilities

D.2 Model Calibration

This appendix section details the back-of-the-envelope calibration of our structural models of (1) the US non-GSE residential mortgages and (2) consumer financial advice. In both cases, we impose common structural assumptions on consumer heterogeneity. The horizontal taste shocks in period one and two, ϵ_{ij}^p and ϵ_{ij}, respectively, follow a type-one extreme value distribution.[37] Consumer preference parameters and marginal costs are distributed independently as $\log \alpha_i \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2)$, $f_i \sim \mathcal{N}(0, \sigma_f^2)$, and $\log mc_i \sim \mathcal{N}(\mu_{mc}, \sigma_{mc}^2)$. The chief challenge for calibration or estimation is to parameterize these distributions. Fortunately, a robust structural finance IO literature has done exactly that.

In the residential mortgage context, we view traditional banks as incumbents and fintechs as entrants. Buchak et al. (2018a) recover the distribution of price sensitivities, $dF(\alpha_i)$, which we adopt here. We calibrate $dF(mc_i)$ through average default rates, and reflecting the findings in Buchak et al. (2018a), we assume that both banks and fintechs face the same marginal cost $mc_{bank} = mc_{fintech}$. Finally, we specialize the value of non-price characteristics, $\delta(g_{ij}, f_i)$, to equal θ if the lender has a previous relationship with the borrower and zero otherwise, reflecting an own-bank preference. Note that this assumption turns off product customization, consistent with the lack of heterogeneity in the highly standardized US mortgage contract. We set the remaining three parameters in the data use stage—the outside option utility, entry cost, and own bank preference—using the method of moments to approximately match the number of lenders, own bank share, and total fraction of households obtaining mortgages.

In the financial adviser context, Di Maggio et al. (2021) recover the distribution of price sensitivities, $dF(\alpha_i)$, which we adopt here. We assume a heterogeneous marginal cost in providing the service, with $mc_{bank} = 1.5\%$ and $mc_{fintech} = 0.35\%$.[38] We specialize $\delta(g_{ij}, f_i) = -\lambda(f_i - g_{ij}j)^2$, where f_i is the "optimal" advice for the consumer, g_{ij} is the (potentially customized) financial advice, and λ is the utility cost of receiving suboptimal advice. Di Maggio et al. (2021) provide the preference parameters over receiving sophisticated financial advice, and we set λ and σ_f^2 so that compared to the average consumer receiving untargeted advice, perfectly targeted advice confers this equivalent utility benefit. As before, we set the outside option utility, the entry cost, and own bank presence to match aggregate market shares, bank market shares, and the number of service providers.

Finally, for the data production stage, we broadly adopt the calibrated parameters in Egan et al. (2017) regarding the number of banks and deposit interest rate sensitivity. As a baseline assumption, we assume that customers are myopic and do not anticipate the value their data brings them in period two. We relax this assumption in comparative statics.

[37] This is a common distributional assumption in models of discrete choice and yields highly tractable market share and pricing equations. See, e.g., in the finance context, Buchak et al. (2018a).

[38] These values reflect reported fees on JPMorgan's website for automated versus particularized financial advice, respectively.