Online Annex 3.1. Case Study on Neobanks

Data description

The data set comprises 37 neobanks and 640 traditional banks across 18 economies, including in Europe (UK, France, Germany, Italy, Spain, Lithuania, Poland, and Russia), Asia (Kazakhstan, Japan, Hong Kong SAR, Indonesia, Korea, and China) and the Western Hemisphere (Canada, US, Brazil, and Mexico).

Assembling the data required the following steps:

Step 1—to identify the universe of neobanks, we employed several sources, including NeoBanks.app (The list of neobanks and digital banks in the world in 2022), WhiteSight (2021), S&P Global Market Intelligence (banks and savings banks tagged as "neobanks"), market reports on both unlisted and listed neobanks and digital banks (Goldman Sachs 2022, Morgan Stanley 2021, and S&P Capital IQ 2021), policy notes (Clerc and others 2020) and extensive conversations with bank analysts, investors and (including former) management of a number of neobanks and traditional banks.

Step 2—the initial sample (250) was narrowed down to those neobanks that make their financial statements available and report them on a comparable format to banks. We then pulled their consolidated financial data from S&P Global Market Intelligence or, alternatively, from their latest publicly available filings (quarterly/semi-annual reports, Pillar-3 reports, etc.) for the two most recent full fiscal years (generally 2019 and 2020). In the case of listed neobanks, we also relied on sell-side research reports and models. The result of this was a sample of 37 neobanks in 18 economies.

Step 3—For each of the 18 economies identified in the previous step, we collected the consolidated financial data for all active banks, both listed and unlisted, that are available in S&P Global Market Intelligence. We excluded neobanks identified in the first step. In the UK, we also excluded consolidated Group data and kept their UK business, whenever we are given this option. In the US, we just kept the "midcap banks" (as defined by Morgan Stanley Research). All in all, we collected 640 traditional banks across 18 economies.

Methodology

We benchmarked each neobank against the universe of its respective local traditional banks, following three steps:

Step 1—We calculated, for each institution, key performance metrics across nine dimensions: yields, spreads, margins, efficiency, credit quality, capital, liquidity, profitability, and growth.

Step 2—We used traditional bank data to compute, for every single metric and individual economy, both the (asset-weighted) average and the standard deviation. We then computed the distance of each neobank to the (asset-weighted) average of its local (traditional) peer group, per metric, as number of standard deviations. This data transformation allows us to make cross-country comparisons.

Step 3—To explore differences across geographies and types of institutions, we also split the sample per region (Europe, Asia, Western Hemisphere), type of economy (AE vs EM), year of origin (pre and post 2010[1]) and return profile (profitable vs loss-making), and compute median values for each of them.

[1] The "neobank" concept originated in the early 2010s alongside the advent of policies supporting open banking around the World (Microsoft, Linklaters and Accenture 2019); it is different to the "direct bank" (branchless bank) concept that had been around since at least the early 1980s (Exton Research 2021, WhiteSight 2021).

Online Annex 3.2. Case Study: US Mortgage Market

The annex describes the data used in the analysis, some additional stylized facts, and presents detailed regression results and robustness checks.

Data

At the heart of the analysis is the detailed Home Mortgage Disclosure Act (HMDA) data. The dataset contains a number of variables (characteristics) for almost all mortgage applications made in the US.[2]

The HMDA data is available on an annual (calendar year) basis from the US Consumer Financial Protection Bureau. The analysis is based on the data from 2007–20, which was the latest available data at the time of publication.[3] The number of recorded data points was significantly expanded in 2018, including the reporting of loan-to-value ratios, the age profile, interest rates and various details on loan terms. Descriptions of the individual data points are available from the dedicated FFIEC website.[4]

The analysis uses only originated mortgages, identified in the data as approved and accepted mortgage applications. The sample is further restricted to mortgages for home purchases or refinancing, first lien mortgages,[5] and loans for 1–4 family homes. This results in a total of about 110 million observations.

Identification of fintechs, non-banks, banks and credit unions

The identification of fintechs (fintech mortgage originators) follows the definition in the literature, including some recent updates. The starting point is the seminal paper of Buchak and others (2019), who identify fintechs as those offering an online application process without the need of human interaction. Those are Quicken Loans (Rocket Mortgage), Amerisave, Guaranteed Rate, Cashcall, Homeward Residential/PHH Mortgage Corporation, and Movement Mortgage. Jagtiani and others (2021) add two additional firms that started to fully operate in 2016 or later: Better Mortgage and SoFi Mortgage. In total, these eight fintechs are responsible for about 11 percent of mortgage origination in 2020 and yield a total of 6.7 million origination observations (all fintechs are non-banks by definition).

Following the literature, non-banks are defined as all non-depository institutions, identified in the dataset as all reporters with an agency code equal to 7 (regulatory agency is the US Department of Housing and Urban Development (HUD)).

Banks are all depository institutions with a regulatory agency that is a bank supervisor (agency codes 1, 2, 3 and 9). Credit Unions are identified as all institutions regulated by the National Credit Union Administration (NCUA).

---

[2] Banks that have no offices in non-metropolitan areas and those below a certain asset threshold do not have to report their mortgage applications. Non-banks (non-depository institutions) are covered as long as they originate a significant number of mortgage loans (currently >100 closed-end mortgages or >200 open-end credit lines, but these thresholds have varied over time). See https://files.consumerfinance.gov/f/documents/cfpb_2022-hmda-insitutional-coverage_03-2021.pdf.

[3] Historical data is available at https://www.consumerfinance.gov/data-research/hmda/historic-data/, whereas data from 2017 is available from https://ffiec.cfpb.gov/data-publication/.

[4] https://ffiec.cfpb.gov/documentation/2020/.

[5] Second or higher-lien mortgage applications represent of very small share of the overall sample (<3.5 percent of observations).

Bank data

For the purpose of this case study, banks are defined as all institutions filing regular US CALL reports (FFIEC 031/041). This excludes most credit unions and thrifts, but it includes almost all other types of depository institutions in the US.

Bank data are taken from the US CALL reports, for which data prior to 2010 is available from the Chicago Fed, and the FFIEC from 2011 onwards.[6] Balance sheet variables are measured at year-end, whereas expense and income variables are aggregated from quarterly to annual values where necessary. The data is merged with the HMDA data using the so-called RSSD identifier, as well as the FDIC certifier, and OCC charter numbers.[7]

Geographical mapping

The HMDA provides the census tract, which is then translated into a ZIP code using both the 2010 US Census Bureau data (from 2011 onwards) and the U.S. Department of Housing and Urban Development ZIP code crosswalk files (for prior observations).[8] The ZIP code with the highest overlap with the census tract (based on percent of population or percent of residential addresses) is used.

The data on bank offices is from the FDIC survey of deposits (SOD), which contains the address and ZIP code of all branches of all FDIC-insured banks in the US. To calculate the number of bank branches within a given radius of a borrower, the sample is restricted to physical branches that offer the full range of services (BRSERTYP=11,12).[9]

Distances between the ZIP code of the borrower and bank branches are calculated using the NBER ZIP code distance database.[10]

Additional stylized facts—business models

Bank and fintech mortgage originators exhibit two additional significant differences.

First, while refinancing mortgages make up the largest share of fintech mortgages, banks tend to originate a more similar proportion of mortgages for home purchases and refinancing (Figure 3.1, panel 1). The literature has not reached a definitive conclusion on the causes for this difference (Jagtiani and others 2021), but mortgage refinancing enables fintechs to grow faster (the market for mortgage refinancing is naturally larger than that for home purchases).

[6] Chicago Fed website: https://www.chicagofed.org/banking/financial-institution-reports/commercial-bank-data-complete-2001–2010; FFIEC website: https://cdr.ffiec.gov/public/pws/downloadbulkdata.aspx.

[7] Prior to 2008, the HMDA data does not provide the necessary bank identifiers. The RSSDs, as well as the OCC Charter and FDIC identifiers can, however, be constructed from the Respondent ID in the HMDA data. See https://s3.amazonaws.com/cfpb-hmda-public/prod/help/2017-hmda-fig.pdf#page=14.

[8] Census relationship file: https://www2.census.gov/geo/docs/maps-data/data/rel/zcta_tract_rel_10.txt; HUD USPS ZIP code crosswalk files: https://www.huduser.gov/portal/datasets/usps_crosswalk.html#codebook.

[9] The SOD data is annual as of end of June for each year, while the HMDA data includes all applications until the end of a calendar year. Branches may have closed between end-June and end of the year or new branches may have opened between the beginning of the year and end-June. Cases of branch openings and closures are, however, infrequent.

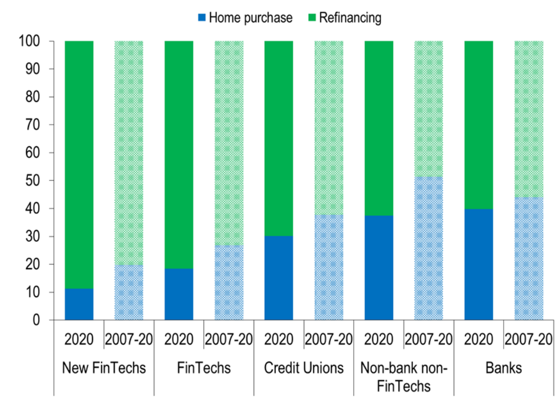
[10] https://www.nber.org/research/data/zip-code-distance-database.

A second key difference is that banks retain a higher share of originated mortgages on their balance sheet (panel 2). This illustrates both the much larger balance sheet and funding capacity of banks as well as the "originate-to-distribute" business model of fintechs. Banks also originate a larger share mortgages with high loans amounts (so-called "jumbo mortgages") which exceed the limits set by the Federal Housing Finance Agency (FHFA) and are therefore not eligible to be purchased, guaranteed, or securitized by the government-backed enterprises Fannie Mae and Freddie Mac.

Figure 3.1 Type of mortgages and securitization

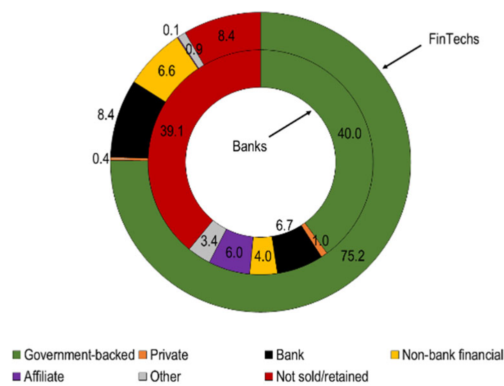
Fintechs specialize in refinancing…

1. Refinancing versus mortgages for home purchases (percent)



…and securitize and sell-off a larger share of originated mortgage loans than banks.

2. Securitization by purchaser type (percent)



Source: US Home Mortgage Disclosure Act data, IMF staff calculations.

Additional stylized facts—access to credit

Fintechs, similar to other non-banks, tend to support access to mortgage credit for potentially underserved borrowers. Non-banks, on average, provided a higher share of loans to borrowers with lower incomes (Figure 3.2, panel 1). This is, at least in part, related to the younger age of borrowers attracted by fintechs. Banks tend to have a particularly high share of borrowers in the top 5 percent of borrower's income, which naturally also tend to have higher loan amounts. A similar picture emerges when looking at average income of borrowers relative to the wider metropolitan area they reside in (panel 2). This is measured at the census tract level, which can be loosely thought of as the wider neighborhood.[11]

[11] A census tract comprises between 1200 and 8000 people. See https://www.census.gov/programs-surveys/geography/about/glossary.html.

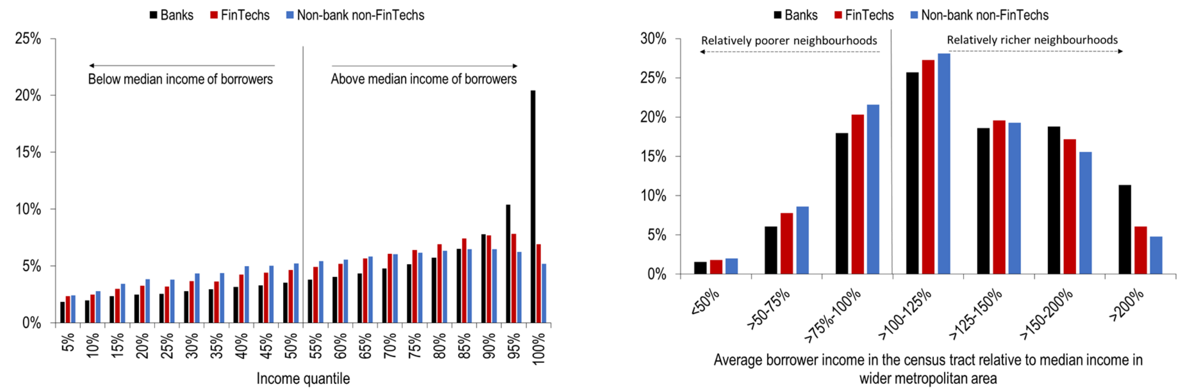
Figure 3.2 Mortgage borrowers by income and affluency of neighborhood

Non-banks, including fintechs, serve a higher share of lower-income borrowers…

1. Share of mortgage origination by borrower income
(percent)

…and borrowers in relatively less affluent neighborhoods (census tracts)

2. Share of mortgage origination by relative income level of the borrower's census tract
(percent)



Source: US Home Mortgage Disclosure Act data, IMF staff calculations.
Note: Based on the entire data sample 2007–2020. Share of originations is based on the loan amount.

Econometric model and variable definitions

The key result for the impact of fintech competitive pressure on banks' mortgage interest income is based on the following model:

$$MortRE_{b,t} = \beta CompPressureFintech_{b,t-1} + \gamma_1 X_{b,t-1} + \varepsilon_{b,t} \quad (1)$$

where $MortRE_{b,t}$ is the return on equity in year t related to bank's b mortgage interest income (not origination income) from loans backed by 1-4 family real estate. $CompPressureFintech_{b,t\leftarrow t-1}$ measures the competitive pressure from fintechs of a given bank's b mortgage origination business, measured at the ZIP code-level z, and aggregated to the bank level *b*. It is defined as $CompPressureFintech_{b,t} = \sum_z (ShareMortgageOrig_{b,z,t} \times \Delta MarketShareFintech_{z,t\leftarrow t-1})$, the share (in percent) of mortgage origination in a given ZIP-code area z for a given bank b in year t multiplied by the change in the market share of fintechs in a ZIP-code area z, aggregated over all ZIP code-areas in which a given bank has mortgage originations. The higher this number, the higher the competitive pressure from fintechs for a given bank. The number varies between -100 (fintechs disappear from all areas where a bank is active) to +100 (fintechs take the entire market in all areas where a given bank is active). X are controls that vary at the bank-level and over time. A key control in X are IT-related expenses (as a share of total expenses).

Results and robustness checks

Table 3.1 presents the key results of the regressions with $MortRE_{b,t}$ as the dependent variable (winsorized at the 1 percent and 99 percent level to limit the effect of outliers). *No Zip codes served* and *HHI* measure a given bank's market diversification and its concentration across ZIP codes respectively. HHI is the Herfindahl-Hirschmann index. Data processing expenditures are bank-level measures (CALL report series RIAD C017), relative to bank equity (consistent with the dependent variable). For models (5) and (6), the sample is restricted to those banks that report these expenditures.[12]

Bank-level controls include the equity ratio (total bank equity to total assets) as well as the deposit ratio (deposits as a share of total non-equity liabilities). Both are highly significant, but do not add much to the explanatory power of the regressions. Unobserved bank-level differences due to, for instance, a persistent difference in size or business model are captured by bank-level fixed-effects. Analogously, common market-related movements in mortgage interest income, due to changes in demand or risk-free rates, are captured by time fixed-effects. The results presented in the main text are based on model (6). The marginal effect of competitive is -0.422. Banks with additional IT expenditures of about 3.37 percent of bank equity (=0.422/0.125) can offset the effect of a 1 percent increase in the fintech composite market share (fintech competitive pressure).

The interaction term in model (7) shows that the marginal effect of competition does not significantly change with data processing expenditures (DPE). Higher DPE does, however, increase mortgage-related income across all specifications and thereby can help to offset the potential impact of competitive pressure from fintechs.[13]

Table 3.1 Baseline regression results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Competitive pressure (t-1)	-3.566*** (1.88e-05)	-2.613*** (0.00336)	-0.422** (0.0482)	-0.395** (0.0441)	-0.391* (0.0502)	-0.422** (0.0380)	-0.417** (0.0397)
No ZIP codes served			0.000443	0.000806*		0.000878*	
HHI			-0.000117	-8.86e-05		-8.70e-05	
Data processing expenditure (DPE)					0.151* (0.0662)	0.125* (0.0555)	0.146* (0.0837)
Competitive pressure (t-1) x DPE							0.0140 (0.388)
Observations	32,411	32,411	22,752	22,289	27,656	21,945	27,656
R-squared	0.014	0.098	0.834	0.850	0.830	0.853	0.830
Year FE		YES	YES	YES	YES	YES	YES
Bank FE		YES	YES	YES	YES	YES	YES
Bank-level controls				YES		YES	

Note: ***, ** and * denotes significance at the 1 percent,5 percent,10 percent level respectively. P-values in parenthesis. Standard errors are double-clustered at the year and bank level.

The additional results for other dependent variables shown in the main text are provided in Table 3.2., models (1) and (4). Models (1)–(3) show the results for the change in the deposit share (percent of total non-equity liabilities) as the dependent variable, whereas models (4)–(6) have the mortgage lending share (percent of total loans) as the left-hand side variable. Otherwise, the specifications are consistent with those in Table 3.1. Even without including proper controls, the effect of competitive pressure from fintechs is not significant.

[12] Generally, if above $100K and >7 percent of a bank’s “other non-interest expenses”.

[13] All results are robust to excluding the period of the Global Financial Crisis (GFC).

Table 3.2 Additional regression results

	(1)	(2)	(3)	(4)	(5)	(6)
Competitive pressure (t-1)	-0.0300 (0.485)	0.0193 (0.483)	-0.0292 (0.626)	-0.721 (0.299)	-0.387 (0.318)	-0.716 (0.386)
No ZIP codes served			-0.000376			0.0135
HHI			3.25e-05			-0.000120
Observations	30,499	30,011	21,872	31,940	31,472	22,288
R-squared	0.063	0.213	0.443	0.000	0.252	0.252
Year FE	YES	YES	YES	YES	YES	YES
Bank FE		YES	YES		YES	YES
Bank-level controls			YES			YES

Note: The dependent variable for models (1)-(3) is the change in the deposit share (percent of total non-equity liabilities) and for models (4)–(6) the change (percent) in the mortgage lending share (percent of total loans).***, ** and * denotes significance at the 1 percent,5 percent,10 percent level respectively. P-values in parenthesis. Standard errors are double-clustered at the year and bank level.

Online Annex 3.3. Risk Analysis on DeFi Lending

This annex describes the data, models, and estimation strategy employed in the analysis of the various risks discussed in the DeFi section, including the details of the event study carried out to assess the risk of cyberattacks.

Probability of Liquidation and Expected Losses

Following Merton (1974) and Black and Cox (1976), a quantitative model is proposed which exploits DeFi user account-level data that includes all assets borrowed and posted as collateral as of end-2021 for selected DeFi platforms: Aave v2, Compound v2, and C.R.E.A.M. Finance on both Ethereum and Polygon network.

Modeling crypto asset portfolio. The dynamics of crypto asset price are modelled as follows.

$$\frac{dp_{k,t}}{p_{k,t}} = \mu_k dt + \sigma_k dW_{k,t} \; k = 1,2,\cdots K, t > 0 \;\; (2)$$

where $\boldsymbol{p}_t = (p_{k,t})_{k=1,2,\cdots,K}$ is the vector of price of crypto asset k at time t, $p_{k,0} > 0, \boldsymbol{\mu} = (\mu_k)_{k=1,2,\cdots,K}$ is the drift, $\boldsymbol{\Sigma} = \mathrm{diag}\big((\sigma_k)_{k=1,2,\cdots,K}\big)$ is the volatility matrix, $W_{k,t}(k = 1,2,\cdots,K)$ is a multi-dimensional standard Brownian motion with $dW_i dW_j = \rho_{i,j} dt$, and $\boldsymbol{\rho} = (\rho_{i,j})_{i,j=1,2,\cdots,K}$ is the correlation matrix.[14]

A user's borrowing value and collateral value can be represented as a portfolio of assets borrowed and posted as collateral. Namely, $V_{b,t} = \sum_{k=1}^{K} n_{b,k,t} p_{k,t}$, where $V_{b,t}$ is the value of borrowing assets, $n_{b,k}$ is the number of borrowing of k-th assets. The dynamics of the borrowing value is $\frac{dV_{b,t}}{V_{b,t}} = \boldsymbol{w}_{b,t} \mathrm{diag}(\boldsymbol{p}_t)^{-1} d\boldsymbol{p}_t$, where $w_{b,t} = (w_{k,b,t})_{k=1,2,\cdots,K}$ is the vector of the share of the borrowed asset k within the total borrowing. Similarly, $V_{c,t} = \sum_{k=1}^{K} n_{c,k,t} p_{k,t}$ where $V_{c,t}$ is the value of collateral assets, $n_{c,k}$ is the number of holding of k-th assets. The dynamics of collateral is $\frac{dV_{c,t}}{V_{c,t}} = \boldsymbol{w}_{c,t} \mathrm{diag}(\boldsymbol{p}_t)^{-1} d\boldsymbol{p}_t$, where $w_{c,t} = (w_{k,c,t})_{k=1,2,\cdots,K}$ is the vector of the share of the collateral k within the total value of collateral posted.

$$\frac{dV_{i,t}}{V_{i,t}} = \mu_i dt + \sigma_i dW_{i,t} \; i \in \{b,c\}, t > 0 \;\; (3)$$

where $\mu_i = \boldsymbol{w}'_{it}\boldsymbol{\mu}$ and $\sigma_i^2 = \boldsymbol{w}'_{i,t}\boldsymbol{\Sigma\Sigma}'\boldsymbol{w}_{i,t}$ for $i \in \{b,c\}$, and $\rho_{bc} = \sqrt{\frac{\boldsymbol{w}'_{b,t}\boldsymbol{\rho\Sigma\Sigma}'\boldsymbol{w}_{c,t}}{\sigma_b^2\sigma_c^2}}$.

Probability of liquidation. Liquidation is triggered when a user fails to maintain the collateral value to meet the collateral requirement. Suppose θ_k is the collateral factor (that is, discounting factor of the collateral set by the DeFi platform), the maximum loanable value of the user is then defined as

[14] It is possible that the drift and volatility may change over time, or there can be jumps in the price that bring discontinuity in the price. Hence the assumption of geometric Brownian motion is strong given the uncertainties regarding crypto asset price. The estimates should be viewed as approximation. By simplifying the model, closed-form formulae of probability of liquidation and expected loss can be derived, which provides economic intuition, such as the relationship between volatilities of assets borrowed and posted as collateral with the riskiness.

$V_{m,t} = \boldsymbol{w}'_{m,t}\boldsymbol{p}_t$ where $V_{m,t}$ is the value of maximum loanable assets, $w_{m,t} = \left(w_{k,c,t}\theta_{k,t}\right)_{k=1,2,\cdots,K}$ is the vector of weight of assets within the total collateral posted by the user multiplied by the collateral factor.

Similar to the collateral value, the dynamics of loanable assets can be written by the following.

$$\frac{dV_{m,t}}{V_{m,t}} = \mu_m dt + \sigma_m dW_{m,t}, t > 0 \quad (4)$$

where $\mu_m = \boldsymbol{w}'_{m,t}\boldsymbol{\mu}$ and $\sigma_m^2 = \boldsymbol{w}'_{m,t}\boldsymbol{\Sigma\Sigma}'\boldsymbol{w}_{m,t}$ and $\rho_{bm} = \sqrt{\frac{\boldsymbol{w}'_{b,t}\boldsymbol{\rho\Sigma\Sigma}'\boldsymbol{w}_{m,t}}{\sigma_b^2\sigma_m^2}}$.

Liquidation is triggered when the health indicator, $Health_t = V_{m,t}/V_{b,t}$ falls below 1. The timing of triggering liquidation, $\tau > t$, where the health falls and passes the threshold, is stochastic. The probability of triggering liquidation before the horizon T ($PL(t,T)$) can be estimable by the following equation,

$$PL(t,T) = P_t[\tau \le T] = \Phi(d_1) - \exp\left\{\frac{-2\log(H_t) + \hat{\mu}_H}{\hat{\sigma}_H^2}\right\}\Phi(d_2) \ (5)$$

where $\Phi(\cdot)$ is a cumulative distribution function of standard normal distribution,

$$d_1 = -\frac{\log(Health_t) + \hat{\mu}_H(T-t)}{\hat{\sigma}_H\sqrt{T-t}}, d_2 = -\frac{\log(Health_t) - \hat{\mu}_H(T-t)}{\hat{\sigma}_H\sqrt{T-t}} \quad (6)$$

for $\hat{\mu}_H = (\mu_m - \mu_b) - \frac{\sigma_m^2 - \sigma_b^2}{2}$, and $\hat{\sigma}_H = \sqrt{\sigma_m^2 + \sigma_b^2 - 2\sigma_m\sigma_b\rho_{mb}}$.

The platform-level probability is estimated by the weighted average of user level estimates of the probability with the weights proportional to the amount of value of borrowing.

Modeling expected loss. In case of liquidation, the platform will call for a liquidator in exchange for a liquidation bonus, say, $\alpha \times 100$ percent of the principal, and receive the repayment from the liquidator. If this process completes instantly, liquidation loss will not happen. However, liquidation may take some time. Suppose liquidation completes at time $T' = \tau + \tau'$, namely τ'-periods after it was triggered). The collateral value may fall below the principal (and the liquidation bonus) due to price fluctuations. The cash flow of the platform is $Payoff_{T'} = \min\{-\alpha V_{b,T'} + V_{c,T''}, V_{b,T'}\} = V_{b,T'}\left(1 - \max\left\{(1+\alpha) - \frac{V_{c,T'}}{V_{b,T'}}, 0\right\}\right)$. The loss rate is then defined as $Loss_{T'} = \frac{V_{b,T'} - Payoff_{T'}}{V_{b,T'}} = \max\left\{(1+\alpha) - \frac{V_{c,T'}}{V_{b,T'}}\right\}$.

Consequently, the expected loss ($EL(t,T)$) can be obtained from the following equation:

$$EL(t,T) = P_t[\tau \le T] \cdot E_t[Loss_{T'}|\tau \le T] = P_t[\tau \le T]\left[(1+\alpha)\Phi(d_3) - e^{\hat{\mu}_{cb}\tau' + \frac{\hat{\sigma}_{cb}^2\tau'}{2}}\Phi(d_4)\right] \quad (7)$$

where

$$d_3 = \frac{\log(1+\alpha) - \log(1/\Theta) - \hat{\mu}_{cb}\tau'}{\hat{\sigma}_{cb}\sqrt{\tau'}}, \qquad d_4 = d_3 - \hat{\sigma}_{cb}\sqrt{\tau'}, \ (8)$$

for $\hat{\mu}_{cb} = (\mu_c - \mu_b) - \frac{\sigma_c^2 - \sigma_b^2}{2}$, $\hat{\sigma}_{cb} = \sqrt{\sigma_c^2 + \sigma_b^2 - 2\sigma_c\sigma_b\rho_{cb}}$, and $\Theta = \frac{V_{m,t}}{V_{c,t}} = \frac{\sum_{k=1}^{K}\theta_k V_{c,k,t}}{\sum_{k=1}^{K} V_{c,k,t}} = \sum_{k=1}^{K}\theta_k w_{ck}$ is the weighted average collateral factor at time t. Since $V_{m,\tau} = V_{b,\tau}$ holds at time τ, the sufficiency of the collateral at time τ is given by $\frac{V_{c,\tau}}{V_{b,\tau}} = \frac{V_{c,\tau}}{V_{m,\tau}} = \frac{1}{\Theta}$.

Parameter estimation. The drift $\boldsymbol{\mu}$, volatility $\boldsymbol{\Sigma}$, and correlation $\boldsymbol{\rho}$ are estimated using daily crypto asset price returns. Euler approximation is applied to the price process:

$$\boldsymbol{R}_{t_n} = \mathrm{diag}(\boldsymbol{p}_{t_{n-1}})^{-1}(\boldsymbol{p}_{t_n} - \boldsymbol{p}_{t_{n-1}}) = \boldsymbol{\mu_d}\Delta t + \boldsymbol{\Sigma_d}\sqrt{\Delta t}\boldsymbol{e}_{t_n} \quad (9)$$

where $\boldsymbol{R}_{t_n} = (R_{k,t_n})_{k=1,2,\cdots K;n=1,2,\cdots,N}$ is the $(K \times 1)$-dimensional time-series of crypto asset price returns at time t_n, $n = 1,2,\cdots N$; $\boldsymbol{e}_{t_n} = (e_{kt_n})_{k\in\{b,c\},n=1,2,\cdots,N}$ is the error term with i.i.d. $(K \times 1)$-dimensional Gaussian distribution $(\sim N(\boldsymbol{0}_K, \boldsymbol{I}_K))$; $\boldsymbol{\mu_d}$ is the $(K \times 1)$-dimensional drift parameter vector denoted by annual rate; and $\boldsymbol{\Sigma_d}\boldsymbol{\Sigma_d}'$ is the $(K \times K)$-dimensional covariance matrix of the annual return. $\Delta t = t_n - t_{n-1}$ is the time interval.

The estimator of drift parameter $\widehat{\boldsymbol{\mu}}_d$ and covariance $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{d}}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{d}}'$ are obtained as

$$\widehat{\boldsymbol{\mu}}_d = \frac{1}{N\Delta t}\sum_{n=1}^{N}\boldsymbol{R}_{t_n} \text{ and } \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{d}}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{d}}' = \frac{1}{(N-1)\Delta t}\sum_{n=1}^{N}(\boldsymbol{R}_{t_n} - \widehat{\boldsymbol{\mu}}_{\boldsymbol{d}})(\boldsymbol{R}_{t_n} - \widehat{\boldsymbol{\mu}}_{\boldsymbol{d}})'. \quad (10)$$

$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{d}}$ is given by Cholesky decomposition. Volatility $(\sigma_k^2)_{k=1,2,\cdots,K}$ is available from the diagonal element, and correlation $(\rho_{i,j})_{i,j=1,2,\cdots,K}$ is available by normalizing (i,j) element of the covariance matrix $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{d}}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{d}}'$ by the product of volatility σ_i^2 and σ_j^2.

The horizon T is set to 1 (year), the time interval is set as $\Delta t = \frac{1}{365}$ (year) to use daily data for all available crypto price used in the selected DeFi lending platform: Aave v2, Compound v2, and C.R.E.A.M Finance as of 2021-12-22. The data source is CoinGecko which spans from January 1, 2019 to December 22, 2021 at daily frequency.[15] User-level borrowing and collateral outstanding data is available for all crypto assets used in the selected platforms from The Graph. The average duration of liquidation τ' are estimated around 0.35 (year), by assuming an exponential distribution $(\tau' \sim Exp(\lambda))$, based on the data of liquidation records of Compound v2 available from The Graph and the weekly snapshots of all account balance during 2020.[16]

Event Analysis on Cyberattacks

A comprehensive dataset regarding the dates and magnitudes of cyberattacks on DeFi platforms was constructed from various sources, including the following: Chainalysis (2021), CryptoSec.info (2022),

[15] The estimation was carried out by winsorizing the daily returns at 1 percent and 99 percent to filter out outliers.

[16] The weekly account level balance data was available only during 2020 from its data API. Given that the number of users has increased significantly after 2021, shorter liquidation duration is likely for more recent period. The estimate of $\tau' = 0.35$ should involve considerable uncertainty.

ImmuneFi (2021), and rekt (2022).[17] We identified 93 incidents as cyberattacks to DeFi during the period of January 1, 2020 to January 15, 2022.

CAR analysis on crypto asset prices

Using this dataset, the cumulative abnormal returns (CAR) analysis is performed with respect to crypto assets price returns and total value locked in the platforms.

This analysis exploits daily data of crypto asset price available from CoinGecko, including 11,722 assets in total, starting from January 2, 2020 to December 22, 2021.[18] To estimate the abnormal returns, the chapter first constructs a dataset of logarithmic market returns which are computed by taking all assets into a basket with weighted average according to the market capitalization.[19] The returns of the basket are regarded as the market portfolio for each platform.

Using the market returns, the following one-factor model is estimated for crypto assets issued by the attacked platforms:

$$Return_{k,t}^{log} - Risk\ free\ rate_t = \alpha_k + \beta_k\left(Market\ return_t^{log} - Risk\ free\ rate_t\right) + AR_{k,t}^{log} \quad (11)$$

where AR^{log} is defined as the abnormal returns.[20] The first two components of the RHS constitute the systemic variation.[21]

$$SR_{k,t}^{log} = \alpha_k + \beta_k\left(Market\ return_t^{log} - Risk\ free\ rate_t\right) \quad (12)$$

Estimation is carried out within the 60 days window prior to the attack. Among 93 episodes, there are episodes where the same platform was attacked multiple times. In such a case, only the first attack is included in the sample, as the first attack often changes the way the market behaves, as shown in the chapter.[22]

The abnormal returns $\widehat{AR}_{k,t}^{log}$ are estimated as:

$$\widehat{AR}_{k,t}^{log} = Return_{k,t}^{log} - Risk\ free\ rate_t - \hat{\alpha}_k + \hat{\beta}_k\left(Market\ return_t^{log} - Risk\ free\ rate_t\right) \quad (13)$$

where $(\hat{\alpha}_k, \hat{\beta}_k)$ are the estimated coefficients. The cumulative abnormal returns are defined as follows:

[17] Multiple sources (including additional news sources, press releases, any announcements, etc.) were used to validate the dates and numbers for final judgment.

[18] Actual number of samples used in estimation varies depending on the availability of the data during the estimation window for each cyberattack incident. On average, more than 5,800 assets are used to construct the market returns.

[19] Individual returns are winsorized at 1 percent and 99 percent to filter out outliers.

[20] Different from simple returns, logarithmic returns have more consistency and higher precision. Given the magnitude of the shock, the chapter use logarithmic returns instead of simple returns. Note $R_t^{log} = \log\left(\frac{Price_t}{Price_{t-1}}\right)$. This suggests $Price_{t+1} = Price_t \times \exp\left(R_t^{log}\right)$, and $Price_{t+days} = Price_t \times \exp\left(\sum_{u=0}^{days} R_{t+u}^{log}\right) = Price_t \times \exp\left(\sum_{u=0}^{days} SR_{t+u}^{log}\right) \times \exp\left(CAR_{t+days}^{log}\right)$, justifying the last term $\exp\left(CAR_{t+days}^{log}\right)$ to be considered as the variation associate with the cumulative abnormal returns.

[21] Hu and others (2018) reports high correlation of crypto asset returns with the market portfolio constituted by crypto assets, which justifies the one-factor model such as Liu and others (2022).

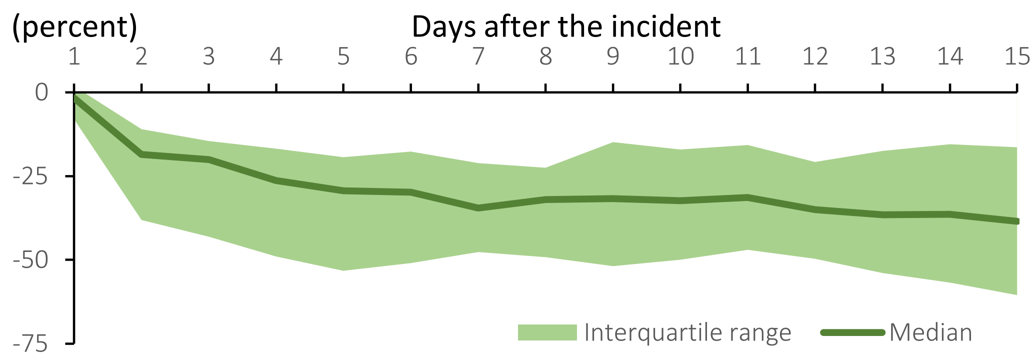
[22] In addition, samples with less than 7 days of datapoints are dropped. Consequently, 28 episodes are selected.

$$\widehat{CAR}_{k,t+days}^{log} = \sum_{u=0}^{days} \widehat{AR}_{k,t+u}^{log} \quad (14)$$

Finally, the logarithmic CAR is converted into CAR: $CAR = \exp(CAR^{log}) - 1$.

Figure 3.3 Cumulative Abnormal Returns

Cumulative Abnormal Returns of Token Prices issued by DeFi Platforms after Cyberattacks



Sources: Chainalysis, CoinGecko, CryptoSec.info, Defi Llama, ImmuneFi, rekt, IMF staff calculations.
Note: In the event analysis on DeFi token prices, cumulative abnormal return (CAR) is computed by a regression of excess returns over the market excess return, a weighted average of all crypto assets minus risk-free rate.

CAR analysis on total value locked

A similar analysis is carried out for total value locked of DeFi platform. In this case, the sample includes all DeFi platforms for which data is available from Defi Llama, including 1,028 DeFi projects in total.[23] Instead of abnormal returns against the market, excess growth of total value locked is measured relative to the total market growth:

$$\widehat{Excess\ Growth}_{k,t}^{log} = Growth_{k,t}^{log} - Market\ Growth_{k,t}^{log} \quad (15)$$

Similar to the CAR analysis on the crypto assets price returns, CAR is estimated as

$$\widehat{CAR}_{k,t+days}^{log} = \sum_{u=0}^{days} \widehat{Excess\ Growth}_{k,t+u}^{log} \quad (16)$$

and converted into simple return format by $CAR = \exp(CAR^{log}) - 1$.

The estimated results are presented in the chapter Figure 3.11 panel 2.

[23] Actual number of samples used in estimation varies dependent on the availability of the data during the estimation window for each cyberattack incident. On average, more than 130 DeFi projects were used to construct the market returns.

Online Annex 3.4. Efficiency Analysis on Financial Institutions and DeFi Platforms

This section provides the estimation method of the price-cost margins of financial intermediaries and their marginal costs as well as the decomposition of marginal costs. This analysis follows the estimation method in Berger and others (2009), who calculate the Lerner Index as a proxy for the market power.

Data Description

Bank/nonbank-level granular data from Fitch Connect comprises 18,011 banks from 137 countries (37 AEs and 100 EMs), 1,136 nonbanks from 46 countries (20 AEs and 26 EMs), spanning from 2007 to 2020 at the annual frequency. The daily data from selected DeFi platforms (Aave and Compound) covers the period between January 2, 2020 to January 28, 2022.[24]

Estimation of Margins and Marginal Costs

The common approach in the literature is the 'intermediation approach' where deposits are an intermediate input in the production of loans (Freixas and Rochet 1997). We take the assumption of a single output, in line with a seminal study by Berger and others (2009) and Weill (2013) using total assets as the output measure, while revenue associated with the output is the interest and non-interest income.

We then estimate the price-cost margins and marginal costs of banks, nonbanks, and DeFi platforms.[25] The price-cost margin $PCM_{i,t}$ is the difference between price and marginal cost, which is calculated as:

$$PCM_{i,t} = P_{i,t} - MC_{i,t} \ (17)$$

where $P_{i,t}$ is the price of total assets proxied by the ratio of total revenues (the sum of interest and non-interest income) to total assets for firm[26] i at time t (PCM would be zero in case of perfect competition).

As marginal costs are not directly observable for an individual firm, we use a trans-log total cost function[27] to estimate the parameters of the cost function and use them to derive the marginal costs.

$$\ln C_{i,t} = \beta_0 + \beta_1 \ln Q_{i,t} + \frac{\beta_2}{2}\left(\ln Q_{i,t}\right)^2 + \sum_{k=1}^{3} \gamma_k \ln W_{k,i,t} + \sum_{k=1}^{3} \delta_k \ln Q_{i,t} \ln W_{k,i,t}$$
$$+ \frac{1}{2}\sum_{k=1}^{3}\sum_{j=1}^{3} \phi_{kj} \ln W_{k,i,t} \ln W_{j,i,t} + \psi_1 T + \frac{\psi_2}{2} T^2 + \sum_{k=1}^{3} \theta_k T \ln W_{k,i,t} + \Gamma X_{i,t} + \varepsilon_{i,t} \ (18)$$

where $C_{i,t}$ is the total cost (or expenses) for firm i at time t; $Q_{i,t}$ is total assets, a proxy for bank output; $W_{k,i,t}$ $(k = 1,2,3)$ are input prices reflecting labor costs, funding costs, and operational costs, respectively.

[24] Data is downloadable from their data API.

[25] Samples are winsorized at 1 percent and 99 percent to filter out outliers.

[26] The term "firm" indicates banks, non-banks, and DeFi platforms.

[27] The translog cost function is frequently used in the banking literature, providing an estimation for the marginal cost of production for banks (Weill 2013).

For input prices, we proxy labor costs using the ratio of personnel expenses to total assets, funding costs using the ratio of interest expenses to total liabilities, and operational costs using the ratio of other operational expenses to total assets.[28] T is a time trend to capture technological changes. $X_{i,t}$ are the firm-characteristic control variables, such as equity to asset ratio and share of non-interest operational income. The cost function is estimated using a panel regression with fixed cross-section effects and clustered errors for each country.

The marginal cost can be derived by differentiating the cost function as follows:

$$MC_{i,t} = \frac{\partial C_{i,t}}{\partial Q_{i,t}} = \frac{\partial \ln C_{i,t}}{\partial \ln Q_{i,t}} \cdot \frac{C_{i,t}}{Q_{i,t}} = \left\{\beta_1 + \beta_2 \ln Q_{i,t} + \sum_{k=1}^{3} \delta_k \ln W_{k,i,t}\right\} \cdot \frac{C_{i,t}}{Q_{i,t}} \; (19)$$

The margin and marginal costs are averaged over time for each firm i, weighted by revenues, and we chose the average value for each country. As Berger, Klapper, and Turk-Ariss (2009) pointed out, these estimates do not capture risk premia in the prices of firms' products or services.

Cost Decomposition

The share of each input cost to total costs can be calculated as follows. We assume the total cost is composed of three input factors, $F_k (k = 1,2,3)$, as follows:

$$\mathrm{C} = W_1 F_1 + W_2 F_2 + W_3 F_3 \; (20)$$

then the share of input factor S_k can be derived with a simple transformation:

$$S_k = W_k \cdot \frac{F_k}{\mathrm{C}} = W_k \cdot \frac{\partial C / \partial W_k}{C} = W_k \cdot \frac{\partial \ln C}{\partial \ln W_k} \cdot \frac{C}{W_k} \cdot \frac{1}{\mathrm{C}} = \frac{\partial \ln C}{\partial \ln W_k}$$

$$= \gamma_{\mathrm{k}} + \delta_{\mathrm{k}} \ln Q + \sum_{k=1}^{3} \sum_{j=1}^{3} \phi_{kj} \ln W_j \; (21)$$

where $\gamma_{\mathrm{k}}, \delta_k, \phi_{kj}$ are the coefficients from the estimation of the total cost function shown above.

[28] Other operational expenses are defined as non-staff related operating expenses incurred through the normal course of business, which includes, depreciation, amortization, administrative expenses, occupancy costs, software costs, operating lease rentals, audit and professional fees, and other operating expenses of an administrative nature.

References

Berger, Allen N., Leora F. Klapper, and Rima Turk-Ariss. 2009. “Bank Competition and Financial Stability: An Empirical Evaluation.” Journal of Financial Services Research 35: 99–118.

Black, Fisher, and John C. Cox, 1976. “Valuing Corporate Securities: Some Effects of Bond Indenture Provisions.” Journal of Finance, Vol. 31, no. 2, (May 1976): 351-367.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2018. "Fintech, Regulatory Arbitrage, and the Rise of Shadow Banks." Journal of Financial Economics 130 (3): 453-483.

Chainalysis. 2021. “The Chainalysis 2021 Crypto Crime Report” (https://go.chainalysis.com/2021-Crypto-Crime-Report.html).

Clerc, Laurent, Arthur Moraglia and Sylvain Perron. 2020. “Neobanks: Stakeholders Seeking Profitability”. ACPR Analyses et syntheses, No 113 – 2020. Study carried out by the Research and Risk Analysis Directorate and the FinTech Unit of the Autorité de Contrôle Prudentiel et de Résolution (ACPR) at the Banque de France. 2020_etude_neobanques_en.pdf (banque-france.fr).

CryptoSec.info. 2022. “DeFi Exploit Timeline” (https://cryptosec.info/defi-hacks/) accessed on January 16, 2022.

Exton Research. 2021. “Report Neobanks 2021: Shifting from Growth to Profitability”. Report-Neobanks-2021.pdf (extonconsulting.com).

Freixas, Xavier and Jean-Charles Rochet. 1997. “The Microeconomics of Banking”, MIT Press, Cambridge MA.

Goldman Sachs. 2022. “Nu Holdings: Leading Fintech in LatAm Well Positioned to Deliver Long-Term Growth”. January 3, 2022.

Hu, Albert, Christine A. Parlour, and Uday Rajan. 2018. “Cryptocurrencies: Stylized Facts on a New Investible Instrument.” Working paper, University of California, Berkeley.

Immunefi. 2021. “Immunefi Crypto Losses Report” (https://immunefi.com/immunefi-crypto-losses-report-2021/)

Jagtiani, Julapa, Lauren Lambie-Hanson, and Timothy Lambie-Hanson. 2021. "Fintech Lending and Mortgage Credit Access." The Journal of FinTech 1 (1): 2050004

Liu, Yukun, Aleh Tsyvinski, Xi Wu. 2022. “Common Risk Factors in Cryptocurrency,” Journal of Finance, Accepted Author Manuscript. https://doi.org/10.1111/jofi.13119.

Merton, Robert C. 1974. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." Journal of Finance Vol. 29, no. 2, (May 1974): 449-470.

Microsoft, Linklaters and Accenture. 2019. "Open Banking: A Shared Opportunity." RE489V8 (microsoft.com)

Morgan Stanley. 2021. “Kakao Bank Corp: Mechanics, Motor & Mettle to Lead Digital Banking”. October 11, 2021.

NeoBanks.app (NeoBanks.app: List of 247 neobanks and digital banks in the world - January 2022) accessed on January 23, 2022.

Rekt. 2022. “leaderboard” (https://rekt.news/leaderboard/) accessed on January 16, 2022.

S&P Capital IQ. 2021. “Strong Interest Income Tips the Scale for Digital Banks in Europe, Asia”. November 10, 2021.

Weill, Laurent. 2013. “Bank Competition in the EU: How Has it Evolved?” *Journal of International Financial Markets, Institutions and Money*, 26:100–112

WhiteSight. 2021. “Back to The Future: The Evolution of Branchless Banking”. October 6, 2021. Back To The Future: The Evolution of Branchless Banking (whitesight.net)