

The Value of Intermediaries for GSE Loans*

Joshua Bosshardt
FHFA

Ali Kakhbod
UC Berkeley

Amir Kermani
*UC Berkeley
& NBER*

March 18, 2023

Abstract

We analyze the costs and benefits of financial intermediaries on access to credit using confidential regulatory data on mortgages securitized by the government-sponsored enterprises (GSEs). We find evidence of lenders pricing for observable and unobservable default risk independently from the GSEs. We explain these findings using a model of competitive mortgage lending with screening in which lenders acquire information beyond the GSEs' underwriting criteria and retain a positive loss given default. The model shows that the discretionary behavior of lenders, relative to a counterfactual in which lenders passively implement the GSEs' underwriting requirements and price competitively, benefits some borrowers with high observable risk at the expense of the majority of borrowers. Finally, the model suggests that the observed differences between banks and nonbanks are more consistent with differences in their expected loss given default rather than screening quality.

Keywords: mortgage lenders · underwriting risk · overlays · nonbanks

JEL Classification: G21 · G23

*We thank Bob Avery, Justin Contat, Len Costa, Will Doerner, Robert Dunskey, Constantine Yanelis, and seminar participants at FHFA for comments. Joshua Bosshardt is with the Federal Housing Finance Agency (FHFA), Washington, DC, USA, email: joshua.bosshardt@fhfa.gov. Ali Kakhbod is with the University of California, Berkeley, Haas School of Business, Berkeley, CA, USA, email: akakhbod@berkeley.edu. Amir Kermani is with the University of California, Berkeley, Haas School of Business, Berkeley, CA, USA, and NBER, email: kermani@berkeley.edu. This paper is also an FHFA Working Paper and available here: <https://www.fhfa.gov/papers/wp2301.aspx>. The views expressed in this paper are solely those of the authors and not necessarily those of the Federal Housing Finance Agency or the U.S. Government. Any errors or omissions are the sole responsibility of the authors.

1 Introduction

Mortgage debt is by far the largest component of household debt in the U.S., accounting for more than 70% of the \$16.5 trillion in household liabilities ([Federal Reserve Bank of New York \(2022\)](#)). Access to mortgage credit partly depends on the prevailing credit profiles in the mortgage market segment supported by the government-sponsored enterprises (GSEs) Fannie Mae and Freddie Mac, which comprises the majority of originations since the financial crisis.^{1,2} In this market segment, the access to credit depends on two factors: the GSEs' underwriting criteria as implemented through their respective automated underwriting systems (Desktop Underwriter for Fannie Mae and Loan Prospector for Freddie Mac) and potential additional restrictions or "overlays" imposed by private mortgage lenders, which serve as intermediaries by originating the loans that the GSEs eventually securitize.³ Given that the GSEs insure the default risk of these mortgages and advancements in the automated underwriting systems, an important question is: what is the added value of intermediaries' discretionary overlays? We specifically focus on a trade-off in which intermediaries reduce the cost of lending by screening out borrowers that are relatively likely to default relative to their easily observed risk characteristics (such as credit score, loan-to-value (LTV) ratio, and debt-to-income (DTI) ratio) but can also charge markups. We further consider which types of borrowers would benefit (or not) from changes in the role or nature of intermediaries, such as a market in which intermediaries passively implement the GSEs' underwriting standards or a market dominated by either bank to nonbank lenders.

We approach these questions by first documenting evidence of an intensive margin of

¹Note that, in this paper, the "GSEs" refers to Fannie Mae and Freddie Mac and not any other government-sponsored enterprises.

²Access to mortgage credit could also be influenced by the credit profiles supported by government-administered housing programs associated with the Federal Housing Administration (FHA), the U.S. Department of Veterans Affairs (VA), and the U.S. Department of Agriculture (USDA). Government loans typically have more flexible credit standards than GSE loans, but they are also associated with various disadvantages, such as required insurance fees, restrictions on the property or loan terms, and less competitive standing compared to other homebuyers. As a result, the credit standards for GSE loans still affect the overall access to housing credit. Access to mortgage credit could also be influenced by the credit standards for loans that are not associated with any government agencies or government-sponsored enterprises, but the share of such loans has been limited since the financial crisis. Additionally, the credit standards for GSE loans have historically also been important even for loans held in portfolio or sold to other secondary market participants ([Johnson \(2022\)](#)).

³Note that we focus on the role of intermediaries at loan origination and not subsequent servicing or other investor activity.

overlays using confidential regulatory data on all loans acquired by the GSEs during part of the post-crisis period. In particular, we find that interest rates net of the g-fees that the GSEs charge for insuring a loan increase with measures of ex-ante observable risk, such as the borrower’s credit score, the LTV ratio, and the DTI ratio. This result suggests that lenders independently add a risk spread to the interest rate. Second, we find that interest rates net of g-fees predict default even after controlling for observable risk, which suggests that this risk spread reflects additional screening by lenders.⁴ Third, we present evidence that interest rates have a relatively modest direct effect on default, which rules out an obvious alternative explanation for our findings. We also find a similar set of results when considering a lender’s total origination revenue instead of the interest rate, where the former incorporates a lender’s income from selling loans on the secondary market as well as closing costs. Finally, on the extensive margin, we show using loan application data that the rate at which lenders deny applications that are accepted by the GSEs also increases with observable risk.

We then develop a model of mortgage lender competition with screening that explains these observations and leverages them to extract insights about the costs and benefits of intermediaries in the GSE segment of the mortgage market. The model has three key ingredients. First, motivated by evidence of lenders pricing for risk, lenders in the model face a positive expected loss given default. For example, this could correspond to penalties imposed by the GSEs, such as repurchases or restrictions on the ability to continue doing business as a counterparty.⁵ Second, motivated by evidence of lenders also pricing for default risk that is not captured by observable risk characteristics, lenders in the model have the ability to implement further screening to determine if they will offer a loan and, if so, how much of a risk spread they will charge in addition to the GSEs’ g-fees. Third, lenders can charge a markup due to limited competition, which is consistent with recent evidence of noncompetitive pricing in the mortgage market (e.g. [Bhutta, Fuster, and Hizmo \(2021\)](#) and [Alexandrov and Koulayev \(2018\)](#)).

⁴Note that this result is robust to capturing observable risk via an extensive set of borrower, loan, and property characteristics, including household demographics, loan purpose and amount, and property value and ZIP code, among others.

⁵A “repurchase”, sometimes also referred to as a “put-back”, refers to when an GSE requires a lender to repurchase a loan based on charges of violating representations and warranties, which can be interpreted as errors in the underwriting process required for delivering a loan to the GSEs. The rate of repurchases increased during the crisis but has since remained low. Nevertheless, [Goodman \(2017\)](#) presents evidence that lenders have increased their investment in careful underwriting and imposed overlays to protect themselves against repurchases.

We leverage the model to compare the status quo in which lenders exercise discretionary overlays, which we refer to as active intermediation, to a counterfactual in which they passively implement the GSEs' underwriting requirements, which we define as approving all applications accepted by the automated underwriting systems and charging a zero-profits interest rate conditional on the borrower's observable risk.⁶ We find that active intermediation is more likely to be associated with higher interest rates for the majority of borrowers due to markups, although it can benefit borrowers with high observable risk due to lenders' additional screening.

While the primary question of this paper is about the costs and benefits of intermediaries in general, an extension of the model with heterogeneous lenders speaks to observed differences between bank and nonbank lenders. We focus on nonbanks because their market share has increased significantly in the years following the crisis (Buchak et al. (2018)), which has raised questions about financial stability since they are associated with a greater liquidity risk and a stronger tendency to lend to relatively riskier borrowers (Kim et al. (2018)). We find that defaults for nonbanks occur at more than 1 and a half times the rate for banks, which reflects a combination of nonbanks lending to observably riskier borrowers as well as exhibiting higher default rates conditional on observable risk. Consistent with their riskier credit profiles, nonbanks are also associated with higher interest rates conditional on observable risk. Finally, we present evidence that the presence of nonbank fintech lenders that use a mostly online application process, which have disproportionately contributed to the increasing market share of nonbanks, has been associated with banks targeting safer borrowers with lower default rates and lower interest rates.

The model suggests the observed differences between banks and nonbanks are more consistent with fundamental differences in the expected loss given default rather than screening quality. In particular, the observation that nonbanks are associated with higher observable risk and higher default rates conditional on observable risk is consistent with them having a lower expected loss given default. By contrast, if nonbanks implemented superior (or worse) screening, they would be expected to focus on observably riskier (safer) borrowers while also having lower (higher) default rates conditional on observable risk. One explanation is that banks more often have an incentive to protect rents

⁶Passive intermediation also produces the same result as a market with perfect competition.

from other business lines, whereas nonbanks, which typically have a monoline business model, may perceive declaring bankruptcy as a less costly limit on losses.⁷

This paper contributes to three major themes in the literature. First, it discusses determinants and implications of access to credit in the U.S. mortgage market. This body of work focuses, for example, on race ([Bhutta, Hizmo, and Ringo \(2021\)](#), [Bartlett et al. \(2022\)](#), and [Giacoletti, Heimer, and Yu \(2022\)](#)), regulations ([Fuster, Plosser, and Vickery \(2021\)](#)), repurchases and servicing costs ([Goodman \(2017\)](#)), fair pricing and credit allocation by region ([Hurst et al. \(2016\)](#) and [Kulkarni \(2016\)](#)), and capacity constraints ([Fuster, Lo, and Willen \(2017\)](#)). We contribute by providing evidence that lenders price for risk on GSE loans in a manner independent of the GSEs' g-fees, consistent with an intensive margin of overlays. Based on our model, we find that intermediaries are more likely to reduce interest rates (compared to a benchmark with passive intermediation) for observably risky borrowers due to their screening incentive, although they result in higher interest rates for the majority of borrowers.

Second, this paper contributes to the literature on the role of nonbanks in mortgage lending ([Buchak et al. \(2018\)](#), [Kim et al. \(2018\)](#), and [Gete and Reher \(2020\)](#)), including fintechs in particular ([Fuster et al. \(2019\)](#) and [Jagtiani, Lambie-Hanson, and Lambie-Hanson \(2021\)](#)). We show that nonbanks are associated with higher default rates and interest rates even conditional on observable risk. Based on our model, we conclude that the observed differences between banks and nonbanks are more consistent with nonbanks having a lower expected loss given default rather than differences in screening quality.

Third, this paper contributes to the literature on noncompetitive pricing in the mortgage market ([Alexandrov and Koulayev \(2018\)](#), [Bhutta, Fuster, and Hizmo \(2021\)](#), [Buchak and Jørring \(2021\)](#), [Malliaris, Retzl, and Singh \(2021\)](#), and [Valentin \(2021\)](#)). It also adds to a more general literature on competition among financial intermediaries, including banks versus banks ([Egan, Hortaçsu, and Matvos \(2017\)](#)), banks versus nonbanks ([Benet-](#)

⁷For example, nonbanks exhibited lower rates of repurchases of risky loans they originated during the housing boom. In particular, within the set of 2007 originations that were delivered to Freddie Mac, 1.08% of the loans delivered by banks have been repurchased compared to only 0.65% of the loans delivered by nonbanks. This may have been due to them failing or being sold to banks during the crisis (e.g. [Buchak et al. \(2018\)](#)) and thus not being liable to further penalties from the GSEs.

ton, Buchak, and Robles-Garcia (2022)), ntechs versus other intermediaries (Di Maggio and Yao (2021)), algorithmic versus human underwriting processes (Jansen, Nguyen, and Shams (2021)), and the relationship between competition and underwriting quality (Yannelis and Zhang (2021)). We provide a model that decomposes mortgage interest rates into components corresponding to the cost of funding, risk spreads, and markups. Based on our model, we conclude that markups account for a relatively greater portion of interest rates for observably safe borrowers compared to observably risky borrowers.

2 Empirical observations

This section first shows that interest rates net of g-fees and origination revenues in the GSE segment of the mortgage market positively correlate with both ex-ante observable risk and ex-post default, suggesting that lenders independently price for risk. It also shows that, compared to banks, nonbanks are associated with greater observable risk, greater default rates conditional on observable risk, and greater interest rates and origination revenue conditional on observable risk, consistent with nonbanks internalizing a lower loss given default.

2.1 Key definitions

We consider relationships involving interest rates, g-fees, origination revenue, ex-post default risk, ex-ante observable risk, and lender types, which we define in the context of our analysis as follows:

Interest rate is the annualized mortgage interest rate at origination.

Guarantee fee or g-fee, refers to the cost that the GSEs charge for acquiring and guaranteeing a mortgage loan. Note that the g-fee for a loan typically contains an ongoing component, which is charged as an annual rate, and an upfront component, which is charged as a percentage of the loan amount. The ongoing component typically depends on the loan's general product type, whereas the upfront component typically depends on the loan's specific risk characteristics. The upfront g-fee is the sum of components de-

scribed in each GSE's respective matrix.⁸ A base component for all loans with terms greater than 15 years depends on the loan-to-value (LTV) ratio and credit score.⁹ Other components can depend on features of the loan (such as the loan purpose) or the property (such as the occupancy type), among other factors. Our measure of the total g-fee expressed as an annualized rate, combines the ongoing and upfront components by converting the upfront component to an annualized rate using the loan's present value multiplier, which is computed by the loan's guaranteeing GSE based on the expected duration of the loan.

Origination revenue refers to a lender's income from originating a loan, expressed as a percentage of the loan amount and not annualized. Origination revenue is the sum of two components: upfront closing costs and secondary marketing income (Zhang (2022)). Closing costs is measured by origination charges which we obtain by merging with the recently expanded HMDA data. Secondary marketing income is the present value of the deviation of a loan's interest rate net of g-fees relative to par, similar to the price of intermediation in Fuster, Lo, and Willen (2017). We compute it by subtracting the current coupon yield on MBS (Bloomberg code "MTGEFNCL") as of the origination date from the interest rate net of the total g-fee and multiplying by the respective present value multiplier (PVM):¹⁰

$$\text{secondary marketing income} = (\text{interest rate} - \text{total g-fee} - \text{MBS yield}) \times \text{PVM} \quad (1)$$

Default in the context of our analysis refers to 90-day delinquency within 2 years of origination.

⁸See <https://singlefamily.fanniemae.com/media/9391/display> for the most recent matrix for Fannie Mae, which refers to the upfront g-fee as loan-level price adjustments. See https://guide.freddiemac.com/euf/assets/pdfs/Exhibit_19.pdf for the most matrix for Freddie Mac, which refers to the upfront g-fee as credit fees. Note that the current matrix may not coincide with the matrix during the sample period. See <https://www.fhfa.gov/AboutUs/Reports/ReportDocuments/GFee-Report-2021.pdf> for general information from the Federal Housing Finance Agency g-fee report.

⁹While the upfront g-fee is generally increasing in default risk, it may not price for risk perfectly. In particular, the matrix is consistent with cross-subsidization of relatively risky borrowers (with high LTV and low credit scores) by relatively less risky borrowers (with low LTV and high credit score). This is not a problem for our analysis, which focuses on the component of interest rates determined by lenders rather than the GSEs.

¹⁰Note that if we split up the g-fee into the ongoing and upfront components then this is also equivalent to: $\text{secondary marketing income} = (\text{interest rate} - \text{ongoing g-fee} - \text{MBS yield}) \times \text{PVM} - \text{upfront g-fee}$.

Observable risks the estimated probability of default based on (that is, is observable with respect to) determinants of the upfront g-fee (credit score and LTV) as well as DTI. Specifically, it is the predicted value of a regression of default (multiplied by 100) on the interaction of credit score bins corresponding to thresholds in the upfront g-fee (less than 620, 620-639, 640-659, 660-679, 680-699, 700-719, 720-739, and 740 or greater), loan-to-value (LTV) bins corresponding to thresholds the upfront g-fee (60% or less, 60.01-70%, 70.01-75%, 75.01-80%, 80.01-85%, 85.01-90%, 90.01-95%, and greater than 95%), and debt-to-income (DTI) bins corresponding to quintiles. Note that the credit score for a loan refers to the "representative credit score" that is used to determine the g-fee. This is defined as the minimum of each borrower's representative score, which is either the lower score if there are two scores or the middle score if there are three.

Banks refers to depositories.

Nonbanks refers to lenders that are not banks.

Fintechs refers to lenders with a mostly online application process. We use the designation of ntechs in [Fuster et al. \(2019\)](#). Note that all ntechs are nonbanks, so we can further distinguish ntechs from nonbank-non ntechs in order to have non-intersecting categories.

2.2 Data

We use data from the Mortgage Loan Integrated System (MLIS), which is a confidential regulatory dataset at the Federal Housing Finance Agency (FHFA) consisting of all loans acquired by the GSEs. The tables and figures in this paper do not contain any confidential or personally identifiable information.

For our baseline sample, we focus on originations during 2016-2017. We start in 2016, which is when we start to have precise data on g-fees, and we end at 2017 because we consider 2-year default rates and do not want to extend into the COVID-19 pandemic. For the results regarding origination revenue, we use the sample of originations during 2018, which we merge with the expanded HMDA data to obtain information on orig-

ination charges.¹¹ Note that observable risk in 2018 is computed based on the model estimated from the baseline 2016-2017 sample rather than the 2018 sample to avoid systematic changes in 2-year default rates associated with the COVID-19 pandemic.

We focus on a subsample of loans where the upfront portion of the g-fee approximately only depends on the LTV ratio and credit score. In particular, we restrict to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses. We also exclude high balance loans exceeding the national baseline conforming loan limit and loans with subordinate financing. Finally, within the resulting set, we restrict to loans where the total upfront g-fee is within 25 basis points of the component determined by LTV and credit score.

Table A.1 in Appendix Section A.1 presents summary statistics for the baseline 2016-2017 sample, and Table A.2 presents summary statistics for the 2018 sample. Note that continuous variables are winsorized at 1%.

2.3 Interest rates net of g-fees and denials increase with default risk

Observations 1-3 show that interest rates net of g-fees are positively associated with observable risk and unobservable risk (i.e. default risk conditional on observable risk) to a degree that exceeds their direct effect on default. Observation 4 shows using application data that denial rates on applications also increase with observable risk.

2.3.1 Observation 1: interest rates net of g-fees increase with observable risk

Figure 1 shows observable risk is positively associated with interest rates, even after subtracting out the g-fee. Similarly, Table 1 column (1) shows that interest rates are positively associated with observable risk while also controlling for ZIP code by year-quarter fixed effects based on the origination date. Decomposing the components of observable risk, column (2) shows that interest rates are negatively associated with credit score and positively associated with the loan-to-value (LTV) ratio and the debt-to-income (DTI) ratio. Column (3) and column (4) show that these results are only partially mitigated by subtracting out the total g-fee. Based on the estimate in column (3), a 1 percentage point increase in the ex-ante probability of default is associated with a 9.4 basis point increase

¹¹We implement an exact merge based on the following characteristics: loan amount rounded to the nearest \$5,000, interest rate, year, loan purpose, term, and census tract. We omit observations in either dataset which are identical based on these characteristics.

in the interest rate net of g-fees. Column (5) shows that the association is stronger for loans with LTV less than or equal to 80%, in which case the GSEs require private mortgage insurance, while column (6) shows that the association is weaker for loans with LTV greater than 80%, in which case the GSEs do not require private mortgage insurance.

These results suggest that lenders independently price for default risk, possibly due to repurchases or other penalties imposed by the GSEs. Section 3 incorporates this result into a model of mortgage lender competition with screening by supposing that lenders bear a positive expected loss given default.

As additional robustness, Figure A.1 in Appendix Section A.2 shows using the 2018 sample that origination revenue is also associated with observable risk. It also shows that this association is primarily driven by secondary market income, whereas closing costs are comparatively constant with respect to observable risk. Table A.3 column (1) and column (2) show that origination revenue is associated with observable risk and its constituent factors while controlling for ZIP code by year-quarter fixed effects. Columns (3)-(6) show that the closing costs portion and the secondary market income are also both increasing in observable risk, although the latter is much stronger and appears to drive the overall association between origination revenue and observable risk.

Additionally, Table C.1 in Appendix Section C.1 shows using an analogous dataset derived from Optimal Blue that the results are qualitatively similar when controlling for lock rate date fixed effects instead of ZIP code by year-quarter fixed effects, which better helps to control for potential changes in the composition of borrowers associated with short-term fluctuations in interest rates. ^{12,13}

¹²Note that Optimal Blue does not have a multiplier to convert the upfront g-fee to an annualized rate, so we instead include fixed effects for the expected upfront g-fee as a function of credit score and LTV based on the first table of the GSEs' g-fee matrix.

¹³Generalizing these results from the GSE segment of the market, Table C.2 shows that risk characteristics also appear to be priced in loans insured by government agencies, including the Federal Housing Administration (FHA), Department of Veterans Affairs (VA), and Department of Agriculture (USDA). Note that we do not include observable risk as a regressor since we estimate observable risk based on GSE loans, which are generally less risky.

Figure 1: Interest rates and observable risk

This figure presents a binned scatterplot of the interest rate and the interest rate net of the total g-fee on observable risk while controlling for year-month fixed effects. Observable risk is the estimated probability of default based on credit score, the loan-to-value ratio, and the debt-to-income ratio as described in Section 2.1. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

Table 1: Interest rates and observable risk

	(1)	(2)	(3)	(4)	(5)	(6)
	IR	IR	IR - g-fee	IR - g-fee	LTV< = 80	LTV> 80
Observable risk	0.206*** (647.97)		0.094*** (340.86)		0.119*** (214.17)	0.059*** (189.56)
Credit score		-0.299*** (-562.96)		-0.125*** (-261.29)		
LTV		0.741*** (440.84)		0.497*** (330.76)		
DTI		0.274*** (106.07)		0.277*** (119.46)		
Observations	2,109,041	2,109,041	2,109,041	2,109,041	851,576	1,219,780
R ²	0.492	0.547	0.496	0.535	0.486	0.539
ZIP Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: Column (1) regresses the interest rate on observable risk while controlling for ZIP code by year-quarter fixed effects. Observable risk is the estimated probability of default based on credit score, the loan-to-value ratio, and the debt-to-income ratio as described in Section 2.1. Column (2) regresses the interest rate on credit score, the loan-to-value (LTV) ratio, and the debt-to-income (DTI) ratio (each divided by 100). Column (3) and column (4) are similar to column (1) and column (2) except that the dependent variable is the interest rate net of the total g-fee. Column (5) and column (6) are similar to column (3) except restricting to loans with LTV less than or equal to 80% or LTV greater than 80%, respectively. T-statistics computed using robust standard errors are reported in parentheses. * indicates statistical significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

2.3.2 Observation 2: interest rates net of g-fees predict default conditional on observable risk

Figure 2 shows that interest rates net of g-fees are positively associated with default rates, even after controlling for observable risk. Similarly, Table 2 column (1) shows that interest rates are predictive of default while also controlling for ZIP code by year-quarter fixed effects, while column (2) shows that this relationship continues to hold even after controlling for observable risk. Column (3) shows that it continues to hold after controlling for a host of additional observable characteristics, including the interaction between 10-point credit score bins (starting at 620, with an additional indicator for all credit scores below 620), 5% loan-to-value bins (starting at 60%, with an additional indicator for all loan-to-value ratios below 60%), and debt-to-income decile indicators (note that this absorbs observable risk); income decile indicators; family type indicators (i.e. single female, single male, or more than 1 borrower); indicators for black and hispanic borrowers; term indicators; appraisal value decile indicators; an indicator for a loan having an interest-only period; an indicator for a refinance loan; loan amount decile indicators; an indicator for self-employed borrowers; an indicator for first-time homebuyers; an indicator for full income documentation; and an indicator for full asset documentation.

Based on the estimate in column (2), a 1 percentage point increase in the interest rate net of g-fees is associated with a 35 basis point increase in the default rate conditional on observable risk, which is substantial compared to the overall default rate of 57 basis points. Column (4) shows that the association between interest rates and default is weaker for relatively safe borrowers with observable risk below the median, while column (5) shows that the association is stronger for riskier borrowers with observable risk above the median. Column (6) shows that the difference between relatively safe and risky borrowers is statistically significant. Finally, column (7) shows that the result is similar when using the interest rate without subtracting out the g-fee.

These results suggest that lenders implement additional screening compared to the determinants of the upfront g-fee, particularly for observably riskier borrowers. This is consistent with existing evidence based on mortgage lender capacity constraints that mortgages for observably riskier borrowers are relatively more time-consuming to underwrite (Sharpe and Sherlund (2016), Fuster et al. (2021)). Section 3 incorporates this

result into a model of mortgage lender competition by supposing that lenders can invest in improving their underwriting practices, which allows them to observe a partially informative signal of the borrower's default risk conditional on observable risk.

As additional robustness, Figure A.2 and Table A.4 in Appendix Section A.3 show that default is also positively associated with origination revenue in 2018, notwithstanding the effects of COVID-19 on loan performance.

Figure 2: Interest rates and default

This figure presents a binned scatterplot of default (multiplied by 100) on the interest rate net of the total g-fee while controlling for year-month fixed effects and observable risk. Observable risk is the estimated probability of default based on credit score, the loan-to-value ratio, and the debt-to-income ratio as described in Section 2.1. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

Table 2: Interest rates and default

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Baseline	Obs. risk	Controls	Safe	Risky	Interact	IR
IR - g-fee	0.927*** (48.99)	0.346*** (18.78)	0.426*** (16.34)	0.129*** (10.13)	0.674*** (17.32)	0.129*** (10.12)	
Obs. risk		0.928*** (63.71)		0.716*** (11.72)	0.943*** (52.69)	0.716*** (11.71)	0.904*** (59.72)
IR - g-fee Risky						0.545*** (13.31)	
Obs. risk Risky						0.227*** (3.57)	
IR							0.275*** (17.46)
Observations	2,109,041	2,109,041	2,109,029	1,030,232	1,040,160	2,070,392	2,109,041
R ²	0.092	0.100	0.103	0.137	0.137	0.140	0.100
ZIP Year-quarter FE	Yes	Yes	Yes	Yes	Yes	No	Yes
ZIP Year-quarter Risky FE	No	No	No	No	No	Yes	No
Controls	No	No	Yes	No	No	No	No

Note: Column (1) regresses an indicator for default (multiplied by 100) on the interest rate net of the total g-fee while controlling for ZIP code by year-quarter fixed effects. Column (2) adds observable risk as a regressor. Observable risk is the estimated probability of default based on credit score, the loan-to-value ratio, and the debt-to-income ratio as described in Section 2.1. Column (3) instead includes the following controls: the interaction between 10-point credit score bins (starting at 620, with an additional indicator for all credit scores below 620), 5% loan-to-value bins (starting at 60%, with an additional indicator for all loan-to-value ratios below 60%), and debt-to-income decile indicators (note that this absorbs observable risk); income decile indicators; family type indicators (i.e. single female, single male, or more than 1 borrower); indicators for black and hispanic borrowers; term indicators; appraisal value decile indicators; an indicator for a loan having an interest-only period; an indicator for a refinance loan; loan amount decile indicators; an indicator for self-employed borrowers; an indicator for first-time homebuyers; an indicator for full income documentation; and an indicator for full asset documentation. Column (4) estimates the specification in column (2) except restricting to relatively safe borrowers with observable risk below the median. Column (5) estimates the specification in column (2) except restricting to relatively risky borrowers with observable risk above the median. Column (6) estimates the specification in column (2) except interacting all of the regressors with a dummy variable Risky indicating borrowers with observable risk above the median. Column (7) estimates the specification in column (2) except using the interest rate (without subtracting out the g-fee) as the dependent variable. T-statistics computed using robust standard errors are reported in parentheses. * indicates statistical significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

2.3.3 Observation 3: interest rates have a relatively small direct effect on default

An alternative explanation for Observation 2 is that higher interest rates might directly increase default risk. To estimate this direct effect, we estimate how default rates vary with changes in interest rates induced by variation in the upfront g-fee for borrowers with similar observable risk. The exogeneity assumption that the variation in the upfront g-fee for a given level of observable risk only affects default rates through the interest rate. For example, it is not correlated with any unobservable risk that might otherwise lead lenders to choose a higher interest rate.

Column (1) of Table 3 shows in a first-stage regression that the upfront g-fee strongly predicts interest rates while controlling for observable risk. Column (2) shows that a 1 percentage point increase in the interest rate induced by the the upfront g-fee is only associated with a 3 basis point and statistically insignificant increase in the default rate, which is small compared to the overall association of 28 basis point reported in Column (7) of Table 2. Column (3) and Column (4) show that this result is mostly similar on the subsets of safe and risky loans, as determined by whether observable risk is below or above the median. Column (5) shows that the difference between the estimates from the safe and risky subsamples is statistically insignificant.

2.3.4 Observation 4: denials conditional on GSE approval increase with observable risk

For this observation, we use a comprehensive dataset of U.S. mortgage applications in 2018 to examine the extent to which lenders deny applications that are accepted by the GSEs' automated underwriting systems (AUSs). Analogous to the MLIS sample, we restrict to applications for conventional, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family non-manufactured houses. We also exclude applications for high balance loans exceeding the national baseline conforming loan limit and restrict to first lien mortgages. We further restrict to mortgages that are processed by exactly one AUS, which is either Desktop Underwriter (for Fannie Mae) or Loan Prospector (for Freddie Mac), and for which the result of the AUS is "Approve/Eligible" or "Accept."¹⁴ To focus on denials that are relatively likely to reflect screening by lenders rather than problems pertaining to the application process, we exclude denials due to

¹⁴Note that about 92.6% of the sample is processed by only one AUS. The results are similar if we restrict to mortgages that receive a response of "Approve/Eligible" or "Accept" by either Desktop Underwriter (for Fannie Mae) or Loan Prospector (Freddie Mac) for at least one AUS submission.

Table 3: G-fee induced variation in interest rates and default

	(1) First stage	(2) IV	(3) Safe	(4) Risky	(5) Interact
Upfront g-fee	0.216*** (398.60)				
Obs. risk	0.068*** (153.65)	0.954*** (36.16)	0.788*** (9.94)	0.970*** (32.22)	0.788*** (9.93)
IR		0.030 (0.31)	0.045 (1.17)	0.051 (0.35)	0.045 (1.17)
IR Risky					0.006 (0.04)
Obs. risk Risky					0.182** (2.15)
Observations	2,109,041	2,109,041	1,030,232	1,040,160	2,070,392
R ²	0.525	0.010	0.000	0.009	0.008
ZIP Year-quarter FE	Yes	Yes	Yes	Yes	No
ZIP Year-quarter Risky FE	No	No	No	No	Yes
Controls	No	No	No	No	No

Note: Column (1), the first stage, regresses the interest rate on the upfront g-fee and observable risk while controlling for ZIP code by year-quarter fixed effects. Observable risk is the estimated probability of default based on credit score, the loan-to-value ratio, and the debt-to-income ratio as described in Section 2.1. Column (2) regresses an indicator for default (multiplied by 100) on the interest rate and observable risk, while the former is instrumented by the upfront g-fee. Column (3) estimates the specification in column (2) except restricting to relatively safe borrowers with observable risk below the median. Column (4) estimates the specification in column (2) except restricting to relatively risky borrowers with observable risk above the median. Column (5) estimates the specification in column (2) except interacting all of the regressors with a dummy variable Risky indicating borrowers with observable risk above the median and instrument the interest rate and interest rate Risky by the upfront g-fee and the upfront g-fee Risky. T-statistics computed using robust standard errors are reported in parentheses. * indicates significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee deviates from the first table of the g-fee matrix by more than 25 basis points.

incomplete applications and insufficient cash.¹⁵ Note that we compute observable risk as a function of credit score, LTV, and DTI based on the model estimated with the MLIS data.¹⁶

Figure 3 shows that denials increase with observable risk, ranging from about 1.85% for borrowers with 0.07% observable risk to 7.65% for borrowers with 3.4% observable risk.

Figure 3: Application denials and observable risk

This figure presents a binned scatterplot of the denial rate for mortgages accepted by the GSEs' automated underwriting systems on observable risk. Observable risk is the estimated probability of default based on credit score, the loan-to-value ratio, and the debt-to-income ratio as described in Section 2.1, which is estimated using the MLIS data. Source: mortgage application data, 2018, restricting to applications accepted by the GSE automated underwriting systems for conventional, purchase or no cash-out refinance, first lien loan applications for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit.

¹⁵See Table C.3 in Appendix Section C.2 for the fraction of denials attributable to each reason before implementing this restriction.

¹⁶Note that we use the combined LTV since the mortgage application data does not have the original LTV for just the loan application.

2.3.5 Observations 1-4 interpretation and robustness

Consistent with an intensive margin of overlays, Observations 1-3 together suggest the interpretation that lenders charge a risk spread on GSE mortgages that is independent of the g-fee and that predicts default without directly increasing the propensity to default to the same degree. Consistent with an extensive margin of overlays, Observation 4 additionally shows that lenders deny riskier applications, even if they are accepted by the GSEs. Section 3 rationalizes these results with a model in which lenders have a positive loss given default as well as a more precise screening technology compared to what is reflected in the g-fee.

Another possible interpretation for Observations 1-3 is that lenders may be pricing for prepayment risk. However, we account for prepayments in the following ways. First, borrowers may use negative discount points to reduce their closing costs while taking on a higher interest rate, which tends to also increase prepayment rates (Zhang (2022)). Figure A.2 and Table A.4 control for the use of discount points by considering total origination revenue, as discount points only shift the closing cost and secondary marketing income components without affecting the total. Second, borrowers with smaller loans may be less likely to refinance because closing costs are a larger fraction of the principal balance, but column (3) of Table 2 controls for the loan amount. Third, Table A.5 in Appendix Section A.4 shows that interest rates are independently associated with both defaults and prepayments, while Table A.6 shows the same for origination revenue. Finally, as a more general consideration, it is unlikely that prepayment risk confounds the relationship between interest rates and default since they are mutually exclusive events.

2.4 Variation in intermediation patterns by lender type

Strikingly, nonbanks exhibit a default rate that exceeds that of banks by more than 70% (0.75% for nonbank-non ntechs and 0.78% for ntechs, respectively, compared to 0.44% for banks, as shown in Table A.1). To analyze this observation more closely, Observation 5 shows that nonbanks are associated with greater observable risk, while Observation 6 shows that nonbanks are also associated with higher default rates conditional on observable risk. Additionally, Observation 7 shows that nonbanks exhibit higher interest rates conditional on observable risk. Finally, Observation 8 shows that the presence of ntechs in particular is associated with a reduction in interest rates and defaults of competing banks.

2.4.1 Observation 5: nonbanks exhibit higher observable risk

Table 4 shows that nonbanks are associated with greater overall observable risk, with nonbank-non ntechs having a 9 basis point higher default probability and ntechs having a 10 basis point higher default probability compared to banks purely on the basis of credit score, LTV, and DTI, which corresponds to a 17% or 19% increase relative to the corresponding probability for banks. In most cases nonbanks are also associated with greater risk as measured by each component of observable risk. Summarizing these results graphically, Figure C.1 in Appendix Section C.3 shows that the kernel density and cumulative distribution function of observable risk are slightly more concentrated at higher values for nonbanks compared to banks. The histograms in Figure C.2 in Appendix Section C.3 show that a similar comparison holds for the components of observable risk.

2.4.2 Observation 6: nonbanks exhibit more defaults conditional on observable risk

Figure 4 shows that nonbanks are associated with greater default rates conditional on observable risk. Similarly, Table 5 column (1) shows that nonbanks are associated with higher default rates, column (2) shows that this relationship continues to hold even after controlling for observable risk, column (3) shows that it continues to hold after controlling for a similar set of additional observable characteristics as in column (3) of Table 2, and column (4) shows that it continues to hold after additionally controlling for the interest rate net of the total g-fee. Based on the estimate in column (2), nonbanks are associated with a 19 basis point increase in the default rate conditional on observable risk, which corresponds to 43% of the 44 basis point default rate of banks. Table C.5b in Appendix Section C.4 shows that the results are similar when comparing banks to either ntechs or nonbank-non ntechs.

2.4.3 Observation 7: nonbanks exhibit higher interest rates conditional on observable risk

Figure 5 shows that nonbanks are associated with higher interest rates net of g-fees conditional on observable risk, consistent with their riskier credit profiles (Observation 6). Similarly, Table 6 column (1) shows that nonbanks are associated with higher interest rates, column (2) shows that this relationship continues to hold even after controlling for the g-fee, and column (3) shows that it continues to hold after controlling for a similar set of additional observable characteristics as in column (3) of Table 2. Based on the estimate in column (2), nonbanks are associated with a 8 basis point increase in

Table 4: Observable risk and lender type

(a) Banks

	Mean	P10	P25	P50	P75	P90
Observable risk (%)	0.53	0.07	0.12	0.24	0.56	1.33
Credit score	753.18	689.00	725.00	762.00	789.00	803.00
Loan-to-value (%)	76.38	50.00	67.00	80.00	90.00	95.00
Debt-to-income (%)	32.72	19.39	25.52	33.32	40.54	44.57

(b) Nonbank-non ntechs

	Mean	P10	P25	P50	P75	P90
Observable risk (%)	0.62	0.07	0.13	0.30	0.70	1.48
Credit score	748.19	683.00	717.00	757.00	785.00	801.00
Loan-to-value (%)	76.84	51.00	68.00	80.00	90.00	95.00
Debt-to-income (%)	34.25	21.04	27.37	35.17	41.83	45.12

(c) Fintechs

	Mean	P10	P25	P50	P75	P90
Observable risk (%)	0.63	0.07	0.13	0.30	0.70	1.60
Credit score	746.94	680.00	714.00	756.00	786.00	802.00
Loan-to-value (%)	75.50	51.00	66.00	80.00	90.00	95.00
Debt-to-income (%)	34.07	20.83	26.98	34.76	41.76	45.60

Note: These tables present summary statistics for observable risk characteristics (credit score, loan-to-value ratio (%), debt-to-income ratio (%), and observable risk) for banks, nonbank-non ntechs, and ntechs. Observable risk is the estimated probability of default based on credit score, the loan-to-value ratio, and the debt-to-income ratio as described in Section 2.1. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront origination fee deviates from the first table of the origination fee matrix by more than 25 basis points.

Figure 4: Default, observable risk, and lender type

This figure presents a binned scatterplot of the default rate on observable risk. Observable risk is the estimated probability of default based on credit score, the loan-to-value ratio, and the debt-to-income ratio as described in Section 2.1. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

Table 5: Default, observable risk, and lender type

	(1)	(2)	(3)	(4)
	Baseline	Obs. risk	Controls	+ IR
Nonbank	0.268*** (23.67)	0.187*** (16.66)	0.174*** (15.55)	0.153*** (13.26)
Observable risk		0.954*** (66.73)		
IR - g-fee				0.315*** (6.91)
Observations	2,109,041	2,109,041	2,109,029	2,109,029
R ²	0.091	0.100	0.103	0.103
ZIP Year-quarter FE	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes

Note: Column (1) regresses an indicator for default (multiplied by 100) on an indicator for nonbanks while controlling for ZIP code by year-quarter fixed effects. Column (2) adds observable risk as a regressor. Observable risk is the estimated probability of default based on credit score, the loan-to-value ratio, and the debt-to-income ratio as described in Section 2.1. Column (3) instead includes the following controls: the interaction between 10-point credit score bins (starting at 620, with an additional indicator for all credit scores below 620), 5% loan-to-value bins (starting at 60%, with an additional indicator for all loan-to-value ratios below 60%), and debt-to-income decile indicators (note that this absorbs observable risk); income decile indicators; family type indicators (i.e. single female, single male, or more than 1 borrower); indicators for black and hispanic borrowers; term indicators; appraisal value decile indicators; an indicator for a loan having an interest-only period; an indicator for a refinance loan; loan amount decile indicators; an indicator for self-employed borrowers; an indicator for first-time homebuyers; an indicator for full income documentation; and an indicator for full asset documentation. Column (4) additionally adds the interest rate net of the total g-fee. T-statistics computed using robust standard errors are reported in parentheses. * indicates statistical significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

the interest rate conditional on observable risk, which corresponds to around 18% of a standard deviation. Table C.5 in Appendix Section C.5 shows that the results are similar when comparing banks to either ntechs or nonbank-non ntechs.

Figure 5: Interest rates, observable risk, and lender type

This figure presents a binned scatterplot of the interest rate net of the total g-fee on observable risk for banks and nonbanks while controlling for year-month fixed effects. Observable risk is the estimated probability of default based on credit score, the loan-to-value ratio, and the debt-to-income ratio as described in Section 2.1. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

2.4.4 Observation 8: the presence of ntechs is associated with relatively lower interest rates and default rates by banks

Table 7 examines interactions between banks and nonbanks by showing how the differences between them varies with the presence of ntechs, which have disproportionately contributed to the increasing market share of nonbanks, over the period 2011-2017. Note that for this time period we relax the sample restriction of requiring the upfront g-fee to be within 25 basis points of the value of the first table of the g-fee matrix since the

Table 6: Interest rates, observable risk, and lender type

	(1)	(2)	(3)
	Baseline	Obs. risk	Controls
Nonbank	0.085*** (163.73)	0.077*** (151.88)	0.066*** (155.16)
Observable risk		0.094*** (328.55)	
Observations	2,109,041	2,109,041	2,109,029
R ²	0.448	0.476	0.656
ZIP Year-quarter FE	Yes	Yes	Yes
Controls	No	No	Yes

Note: Column (1) regresses the interest rate net of the total g-fee on an indicator for nonbanks while controlling for ZIP code by year-quarter fixed effects. Column (2) adds observable risk as a regressor. Observable risk is the estimated probability of default based on credit score, the loan-to-value ratio, and the debt-to-income ratio as described in Section 2.1. Column (3) instead includes the following controls: the interaction between 10-point credit score bins (starting at 620, with an additional indicator for all credit scores below 620), 5% loan-to-value bins (starting at 60%, with an additional indicator for all loan-to-value ratios below 60%), and debt-to-income decile indicators (note that this absorbs observable risk); income decile indicators; family type indicators (i.e. single female, single male, or more than 1 borrower); indicators for black and hispanic borrowers; term indicators; appraisal value decile indicators; an indicator for a loan having an interest-only period; an indicator for a refinance loan; loan amount decile indicators; an indicator for self-employed borrowers; an indicator for first-time homebuyers; an indicator for full income documentation; and an indicator for full asset documentation. Column (4) additionally adds the interest rate net of the total g-fee. T-statistics computed using robust standard errors are reported in parentheses. * indicates statistical significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

precise data on g-fees is not consistently available before 2016. We also estimate observable risk based on the sample in each year rather than applying the model estimated on 2016-2017.

Similar to Table 5, column (1) shows that banks are associated with lower default rates during this longer time period. Column (2), which adds the interaction of the bank indicator with the share of origination volume in a county in the last year attributable to ntechs (scaled from 0 to 1), shows that the difference between banks and nonbanks widens as ntechs occupy a greater market share. Similar to Table 6, column (3) shows that banks are associated with lower interest rates, while column (4) shows that the difference between banks and nonbanks widens as ntechs occupy a greater market share. Similarly, Figure C.3 in Appendix Section C.6 shows that nonbanks exhibited a relative increase in default rates and interest rates compared to banks during this period, which was relatively more pronounced in counties with a greater presence of ntechs in 2017. These results are consistent with nonbanks competing more aggressively for the riskier borrowers for a given level of observable risk, resulting in banks focusing on relatively safer borrowers.

2.4.5 Observations 5-8 interpretation

Section 3 rationalizes Observations 5-8 with a model in which nonbank lenders have a lower loss expected given default, which gives them a greater incentive to lend to borrowers with greater observable and unobservable risk while also collecting a higher interest rate and causing rival lenders to shift to safer borrowers with lower interest rates. This lower expected loss given default could be attributable to nonbanks typically being monolines and therefore having less of a concern to protect profits from other product offerings.

3 Model

The preceding section provides evidence that lenders price for credit risk in a manner indicative of performing additional screening relative to the GSEs' g-fees. This section develops a model of mortgage lender competition which is consistent with these observations and which additionally shows that active intermediation can lead to either higher or lower interest rates, depending on a borrower's observable risk, compared to a counterfactual in which lenders passively implement the GSEs' underwriting criteria. We also use the model to show that the differences between banks and nonbanks are

Table 7: Interest rates, default, observable risk, lender type, and ntech market share

	(1)	(2)	(3)	(4)
	Default	Default	IR	IR
Bank	-0.129*** (-22.34)	-0.077*** (-5.72)	-0.078*** (-250.97)	-0.073*** (-98.10)
Obs. risk	1.028*** (146.99)	1.027*** (146.96)	0.180*** (947.05)	0.180*** (946.83)
Lag ntech share		-0.531*** (-4.06)		-0.047*** (-6.72)
Observations	8,678,978	8,678,960	8,678,978	8,678,960
R ²	0.030	0.030	0.453	0.453
County Year-quarter FE	Yes	Yes	Yes	Yes
Controls	No	No	No	No

Note: Column (1) regresses an indicator for default (multiplied by 100) on an indicator for banks while controlling for observable risk and county by year-quarter fixed effects. Column (2) adds the interaction of the bank indicator with the share of origination volume in a county in the last year attributable to ntechs (scaled from 0 to 1). Column (3) and column (4) are similar to column (1) and column (2) except using the interest rate as the dependent variable. T-statistics computed using robust standard errors are reported in parentheses. * indicates statistical significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2011-2017, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, and loans with subordinate financing.

consistent with the latter having a lower expected loss given default and consider the implications for the increasing market share of nonbanks.

3.1 Agents

There are two types of agents: consumers and lenders. All agents are risk neutral.

A consumer can either buy a house requiring 1 unit of external capital or take an outside option whose value is normalized to zero. Consumers are willing to pay up to A in financing costs. There are two quality types q of consumers: type d consumers default, while type r consumers repay the loan.¹⁷ Lenders cannot perceive the type of an individual consumer, but they know the frequencies of the two types in the population, $l_d = \frac{1}{2}$ and $l_r = 1 - l_d$.

3.2 Timeline overview

In period $t = 0$, lenders invest in underwriting technology, which could involve improving their risk assessment models, investing labor hours in careful loan processing, and potentially also collecting additional information about applicants beyond what is required for the GSEs' underwriting criteria.

In period $t = 1$, a consumer applies for a loan with a set of lenders. To focus on the discretionary behavior of lenders as distinct from the underwriting processes of the GSEs, we specifically consider a loan that satisfies the GSEs' underwriting criteria. The lenders first estimate the consumer's default risk, which is represented by allowing each lender to independently draw a signal whose informativeness depends on the quality of its screening technology. Then, lenders that perceive the consumer as too risky reject the consumer's application, while the remaining lenders compete with each other.

In period $t = 2$, the consumer receives the outside option payoff of zero if it did not obtain funding, otherwise, it either repays the loan or defaults.

An elaboration of the model follows in approximately backward order.

¹⁷For simplicity, we abstract from prepayment risk, which is more pertinent in the context of servicing. We focus instead on a lender's losses from originating mortgages that default, such as repurchases or other penalties imposed by the GSEs. In particular, prepayments could be included in the set of repaid loans.

3.3 Risk estimates

This section shows how lenders estimate the default risk of a consumer, which can be described in two parts. First, each lender draws a signal from a distribution that depends on the quality of its underwriting and the quality of the consumer. Second, a lender then adjusts this estimate to take into account the additional information that it would learn conditional on being chosen by the borrower.

3.3.1 Risk estimate conditional on a lender's own signal

Consider a consumer that applies for a loan with n lenders. Suppose that each lender i has some information about the consumer, which is represented by the information level $y_i \in [0, \bar{y}]$ that summarizes the quality of the screening process.¹⁸ For simplicity, we focus on a symmetric equilibrium in which all lenders have the same information level y . See Section 5 for a version of the model where lenders can have different information levels.

At the beginning of the loan application phase, each lender independently draws a privately observed signal $s_i \in [0, 1]$ that depends on the consumer's quality type and the information level of the lenders according to the pdf¹⁹

$$f(s|d; y) = \frac{1}{2} \left(1 + s \frac{y}{l_d} \right) \quad (2)$$

$$f(s|r; y) = \frac{1}{2} \left(1 - s \frac{y}{l_r} \right) \quad (3)$$

The information level corresponds to the precision of the signal. For example, if $y = 0$ then both types produce a uniformly distributed signal, whereas the signal distributions become more differentiated, and the signal therefore becomes more informative, as y increases.

The posterior risk of default conditional on receiving signal s with information y can be

¹⁸Note that the information level is bounded to ensure that the signal distributions described in (2) and (3) are nonnegative.

¹⁹Up to a first order approximation in y , this distribution system can be assumed without loss of generality conditional on the following set of intuitive properties: the predictive distribution does not depend on the information level, the conditional distributions converge to the predictive distribution when the information level is equal to zero, and the first order effect of information on the conditional pdf for a good signal is given by the probability of receiving as high a signal under the predictive distribution. See Appendix Section D.2 for details.

expressed as

$$\begin{aligned} D(s; y) &= \Pr(d|s; y) \\ &= l_d + \frac{1}{2} s y \end{aligned} \quad (4)$$

The properties of the posterior risk are represented graphically in Figure 6. The posterior risk is decreasing in the signal and equal to the prior l_d at the threshold point $s = \frac{1}{2}$. The strength of a signal in shifting the prior is increasing in its distance from this threshold as well as the information level.

Figure 6: Posterior risk conditional on one signal

This figure shows how a lender's posterior risk $D(s; y)$ varies with its signal s and information level y .

3.3.2 Adjustment of risk estimate conditional on supplying the loan

Conditional on the signals, competition among the lenders is formally represented as a second-price sealed-bid auction where the bids correspond to interest rate offers.²⁰ Note that the supplying lender, or the lender with the most competitive offer, can make an inference about the signal of the next most competitive lender based on the equilibrium outcome, resulting in an adjustment of its estimated posterior risk of default.²¹ By a general result for common value auctions from [Milgrom \(1981\)](#), there is a symmetric equilibrium in which each lender's interest rate offer is based on the minimum posterior risk of default that it could have conditional on supplying the loan and updating its posterior risk based on the equilibrium outcome.²² Conditioning on winning the auction accounts for the “winner's curse”, or the tendency for the winner of a common value auction to have an over-optimistic assessment.

Denote the j th order statistic of k signal draws by $s_{j:k}$. To capture the additional information acquired after observing the equilibrium outcome, it is helpful to consider the posterior risk conditional on the lender's own signal s and inferring from the equilibrium interest rate the signal of the next most competitive lender t .²³

$$D(s, t; y, n) = \Pr(d | s_{n:n} = s, s_{n-1:n} = t; y, n) \Big|_{y=0} \approx 1 + \frac{1}{2}(n-2s-nt)y \quad (5)$$

The minimum posterior risk conditional on winning with signal s occurs when the next

²⁰The auction is analogous to Bertrand competition except that the perceived costs of producing loans are based on estimates of a common cost based on the borrower's quality. For simplicity, the auction is assumed to be sealed-bid so that the condition of winning the auction and the equilibrium interest rate are the only sources of information about the signals of the other lenders. The assumption that banks cannot observe each other's offers is similar to other models of bank competition with screening, such as [Broecker \(1990\)](#), [Cao and Shi \(2001\)](#), and [Ruckes \(2004\)](#).

²¹Note that the supplying lender can infer the signal of the next most competitive lender exactly if any other lender makes an offer, as it will be reflected in the equilibrium interest rate. If no other lenders make an offer, then the supplying lender can only infer that all the other lenders received a small enough signal to discourage lending.

²²See Appendix Section [B.1](#) for a proof in this environment.

²³See Appendix Section [B.2](#) for a calculation. The notation $\Big|_{y=0}$ indicates that the expression is a first order approximation around $y = 0$.

most competitive lender has the same posterior risk,²⁴ which can be expressed as

$$D(s, s; y, n) = l_d + \frac{1}{2}(n - (n + 2)s)y \quad (6)$$

The minimum posterior risk conditional on winning the auction qualitatively inherits some of the properties of the posterior risk conditional on just the signal, $D(s; y)$. Specifically, the minimum posterior risk is decreasing in the signal, and information increases the strength of the signal.

3.4 Interest rate offers

A lender that is willing to lend to the consumer participates in the auction by offering an interest rate R . Again, to focus on the decisions of lenders as distinct from the GSEs, we assume that the interest rate is net of g-fees. Suppose that the cost of funding is equal to r . If the loan defaults, then the lender incurs an expected loss given default of $w > 0$ due to, for example, repurchase risk (Goodman (2017)). A lender's expected profits upon winning the auction can therefore be expressed as

$$(1 + R) - wD - (1 + r) = R - (wD + r) \quad (7)$$

The zero-profit interest rate can be written as a markup over the cost of funds that corresponds to the risk

$$\underline{R}(D) = wD + r \quad (8)$$

A lender's interest rate offer is equal to the zero-profit interest rate corresponding to the minimum posterior risk conditional on being chosen by the consumer (see equation (6)).²⁵

Proposition 1. A lender's interest rate offer is equal to

$$\underline{R}(D(s, s; y, n)) = w l_d + \frac{1}{2}(n - (n + 2)s)y + r \quad (9)$$

²⁴This is because a lender wins only if it has a lower posterior risk. Hence, the information contained in the other lender's action can only lead the supplying lender to increase its posterior risk estimate. This adjustment is relatively small when the next most competitive lender has a similar estimate of the borrowers' default risk and vanishes in the limiting case where the estimated default risk of the two lenders is the same.

²⁵See Appendix Section B.1 for a proof. Note that, as shown in the proof, a lender making this offer always obtains a nonnegative expected payoff upon supplying the loan.

If only one lender has a sufficiently optimistic signal to offer a loan, then it fully appropriates any potential surplus by charging an interest rate that is equal to the consumer's willingness to pay, A . This formal convention is consistent with the motivating intuition that a consumer's bargaining power derives from leveraging competing offers from other informed lenders.²⁶

3.5 Participation decision

A lender offers a loan on the condition that its expected profits after learning about the action of the next most competitive lender from the equilibrium outcome will be non-negative. Consider a symmetric equilibrium in which there exists a threshold \underline{s} such that a lender makes an offer when $s_i \geq \underline{s}$.^{27,28}

Proposition 2. A lender's participation threshold is equal to

$$\underline{s} = \frac{n}{n+1} + \frac{2(wl_d + r - A)}{(n+1)wy} \quad (10)$$

Note that lenders make all of their decisions, including whether to participate and any interest rate offer, simultaneously.

3.6 Equilibrium summary

The resolution of the auction is summarized as follows. If more than one lender is willing to lend, then the lenders offer $\underline{R}(D(s_i, s_i; y, n))$ and the lender with the lowest offer supplies the loan at the second lowest offer, $\underline{R}(D(s_{n-1:n}, s_{n-1:n}; y, n))$. If only one lender is willing to lend, then it charges the maximum possible interest rate A . If no lender obtains a sufficiently optimistic assessment to be willing to offer credit (i.e. $s_i < \underline{s}$ for each lender i), then the consumer takes the outside option.

Note that a few key properties of the model are as follows:

²⁶Note that the resolution of the equilibrium in the case where only one lender is willing to lend is a modeling choice since the second-price auction is not defined.

²⁷See Appendix Section B.3 for a calculation.

²⁸Note that \underline{s}_i need not occur in the support of the signal, $[0, 1]$, but the conclusion is still the same. That is, if $\underline{s}_i = 0$ then the lender always makes an offer, and if $\underline{s}_i = 1$ then the lender never makes an offer.

1. The probability of receiving credit (averaged over both types of borrowers) is given by

$$l_d \Pr(s_{n:n} > \underline{s}_d; y) + l_r \Pr(s_{n:n} > \underline{s}_r; y). \quad (11)$$

2. The default rate is the fraction of borrowers receiving credit that default:

$$\frac{l_d \Pr(s_{n:n} > \underline{s}_d; y)}{l_d \Pr(s_{n:n} > \underline{s}_d; y) + l_r \Pr(s_{n:n} > \underline{s}_r; y)}. \quad (12)$$

3. The expected interest rate (averaged over both types of borrowers) is given by:

$$R_{\text{exp}} = l_d E_{s_{n-1:n}} \left[\frac{h}{h} \underline{R}(D(s_{n-1:n}, s_{n-1:n}; y, n)) \mathbb{1}_{s_{n-1:n} > \underline{s}_d; y} \right] + l_r E_{s_{n-1:n}} \left[\underline{R}(D(s_{n-1:n}, s_{n-1:n}; y, n)) \mathbb{1}_{s_{n-1:n} > \underline{s}_r; y} \right]. \quad (13)$$

3.7 A lender's information acquisition decision

Lenders can acquire information $h \geq 0$ with a convex acquisition cost $\frac{m^2 h^2}{2}$, which could represent lenders developing more sophisticated risk assessment models or investing labor hours in careful loan processing. In general, we can also allow for there to be information that is relatively costless to process $z \geq 0$, in which case the total information level is the sum $y = z + h$. However, we focus on the case where z is equal to 0.

The value of information to a lender consists in efficiently providing credit to applicants that are likely to repay as well as undercutting competitors with noisier signals. For simplicity, consider a symmetric equilibrium in which all lenders commit to acquire the same information level. In particular, lenders choose the level of information $y \in [0, \bar{y}]$ to maximize their total expected profit

$$E_{s, s_{n-1:n}} \left[\frac{A}{2} \left(\frac{\underline{R}(D(s, s_{n-1:n}; y, n))}{z} \right) \mathbb{1}_{s = s_{n:n} \leq s_{n-1:n} \leq \underline{s}_d} \right] + E_{s, s_{n-1:n}} \left[\frac{A}{2} \left(\frac{\underline{R}(D(s, s_{n-1:n}; y, n))}{z} \right) \mathbb{1}_{s = s_{n:n} \leq s_{n-1:n} \leq \underline{s}_r} \right] - \frac{m(y-z)^2}{2}, \quad (14)$$

= R as in (7), with no competing lenders

= R as in (7), with competing lenders

where the expectation averages over cases where the lender wins the auction and obtains a profit corresponding to the difference between the interest rate that it collects and its zero-profit interest rate.²⁹

4 Comparing active and passive intermediation

4.1 Definitions

We define active intermediation as the system described in the model. We define passive intermediation as a system in which there is no additional screening for an application has satisfied the GSEs' underwriting criteria. In particular, all applications that are accepted by the GSE underwriting criteria are offered a loan, and the interest rate for a given level of observable risk is determined by a simple zero-profit condition. Note that passive intermediation yields the same result as a market with perfect competition. We can summarize some key properties of passive intermediation as follows:

1. Interest rate = $\underline{R}(l_d) = r + w l_d$
2. Probability of receiving credit = 1 if $\underline{R}(l_d) \leq A$ or 0 if $\underline{R}(l_d) > A$ since the interest rate exceeds the borrower's willingness to pay
3. Default rate = l_d

We focus on which system results in a lower average interest rate. Under active intermediation, the interest rate can be decomposed into 3 components: the ex-ante information cost, an origination cost that corresponds to the zero-profit interest rate $\underline{R}(D) = r + wD$, and a residual markup that corresponds to a lender's total expected profits. In particular, if the expected equilibrium interest payment under active intermediation is R_{exp} , then

²⁹See Appendix Section B.4 for a calculation of a lender's problem.

the components can be written as follows:

$$\begin{aligned}
 R_{\text{exp}} = & \underbrace{nm \frac{(y - z)^2}{z^2}}_{\text{total information cost}} \\
 & + \underbrace{E_{s_{n:n}, s_{n-1:n}} [R(D(s_{n:n}, s_{n-1:n}; y, n))]}_{\text{origination cost}} \\
 & + \underbrace{R_{\text{exp}} \frac{nm \frac{(y - z)^2}{2} + E_{s_{n:n}, s_{n-1:n}} [R(D(s_{n:n}, s_{n-1:n}; y, n))]}{z}}_{\text{markup}} \quad (15)
 \end{aligned}$$

On the one hand, active intermediaries may exhibit lower origination costs since they lend to borrowers that are more likely to repay. On the other hand, active intermediaries also have information acquisition costs and an opportunity to charge a markup, i.e.

$$\begin{aligned}
 R_{\text{exp}} \frac{R(l_d)}{0} = & \underbrace{\text{origination cost}}_{0} \frac{R(l_d)}{z} \\
 & + \underbrace{\text{information cost} + \text{markup}}_{0} \frac{R(l_d)}{z} \quad (16)
 \end{aligned}$$

4.2 Model simulation

In our simulations, we focus on how model outcomes vary with l_d . We therefore normalize the cost of funding r to be zero, as it only serves to create a level shift of interest rates that can be used to capture time-varying factors that are not the focus of this exercise. We select fixed values for the other parameters (n, A, w, m) based on the following considerations.

The number of lenders n is directly selected to be 2, which is the median number of lenders that are seriously considered by borrowers according to the National Survey of Mortgage Originations ([Bhutta, Fuster, and Hizmo \(2021\)](#), [Alexandrov and Koulayev \(2018\)](#)).

The remaining parameters (A, w, m) are selected to match the intensive and extensive margins of overlays:

1. The intensive margin of overlays is represented by column (3) of Table 1, which shows that an increase in observable risk is on average associated with a .094 percent increase in interest rates net of g-fees. Note that we determine the model analog by computing the average interest rate (over repaying and defaulting borrowers) for each level of observable risk and then taking a weighted average over observable risk based on the empirical distribution.
2. The extensive margin of overlays is represented by Figure 3, which shows the rejection rate of applications accepted by the GSEs. We select the parameters to match the average acceptance rate of .963 and the slope with respect to observable risk of -.0173, again weighting over observable risk based on the empirical distribution.

Table 8 presents the selected parameters, while Table 9 compares the empirical and model-generated values of the matched characteristics.

Table 8: Selected parameters

Parameter	Value
Lenders (n)	2
Borrower willingness to pay (λ)	0.0026
Loss given default (w)	0.065
Information cost (m_b)	0.1

Table 9: Comparison of empirical and model-generated variables

Variable	Empirical	Model
Average effect of observable risk on interest rate - g-fee	0.094	0.091
Average acceptance rate	0.963	0.969
Average effect of observable risk on acceptance rate	-.017	-.02

Figure 7 compares active and passive intermediation for varying degrees of I_d . For low levels of risk, active intermediaries do not reject any applications. In that case, active intermediation exhibits the same default rate and origination cost as passive intermediation. However, lenders still screen the borrower to determine the interest rate, and disparities in the assessment of the borrower's risk provide an opportunity for lenders to obtain markups. Therefore, active intermediation is associated with a higher interest rate compared to passive intermediation. The figure also illustrates that the opportunity

to obtain markups creates an incentive for lenders to improve their screening processes, which allows them to obtain higher markups as borrower risk increases.³⁰

For sufficiently high I_d , active intermediaries start to reject applications, resulting in a reduction of the default rate compared to passive intermediation.³¹ This in turn leads to a reduction in the origination cost compared to passive intermediation, which can be large enough to also lead to a relative reduction in the overall interest rate compared to passive intermediation.

Additionally, as I_d becomes sufficiently high, the ability to obtain markups generally decreases since the origination cost increases while the maximum interest rate that a lender can charge is fixed at the consumer's willingness to pay A . This in turn dampens the incentive for lenders to acquire information. If I_d becomes too high, then lenders lose the incentive to invest in screening technology at all, causing them to leave the market. However, for borrowers with high observable risk, active intermediaries are more likely to offer lower interest rates or lend at all compared to passive intermediaries.

Figure 8 illustrates the role of the lender's loss given default w in determining overlays. As w increases, lenders deny more applications and have lower default rates. They also charge higher interest rates, which is partly driven by the direct effect of w on origination costs and partly driven by the fact that lenders can charge higher markups since there is less often competition from rival lenders, as shown in Figure D.2 in Appendix Section D.

Finally, Figure 9 illustrates the role of competition. As the number of lenders increases, markups naturally decrease. This in turn reduces the incentive to invest in screening, resulting in a higher default rate and origination cost. Therefore, increasing competition is more likely to lead to lower interest rates in the safer segments of the market where the markup occupies a greater share of the interest rate, but it is also more likely to lead to higher interest rates in the risky segments of the market where the origination cost occupies a greater share of the interest rate. This result is consistent with Yannelis and

³⁰See Figure D.5 in Appendix Section D for the direct association between interest rates and default as functions of I_d .

³¹See Figure D.1 in Appendix Section D for a more detailed decomposition of the number of offers received by a borrower.

Figure 7: Active and passive intermediation

These figures show various features of the model for active intermediation (the baseline model in which lenders screen the applicant, approve or deny the application, and engage in imperfect competition to determine the interest rate) and passive intermediation (a setting where all applications approved by the AUS are originated and the interest rate is determined by a zero-profits condition) for various levels of λ . The probability of receiving credit is the probability that at least one lender approves the application. The default rate is the fraction of approved applications that consist of defaulting borrowers. The interest rate is the average interest payment divided by the probability of receiving credit. The interest rate for active intermediation is decomposed as the origination cost (which is the zero-profits interest rate of the supplying lender conditional on its own signal and inferring from the equilibrium the signal of the next most competitive lender), the information cost (which is the cost associated with the parameter γ corresponding to the quality of screening), and a residual markup (which corresponds to a lender's total expected total profits). $\Pr(1 \text{ offer} | \text{receiving credit})$ is the probability that the consumer receives only one offer conditional on receiving credit. Parameters: $r = 0$, $n = 2$, $A = .0026$, $w = .065$, $m = .1$.

Figure 8: Active intermediation as w varies

These figures show various features of the model for a low loss given default w and a high loss given default. The probability of receiving credit is the probability that at least one lender approves the application. The default rate is the fraction of approved applications that consist of defaulting borrowers. The average interest rate is the average interest payment divided by the probability of receiving credit. The average origination cost is the average zero-profits interest rate of the supplying lender conditional on its own signal and inferring from the equilibrium the signal of the next most competitive lender. The average information cost is the lenders' combined cost associated with the parameter γ corresponding to the quality of screening divided by the probability of receiving credit. The average markup is the lenders' combined expected profits (average interest rate - average origination cost - information cost) divided by the probability of receiving credit. Parameters: $r = 0$, $n = 2$, $A = .0026$, $w = .065$ and $.09$, $m = .1$.

Zhang (2021), who provide evidence of a similar effect of competition in the context of auto loans. To summarize, stronger competition tends to reduce the differences between active and passive intermediation, and in fact active intermediation converges to passive intermediation as n increases.

5 Heterogeneous lenders

This section presents versions of the model with heterogeneous lenders. We find that the observed differences between banks and nonbanks are more consistent with differences in the implied loss given default w than screening quality y .

5.1 Heterogeneous y

For simplicity, we assume that there are only 2 lenders with exogenous information levels y_1 and $y_2 < y_1$. Since the information levels are exogenous, we abstract from the information cost. The presentation of the model in this section is brief and focuses on differences relative to the original model in Section 3.

The posterior risk of default conditional on receiving signal s_i with information level y_i is directly analogous to equation (4) and can be expressed as

$$\begin{aligned} D(s_i; y_i) &= \Pr(d|s_i; y_i) \\ &= l_d + \frac{1}{2} (s_i - y_i) \end{aligned} \quad (17)$$

The posterior risk conditional on the lender's own signal and inferring from the equilibrium interest rate the signal of the competing lender becomes³²

$$\begin{aligned} D(s_1, s_2; y_1, y_2) &= \Pr(d|s_1, s_2; y_1, y_2) \\ &= l_d + \frac{1}{2} (s_1 - y_1) + \frac{1}{2} (s_2 - y_2) \end{aligned} \quad (18)$$

As before, there is an equilibrium in which each lender's interest rate offer is based on the minimum posterior risk of default that it could have conditional on supplying the

³²See Section B.5 for a calculation.

Figure 9: Active intermediation as n varies

These figures show various features of the model when the number of lenders n is equal to 2 or 3. The probability of receiving credit is the probability that at least one lender approves the application. The default rate is the fraction of approved applications that consist of defaulting borrowers. The average interest rate is the average interest payment divided by the probability of receiving credit. The average origination cost is the average zero-profits interest rate of the supplying lender conditional on its own signal and inferring from the equilibrium the signal of the next most competitive lender. The average information cost is the lenders' combined cost associated with the parameter γ corresponding to the quality of screening divided by the probability of receiving credit. The average markup is the lenders' combined expected profits (average interest rate - average origination cost - information cost) divided by the probability of receiving credit. Parameters: $r = 0$, $n = 2$ and 3 , $A = .0026$, $w = .065$, $m = .1$.

loan and updating its posterior risk based on the equilibrium outcome.³³ The minimum posterior risk conditional on winning for lender i occurs when the competing lender has the same posterior risk, which can be expressed as³⁴

$$D(s_i, s_i; y_i, y_i) = l_d + (1 - 2s_i) y_i \quad (19)$$

There is an equilibrium in which each lender's interest rate offer is equal to the zero-profits interest rate corresponding to the minimum posterior risk conditional on being equal to the zero-profits interest rate corresponding to the minimum posterior risk conditional on being chosen by the consumer (see equation (19)).³⁵

Proposition 3. A lender's interest rate offer is equal to

$$\underline{R}(D(s_i, s_i; y_i, y_i)) = w[l_d + (1 - 2s_i) y_i] + r \quad (20)$$

If only one lender makes an offer, it charges the maximum possible interest rate, A .

Each lender offers a loan on the condition that it will achieve nonnegative expected profits after learning about the action of the other lender from the equilibrium outcome. We consider an equilibrium in which each lender has a threshold \underline{s}_i such that it makes an offer when $s_i \geq \underline{s}_i$. We determine \underline{s}_i as the signal at which a lender would achieve zero expected profits assuming the other lender does not make an offer. This determines the following participation threshold.³⁶

Proposition 4. A lender's participation threshold is equal to

$$\underline{s}_i = \frac{2}{3} \frac{y_1 + y_2}{2y_i} + \frac{w[l_d + r] - A}{wy_i} \quad (21)$$

³³See Appendix Section B.6 for a proof.

³⁴Recall from equation (23) that the posterior risk for lender i is $l_d + \frac{1}{2} s_i y_i$. Therefore, if lender i has the same posterior risk as lender i then we must have $\frac{1}{2} s_i y_i = \frac{1}{2} s_i y_i$. Substituting this into equation (18) obtains equation (19).

³⁵See Appendix Section B.6 for a proof.

³⁶See Appendix Section B.7 for a proof.

If $s_i = 1$ (where s_i is the signal of the other lender), then π_i is instead given by

$$s_i = \frac{1}{2} + \frac{w_l d + r - A}{w y_i} \quad (22)$$

Figure 10 shows the probability of receiving credit from the more or less informed lender as well as outcomes associated with the two types of lenders for varying levels of l_d .³⁷ The more informed lender is associated with a greater willingness to provide credit, which generally becomes more pronounced for riskier borrowers.³⁸ The more informed lender is also associated with a lower default rate and origination cost as well as a higher markup. In this case, the more informed lender is associated with a lower average interest rate for low l_d borrowers but a higher interest rate for high l_d borrowers.³⁹

5.2 Heterogeneous w

In this section we assume there are 2 lenders with the same exogenous but different degrees of the expected loss given default: $w_2 < w_1$. Furthermore, we assume that the lender with lower loss given default has a higher cost of funding in order to maintain the property that the two lenders are equally competitive in the benchmark case of no screening, i.e. $r_2 = r_1 + l_d(w_1 - w_2)$. In particular, this implies that the expected origination costs without screening for the two lenders satisfy $r_1 + w_1 l_d = r_2 + w_2 l_d$. As in Section 5, we abstract from the information cost since the information level is exogenous.

The posterior risk of default conditional on receiving signal s_i with information level y is directly analogous to equation (4) and can be expressed as

$$\begin{aligned} D(s_i; y) &= \Pr(d|s_i; y) \\ &= l_d + \frac{1}{2} s_i y \end{aligned} \quad (23)$$

The posterior risk conditional on the lender's own signal and inferring from the equilib-

³⁷Note that l_d starts a point greater than 0 due to the constraint $y \geq 2l_d$ as described in Section 3.3.1.

³⁸See Figure D.3 in Appendix Section D for a more detailed decomposition of the number of offers received by a borrower.

³⁹See also Figure D.6 in Appendix Section D for the direct association between interest rates and default as functions of l_d .

Figure 10: Active intermediation with heterogeneous γ

These figures show various features of the version of the model with 2 lenders with exogenous and different information levels γ (described in Section 5.1). The probability of receiving credit is the probability that at least one lender approves the application. The default rate is the fraction of approved applications that consist of defaulting borrowers. The average interest rate is the average interest payment divided by the probability of receiving credit. The average origination cost is the average zero-profit interest rate of the supplying lender conditional on its own signal and inferring from the equilibrium the signal of the next most competitive lender. The average markup is a lender's total expected profits (average interest rate - average origination cost). $\Pr(1 \text{ offer} | \text{receiving credit})$ is the probability that the consumer receives only one offer conditional on receiving an offer. Parameters: $r = 0$, $n = 2$, $A = .0026$, $w = .065$, $m = .1$, $\gamma = .015$ and $.02$.

rium interest rate the signal of the competing lender becomes ⁴⁰

$$D(s_1, s_2; y) = \Pr(d_j s_1, s_2; y) \quad (24)$$

Lender i 's zero-profit interest rate is then

$$\underline{R}_i(D(s_1, s_2; y)) = w_i [l_d + (1 - s_1 - s_2) y] + r_i \quad (25)$$

In contrast to the baseline model and the version with heterogeneous y , the version with heterogeneous w is not a common value auction. As a result, we take a different strategy to derive the bid functions. In particular, we consider the space of linear bid functions $B_i(s_i) = a_i + b_i s_i$ and suppose that each lender chooses a_i and b_i in order to maximize its expected profits over realizations of the other lender's bid:

$$E_{s_{-i}} \left(B_i(s_i) - \underline{R}_i(D(s_1, s_2; y)) \right) 1_{a_i + b_i s_i < B_{-i}(s_{-i})} \quad (26)$$

This determines the following equilibrium. ⁴¹

Proposition 5. There is an equilibrium in which the bidding strategies are given by

$$B_i(s_i) = w_i l_d + w_i \frac{1}{2} s_i y + r_i \quad (27)$$

Additionally, each lender offers a loan on the condition that it will achieve nonnegative expected profits after learning about the action of the other lender from the equilibrium outcome. We consider an equilibrium in which each lender has a threshold \underline{s}_i such that it makes an offer when $s_i \geq \underline{s}_i$. We determine \underline{s}_i as the signal at which a lender would achieve zero expected profits assuming the other lender does not make an offer. This determines the following participation threshold. ⁴²

⁴⁰This follows from the proof in Appendix Section B.2 for the case of 2 lenders.

⁴¹See Appendix Section B.8 for a proof.

⁴²See Appendix Section B.9 for a proof.

Proposition 6. A lender's participation threshold is equal to

$$\underline{s}_i = \frac{2}{3} + 1 + \frac{2w_i - w_i}{w_1 w_2} \frac{(w_i l_d + r_i - A)}{y} \quad (28)$$

If $\underline{s}_i < 1$ (where \underline{s}_i is the signal of the other lender), then \underline{s}_i is instead given by

$$\underline{s}_i = \frac{1}{2} + \frac{w_i l_d + r_i - A}{w_i y} \quad (29)$$

Figure 11 shows the probability of receiving credit from the lender with greater or lower loss given default as well as outcomes associated with the two types of lenders for varying levels of l_d . In this case, the lender with lower loss given default is associated with a greater willingness to provide credit, which generally becomes more pronounced for riskier borrowers.⁴³ The lender with lower loss given default is also associated with a higher default rate, origination cost, and overall interest rate.⁴⁴ Figure D.8 in Appendix Section D shows how the participation thresholds \underline{s}_i vary with l_d . It suggests that the lender with greater loss given default restricts to sufficiently strong borrowers, leaving many borrowers with weaker credit assessments to be serviced by the lender with lower loss given default. Figure D.9 in Appendix Section D compares the case of 2 lenders with different w to the case of 2 homogeneous lenders with the same w .

5.3 Comparison with empirical observations

The empirical observations comparing banks and nonbanks are most consistent with nonbanks having a lower expected loss given default w :

1. The upper left sub figure of Figure 11 shows that lenders with lower w tend to lend relatively more to observably risky borrowers, which matches the observation that nonbanks exhibit higher observable risk (Observation 5 in Section 2.4.1).
2. The upper right sub figure of Figure 11 shows that lenders with lower w tend to have higher default rates conditional on observable risk, which matches the observation that nonbanks exhibit more defaults conditional on observable risk (Observation 6 in Section 2.4.1).

⁴³See Figure D.4 in Appendix Section D for a more detailed decomposition of the number of offers received by a borrower.

⁴⁴See also Figure D.6 in Appendix Section D for the direct association between interest rates and default as functions of l_d .

Figure 11: Active intermediation with heterogeneous w

These figures show various features of the version of the model with 2 lenders with exogenous and the same information levels y but different losses given default w (described in Section 5.2). The probability of receiving credit is the probability that at least one lender approves the application. The default rate is the fraction of approved applications that consist of defaulting borrowers. The average interest rate is the average interest payment divided by the probability of receiving credit. The average origination cost is the average zero-profits interest rate of the supplying lender conditional on its own signal and inferring from the equilibrium the signal of the next most competitive lender. The average markup is a lender's total expected profits (average interest rate - average origination cost). $\Pr(1 \text{ offer} | \text{receiving credit})$ is the probability that the consumer receives only one offer conditional on receiving an offer. Parameters: $n = 2$, $A = .0026$, $w = .065$ and $.04$, $m = .01$, $y = .02$.

vation 6 in Section 2.4.2).

3. The middle left subfigure of Figure 11 shows that lenders with lower γ tend to have a higher interest rate, which matches the observation that nonbanks exhibit higher interest rates conditional on observable risk (Observation 7 in Section 2.4.3).

By contrast, the empirical results are not consistent with banks and nonbanks having heterogeneous γ . In particular, Figure 10 shows that lenders with lower γ tend to lend relatively more to observably risk borrowers but also tend to have lower default rates conditional on observable risk, which doesn't fit the profile of either banks or nonbanks.

The model yields notable implications of the increasing market share of nonbanks. First, Figure D.8 in Appendix Section D shows that the presence of the low γ lender tends to induce “cream-skimming” behavior in which the high γ lender takes only the safest borrowers for a given level of observable risk, which is consistent with the results in Section 2.4.4 that a larger market share of fintechs is associated with banks having lower default rates and interest rates compared to nonbanks. Second, Figure 8 shows that if the market eventually becomes dominated by nonbanks, then default rates will be higher but interest rates may actually be lower compared to a market dominated by banks, as the former have lower origination costs and compete more aggressively against each other.

6 Conclusion

We provide evidence of active intermediation by lenders of GSE mortgages. Specifically, we show that mortgage interest rates net of g-fees increase with observable risk, consistent with discretionary pricing for risk. Interest rates also predict default conditional on observable risk, consistent with lender screening. We develop a model of mortgage lender competition with screening that explains these observations by supposing that lenders of GSE mortgages face a positive expected loss given default. The model additionally shows that active intermediation can lead to directionally different effects on interest rates relative to a counterfactual passive intermediation scenario for borrowers with different levels of observable risk. This is due to the counteracting effects of higher markups and lower origination costs.

We also show that nonbanks, which comprise an increasing share of the mortgage mar-

ket, exhibit different intermediation patterns compared to banks, such as higher default rates conditional on observable risk, and higher interest rates compared to banks. Non-banks also affect the intermediation activity of banks, as the growing presence of fintechs has been associated with a decrease in interest rates and default rates for banks. Through the lens of the model, the behavior of nonbanks is consistent with them having a lower expected loss given default. The model suggests that the increasing market share of non-banks may lead to an increase in default rates but, in the long run, an overall reduction in interest rates.

From a policy perspective, these results suggest that the added value of implementing the GSE segment of the mortgage market through private intermediaries consists of decreasing the cost of housing credit for observably risky borrowers, albeit at the expense of increasing markups for the majority of borrowers. Additionally, while the increasing presence of nonbank lenders that are more associated with observably risky borrowers could further improve the access to credit, it could also lead to a riskier pool of borrowers, albeit still within the underwriting requirements of the GSEs.

References

- Alexandrov, Alexei and Sergei Koulayev. 2018. “No Shopping in the U.S. Mortgage Market: Direct and Strategic Effects of Providing Information.” Working Paper.
- Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace. 2022. “Consumer-lending discrimination in the FinTech Era.” *Journal of Financial Economics* 143(1): 30–56.
- Benetton, Matteo, Greg Buchak, and Claudia Robles-Garcia. 2022. “Wide or Narrow? Competition and Scope in Financial Intermediation.” Working paper.
- Bhutta, Neil, Andreas Fuster, and Aurel Hizmo. 2021. “Paying Too Much? Borrower Sophistication and Overpayment in the US Mortgage Market.” Working Paper.
- Bhutta, Neil, Aurel Hizmo, and Daniel Ringo. 2021. “How Much Does Racial Bias Affect Mortgage Lending? Evidence from Human and Algorithmic Credit Decisions.”
- Broecker, Thorsten. 1990. “Credit-Worthiness Tests and Interbank Competition.” *Econometrica* 58(2): 429–452.
- Buchak, Greg and Adam Jørring. 2021. “Do Mortgage Lenders Compete Locally? Implications for Credit Access.” Working Paper.

- 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.
- Cao, Melanie and Shouyong Shi. 2001. "Screening, Bidding, and the Loan Market Tightness." *European Finance Review* 5(1-2): 21–61.
- Di Maggio, Marco and Vincent Yao. 2021. "Fintech Borrowers: Lax Screening or Cream-Skimming?" *The Review of Financial Studies* 34(10): 4565–4618.
- Egan, Mark, Ali Hortaçsu, and Gregor Matvos. 2017. "Deposit Competition and Financial Fragility: Evidence from the US Banking Sector." *The American Economic Review* 107(1): 169–216.
- Federal Reserve Bank of New York. 2022. "Quarterly Report on Household Debt and Credit (2022Q3)."
- Fuster, Andreas, Aurel Hizmo, Lauren Lambie-Hanson, James Vickery, and Paul S. Willen. 2021. "How Resilient Is Mortgage Credit Supply? Evidence from the COVID-19 Pandemic." NBER Working Paper 28843.
- Fuster, Andreas, Matthew Plosser, and James Vickery. 2021. "Does CFPB Oversight Crimp Credit?" Federal Reserve Bank of New York Staff Reports, no. 857.
- Fuster, Andreas, Matthew Plosser, Philipp Schnabl, and James Vickery. 2019. "The Role of Technology in Mortgage Lending." *The Review of Financial Studies* 32(5): 1854–1899.
- Fuster, Andreas, Stephanie H. Lo, and Paul S. Willen. 2017. "The Time-Varying Price of Financial Intermediation in the Mortgage Market." NBER Working Paper 23706.
- Gete, Pedro and Michael Reher. 2020. "Mortgage Securitization and Shadow Bank Lending." *The Review of Financial Studies* 34(5): 2236–2274.
- Giacoletti, Marco, Rawley Heimer, and Edison G. Yu. 2022. "Using High-Frequency Evaluations to Estimate Disparate Treatment: Evidence from Mortgage Loan Officers." Working Paper.
- Goodman, Laurie. 2017. "Quantifying the Tightness of Mortgage Credit and Assessing Policy Actions." *Boston College Journal of Law & Social Justice* 37(2): 235–265.

- Hurst, Erik, Benjamin J. Keys, Amit Seru, and Joseph Vavra. 2016. "Regional Redistribution through the US Mortgage Market." *American Economic Review* 106(10): 2982–3028.
- Jagtiani, Julapa, Lauren Lambie-Hanson, and Timothy Lambie-Hanson. 2021. "Fintech Lending and Mortgage Credit Access." *The Journal of FinTech* 1(1).
- Jansen, Mark, Hieu Quang Nguyen, and Amin Shams. 2021. "Rise of the Machines: The Impact of Automated Underwriting." Working paper.
- Johnson, Stephanie. 2022. "Financial Technology and the 1990s Housing Boom."
- Kim, You Suk, Steven M. Laufer, Karen Pence, Richard Stanton, and Nancy Wallace. 2018. "Liquidity Crises in the Mortgage Market." *Brookings Papers on Economic Activity*, Spring 2018 Edition, 347–428.
- Kulkarni, Nirupama. 2016. "Are Uniform Pricing Policies Unfair? Mortgage Rates, Credit Rationing, and Regional Inequality." Working Paper.
- Malliaris, Steven, Daniel A. Rettl, and Ruchi Singh. 2021. "Is competition a cure for confusion? Evidence from the residential mortgage market." *Real Estate Economics* 50(1): 206–246.
- Milgrom, Paul R. 1981. "Rational Expectations, Information Acquisition, and Competitive Bidding." *Econometrica* 49(4): 921–943.
- Ruckes, Martin. 2004. "Bank Competition and Credit Standards." *The Review of Financial Studies* 17(4): 1073–1102.
- Sharpe, Steve A. and Shane M. Sherlund. 2016. "Crowding Out Effects of Renance on New Purchase Mortgages." *Review of Industrial Organization* 48(2): 209–239.
- Valentin, Maxence. 2021. "Interest Deductibility, Market Frictions, and Price Discrimination." Working Paper.
- Yannelis, Constantine and Anthony Lee Zhang. 2021. "Competition and Selection in Credit Markets." NBER Working Paper 29169.
- Zhang, David. 2022. "Closing Costs, Renancing, and Ineficiencies in the Mortgage Market." Working Paper.

Appendix

A Supplemental material for Section 2

A.1 Supplemental material for Section 2.2

Table A.1: Summary statistics for the 2016-2017 sample

(a) Full sample

	N	Mean	SD	P25	P75
Default (%)	2,141,661	0.57	7.52	0.00	0.00
Credit score	2,141,661	751.00	44.78	721.00	787.00
Loan-to-value (%)	2,141,661	76.48	17.45	67.00	90.00
Debt-to-income (%)	2,141,661	33.35	9.51	26.25	41.11
Observable risk (%)	2,141,661	0.57	0.80	0.13	0.63
Interest rate (%)	2,141,661	3.87	0.45	3.62	4.12
G-fee (%)	2,141,661	0.47	0.14	0.41	0.52
Bank	2,141,661	0.58	0.49	0.00	1.00
Nonbank-non ntech	2,141,661	0.35	0.48	0.00	1.00
Fintech	2,141,661	0.07	0.25	0.00	0.00
Income (\$1000s)	2,141,661	78.74	43.65	47.86	98.67
Single female	2,141,661	0.20	0.40	0.00	0.00
Single male	2,141,661	0.29	0.45	0.00	1.00
> 1 borrower	2,141,661	0.52	0.50	0.00	1.00
Hispanic	2,141,661	0.08	0.27	0.00	0.00
Black	2,141,661	0.02	0.16	0.00	0.00
Term (months)	2,141,661	328.26	66.10	360.00	360.00
Appraisal value (\$1000s)	2,141,661	306.27	158.09	188.00	390.50
Re finance	2,141,661	0.39	0.49	0.00	1.00
Interest only	2,141,661	0.00	0.00	0.00	0.00

Note: These tables present summary statistics for the 2016-2017 sample (Table A.1a), the subsample of loans originated by banks (Table A.1b), the subsample of loans originated by nonbank-non ntechs (Table A.1c), and the subsample of loans originated by ntechs (Table A.1d). Default indicates 90-day delinquency within 2 years of origination (multiplied by 100). Credit score is the representative credit score, i.e. the minimum of each borrower's representative score, which is either the lower score if there are two scores or the middle score if there are three. Loan-to-value (LTV) is the ratio of the loan amount to the lesser of the appraised value and the selling price. Debt-to-income (DTI) is the ratio of all debt payments to household income. Observable risk is the estimated probability of default based on credit score, LTV, and DTI as described in Section 2.1. Interest rate is the interest rate at origination. G-fees is the total guarantee fee expressed as an annualized rate. Note that the upfront component of the g-fee is converted to an annualized rate using the respective present value multiplier before being added to the ongoing component of the g-fee. Bank is an indicator for depositories. Nonbank-non ntech is an indicator for lenders that are neither banks nor ntechs. Fintech is an indicator for lenders with a mostly online application process based on the designation of ntechs in Fuster et al. (2019)). Income is the gross income of all borrowers. Single female, single male, and > 1 borrower indicate the number of borrowers and, in the case of 1 borrower, the gender. Hispanic and Black refer to the ethnicity and race of the primary borrower. Term is the number of monthly payments from the origination date until the maturity date of the loan specified as of the origination date. Appraisal value is the appraised value of the collateral for the mortgage. Re finance indicates re finance loans. Interest-only indicates loans with an interest-only portion. Continuous variables are winsorized at 1%. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out re finance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

(b) Banks

	N	Mean	SD	P25	P75
Default (%)	1,240,402	0.44	6.59	0.00	0.00
Credit score	1,240,402	753.18	44.02	725.00	789.00
Loan-to-value (%)	1,240,402	76.38	17.67	67.00	90.00
Debt-to-income (%)	1,240,402	32.72	9.55	25.52	40.54
Observable risk (%)	1,240,402	0.53	0.76	0.12	0.56
Interest rate (%)	1,240,402	3.82	0.44	3.50	4.12
G-fee (%)	1,240,402	0.47	0.13	0.41	0.52
Income (\$1000s)	1,240,402	78.79	44.23	47.45	98.83
Single female	1,240,402	0.19	0.39	0.00	0.00
Single male	1,240,402	0.28	0.45	0.00	1.00
> 1 borrower	1,240,402	0.53	0.50	0.00	1.00
Hispanic	1,240,402	0.07	0.25	0.00	0.00
Black	1,240,402	0.02	0.15	0.00	0.00
Term (months)	1,240,402	326.49	67.80	360.00	360.00
Appraisal value (\$1000s)	1,240,402	295.81	155.29	180.00	378.00
Re finance	1,240,402	0.37	0.48	0.00	1.00
Interest only	1,240,402	0.00	0.00	0.00	0.00

Note: These tables present summary statistics for the 2016-2017 sample (Table A.1a), the subsample of loans originated by banks (Table A.1b), the subsample of loans originated by nonbank-non ntechs (Table A.1c), and the subsample of loans originated by ntechs (Table A.1d). Default indicates 90-day delinquency within 2 years of origination (multiplied by 100). Credit scores the representative credit score, i.e. the minimum of each borrower's representative score, which is either the lower score if there are two scores or the middle score if there are three. Loan-to-value(LTV) is the ratio of the loan amount to the lesser of the appraised value and the selling price. Debt-to-income(DTI) is the ratio of all debt payments to household income. Observable risks the estimated probability of default based on credit score, LTV, and DTI as described in Section 2.1. Interest rate is the interest rate at origination. G-fee is the total guarantee fee expressed as an annualized rate. Note that the upfront component of the g-fee is converted to an annualized rate using the respective present value multiplier before being added to the ongoing component of the g-fee. Bank is an indicator for depositories. Nonbank-non ntech is an indicator for lenders that are neither banks nor ntechs. Fintech is an indicator for lenders with a mostly online application process based on the designation of ntechs in Fuster et al. (2019)). Income is the gross income of all borrowers. Single female, single male, and > 1 borrower indicate the number of borrowers and, in the case of 1 borrower, the gender. Hispanic and Black refer to the ethnicity and race of the primary borrower. Term is the number of monthly payments from the origination date until the maturity date of the loan specified as of the origination date. Appraisal value is the appraised value of the collateral for the mortgage. Re finance indicates re finance loans. Interest-only indicates loans with an interest-only portion. Continuous variables are winsorized at 1%. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out re finance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

Table A.1: Summary statistics (continued.)

(c) Nonbank-non ntechs

	N	Mean	SD	P25	P75
Default (%)	754,848	0.75	8.60	0.00	0.00
Credit score	754,848	748.19	45.35	717.00	785.00
Loan-to-value (%)	754,848	76.84	17.14	68.00	90.00
Debt-to-income (%)	754,848	34.25	9.36	27.37	41.83
Observable risk (%)	754,848	0.62	0.84	0.13	0.70
Interest rate (%)	754,848	3.92	0.46	3.62	4.25
G-fee (%)	754,848	0.48	0.14	0.42	0.53
Income (\$1000s)	754,848	78.59	42.80	48.50	98.27
Single female	754,848	0.20	0.40	0.00	0.00
Single male	754,848	0.31	0.46	0.00	1.00
> 1 borrower	754,848	0.49	0.50	0.00	1.00
Hispanic	754,848	0.10	0.30	0.00	0.00
Black	754,848	0.03	0.16	0.00	0.00
Term (months)	754,848	333.14	61.76	360.00	360.00
Appraisal value (\$1000s)	754,848	325.61	162.14	205.00	415.00
Re nance	754,848	0.39	0.49	0.00	1.00
Interest only	754,848	0.00	0.00	0.00	0.00

Note: These tables present summary statistics for the 2016-2017 sample (Table A.1a), the subsample of loans originated by banks (Table A.1b), the subsample of loans originated by nonbank-non ntechs (Table A.1c), and the subsample of loans originated by ntechs (Table A.1d). Default indicates 90-day delinquency within 2 years of origination (multiplied by 100). Credit scores the representative credit score, i.e. the minimum of each borrower's representative score, which is either the lower score if there are two scores or the middle score if there are three. Loan-to-value(LTV) is the ratio of the loan amount to the lesser of the appraised value and the selling price. Debt-to-income(DTI) is the ratio of all debt payments to household income. Observable risks the estimated probability of default based on credit score, LTV, and DTI as described in Section 2.1. Interest rate is the interest rate at origination. G-fee is the total guarantee fee expressed as an annualized rate. Note that the upfront component of the g-fee is converted to an annualized rate using the respective present value multiplier before being added to the ongoing component of the g-fee. Bank is an indicator for depositories. Nonbank-non ntech is an indicator for lenders that are neither banks nor ntechs. Fintech is an indicator for lenders with a mostly online application process based on the designation of ntechs in Fuster et al. (2019)). Income is the gross income of all borrowers. Single female, single male, and > 1 borrower indicate the number of borrowers and, in the case of 1 borrower, the gender. Hispanic and Black refer to the ethnicity and race of the primary borrower. Term is the number of monthly payments from the origination date until the maturity date of the loan specified as of the origination date. Appraisal value is the appraised value of the collateral for the mortgage. Re nance indicates re nance loans. Interest-only indicates loans with an interest-only portion. Continuous variables are winsorized at 1%. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out re nance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

(d) Fintechs

	N	Mean	SD	P25	P75
Default (%)	146,411	0.78	8.82	0.00	0.00
Credit score	146,411	746.94	47.13	714.00	786.00
Loan-to-value (%)	146,411	75.50	17.06	66.00	90.00
Debt-to-income (%)	146,411	34.07	9.49	26.98	41.76
Observable risk (%)	146,411	0.63	0.86	0.13	0.70
Interest rate (%)	146,411	3.94	0.45	3.67	4.25
G-fee (%)	146,411	0.48	0.14	0.40	0.54
Income (\$1000s)	146,411	79.06	43.12	48.33	99.57
Single female	146,411	0.18	0.39	0.00	0.00
Single male	146,411	0.29	0.45	0.00	1.00
> 1 borrower	146,411	0.53	0.50	0.00	1.00
Hispanic	146,411	0.06	0.24	0.00	0.00
Black	146,411	0.03	0.18	0.00	0.00
Term (months)	146,411	318.07	71.06	240.00	360.00
Appraisal value (\$1000s)	146,411	295.16	150.86	184.00	375.00
Re finance	146,411	0.60	0.49	0.00	1.00
Interest only	146,411	0.00	0.00	0.00	0.00

Note: These tables present summary statistics for the 2016-2017 sample (Table A.1a), the subsample of loans originated by banks (Table A.1b), the subsample of loans originated by nonbank-non ntechs (Table A.1c), and the subsample of loans originated by ntechs (Table A.1d). Default indicates 90-day delinquency within 2 years of origination (multiplied by 100). Credit scores the representative credit score, i.e. the minimum of each borrower's representative score, which is either the lower score if there are two scores or the middle score if there are three. Loan-to-value(LTV) is the ratio of the loan amount to the lesser of the appraised value and the selling price. Debt-to-income(DTI) is the ratio of all debt payments to household income. Observable risks the estimated probability of default based on credit score, LTV, and DTI as described in Section 2.1. Interest rate is the interest rate at origination. G-fee is the total guarantee fee expressed as an annualized rate. Note that the upfront component of the g-fee is converted to an annualized rate using the respective present value multiplier before being added to the ongoing component of the g-fee. Bank is an indicator for depositories. Nonbank-non ntech is an indicator for lenders that are neither banks nor ntechs. Fintech is an indicator for lenders with a mostly online application process based on the designation of ntechs in Fuster et al. (2019)). Income is the gross income of all borrowers. Single female, single male, and > 1 borrower indicate the number of borrowers and, in the case of 1 borrower, the gender. Hispanic and Black refer to the ethnicity and race of the primary borrower. Term is the number of monthly payments from the origination date until the maturity date of the loan specified as of the origination date. Appraisal value is the appraised value of the collateral for the mortgage. Re finance indicates re finance loans. Interest-only indicates loans with an interest-only portion. Continuous variables are winsorized at 1%. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out re finance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

Table A.2: Summary statistics for the 2018 sample

(a) Full sample

	N	Mean	SD	P25	P75
Default (%)	407,481	0.41	6.41	0.00	0.00
Credit score	407,481	746.26	46.10	715.00	784.00
Loan-to-value (%)	407,481	77.86	17.60	70.00	95.00
Debt-to-income (%)	407,481	35.70	9.57	28.66	43.71
Observable risk (%)	407,481	0.72	0.98	0.17	0.76
Interest rate (%)	407,481	4.67	0.44	4.38	4.88
G-fee (%)	407,481	0.48	0.15	0.40	0.53
Bank	407,481	0.54	0.50	0.00	1.00
Nonbank-non ntech	407,481	0.38	0.48	0.00	1.00
Fintech	407,481	0.08	0.27	0.00	0.00
Income (\$1000s)	407,481	78.37	41.94	48.22	98.77
Single female	407,481	0.21	0.40	0.00	0.00
Single male	407,481	0.30	0.46	0.00	1.00
> 1 borrower	407,481	0.49	0.50	0.00	1.00
Hispanic	407,481	0.09	0.29	0.00	0.00
Black	407,481	0.03	0.18	0.00	0.00
Term (months)	407,481	337.91	55.26	360.00	360.00
Appraisal value (\$1000s)	407,481	303.69	140.31	196.00	390.00
Re nance	407,481	0.24	0.43	0.00	0.00
Interest only	407,481	0.00	0.00	0.00	0.00
Origination revenue (%)	402,561	4.19	2.06	2.83	5.38
Closing costs (%)	402,561	0.88	0.77	0.35	1.20
Secondary marketing income (%)	407,481	3.31	1.93	2.04	4.49

Note: These tables present summary statistics for the 2018 sample (Table A.2a), the subsample of loans originated by banks (Table A.2b), the subsample of loans originated by nonbank-non ntechs (Table A.2c), and the subsample of loans originated by ntechs (Table A.2d). Default indicates 90-day delinquency within 2 years of origination (multiplied by 100). Credit score is the representative credit score, i.e. the minimum of each borrower's representative score, which is either the lower score if there are two scores or the middle score if there are three. Loan-to-value (LTV) is the ratio of the loan amount to the lesser of the appraised value and the selling price. Debt-to-income (DTI) is the ratio of all debt payments to household income. Observable risk is the estimated probability of default based on credit score, LTV, and DTI as described in Section 2.1. Interest rate is the interest rate at origination. G-fee is the total guarantee fee expressed as an annualized rate. Note that the upfront component of the g-fee is converted to an annualized rate using the GSE's respective present value multiplier before being added to the ongoing component of the g-fee. Origination revenue is origination charges plus secondary marketing income (as defined in Section 2.1) divided by the loan amount. Bank is an indicator for depositories. Nonbank-non ntech is an indicator for lenders that are neither banks nor ntechs. Fintech is an indicator for lenders with a mostly online application process based on the designation of ntechs in Fuster et al. (2019)). Income is the gross income of all borrowers. Single female, single male, and > 1 borrower indicate the number of borrowers and, in the case of 1 borrower, the gender. Hispanic and Black refer to the ethnicity and race of the primary borrower. Term is the number of monthly payments from the origination date until the maturity date of the loan specified as of the origination date. Appraisal value is the appraised value of the collateral for the mortgage. Re nance indicates re nance loans. Interest-only indicates loans with an interest-only portion. Origination revenue is a lender's income from originating a loan, expressed as a percentage of the loan amount (see Section 2.1). Closing costs is origination charges as a percentage of the loan amount. Secondary marketing income is the present value of the deviation of a loan's interest rate net of g-fees relative to par. Continuous variables are winsorized at 1%. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out re nance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

(b) Banks

	N	Mean	SD	P25	P75
Default (%)	220,698	0.37	6.09	0.00	0.00
Credit score	220,698	748.93	45.60	720.00	786.00
Loan-to-value (%)	220,698	77.30	17.81	70.00	92.00
Debt-to-income (%)	220,698	35.07	9.63	27.94	43.11
Observable risk (%)	220,698	0.67	0.94	0.16	0.70
Interest rate (%)	220,698	4.61	0.42	4.38	4.88
G-fee (%)	220,698	0.48	0.14	0.40	0.52
Income (\$1000s)	220,698	78.83	42.61	48.03	99.67
Single female	220,698	0.20	0.40	0.00	0.00
Single male	220,698	0.29	0.46	0.00	1.00
> 1 borrower	220,698	0.50	0.50	0.00	1.00
Hispanic	220,698	0.09	0.28	0.00	0.00
Black	220,698	0.03	0.17	0.00	0.00
Term (months)	220,698	336.49	56.82	360.00	360.00
Appraisal value (\$1000s)	220,698	298.55	140.73	190.00	385.00
Re finance	220,698	0.23	0.42	0.00	0.00
Interest only	220,698	0.00	0.00	0.00	0.00
Origination revenue (%)	218,264	3.83	1.94	2.54	4.97
Closing costs (%)	218,264	0.79	0.66	0.35	1.06
Secondary marketing income (%)	220,698	3.05	1.83	1.84	4.17

Note: These tables present summary statistics for the 2018 sample (Table A.2a), the subsample of loans originated by banks (Table A.2b), the subsample of loans originated by nonbank-non ntechs (Table A.2c), and the subsample of loans originated by ntechs (Table A.2d). Default indicates 90-day delinquency within 2 years of origination (multiplied by 100). Credit score is the representative credit score, i.e. the minimum of each borrower's representative score, which is either the lower score if there are two scores or the middle score if there are three. Loan-to-value(LTV) is the ratio of the loan amount to the lesser of the appraised value and the selling price. Debt-to-income(DTI) is the ratio of all debt payments to household income. Observable risk is the estimated probability of default based on credit score, LTV, and DTI as described in Section 2.1. Interest rate is the interest rate at origination. G-fees is the total guarantee fee expressed as an annualized rate. Note that the upfront component of the g-fee is converted to an annualized rate using the GSE's respective present value multiplier before being added to the ongoing component of the g-fee. Origination revenue is origination charges plus secondary marketing income (as defined in Section 2.1) divided by the loan amount. Bank is an indicator for depositories. Nonbank-non ntech is an indicator for lenders that are neither banks nor ntechs. Fintech is an indicator for lenders with a mostly online application process based on the designation of ntechs in Fuster et al. (2019)). Income is the gross income of all borrowers. Single female, single male, and > 1 borrower indicate the number of borrowers and, in the case of 1 borrower, the gender. Hispanic and Black refer to the ethnicity and race of the primary borrower. Term is the number of monthly payments from the origination date until the maturity date of the loan specified as of the origination date. Appraisal value is the appraised value of the collateral for the mortgage. Re finance indicates re finance loans. Interest-only indicates loans with an interest-only portion. Origination revenue is a lender's income from originating a loan, expressed as a percentage of the loan amount (see Section 2.1). Closing costs is origination charges as a percentage of the loan amount. Secondary marketing income is the present value of the deviation of a loan's interest rate net of g-fees relative to par. Continuous variables are winsorized at 1%. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

Table A.2: Summary statistics (continued.)

(c) Nonbank-non ntechs

	N	Mean	SD	P25	P75
Default (%)	154,043	0.44	6.64	0.00	0.00
Credit score	154,043	744.11	46.11	713.00	781.00
Loan-to-value (%)	154,043	78.56	17.27	71.00	95.00
Debt-to-income (%)	154,043	36.45	9.44	29.59	44.34
Observable risk (%)	154,043	0.77	1.01	0.18	0.91
Interest rate (%)	154,043	4.74	0.44	4.50	5.00
G-fee (%)	154,043	0.50	0.15	0.42	0.54
Income (\$1000s)	154,043	78.48	41.05	49.27	98.43
Single female	154,043	0.21	0.41	0.00	0.00
Single male	154,043	0.31	0.46	0.00	1.00
> 1 borrower	154,043	0.48	0.50	0.00	1.00
Hispanic	154,043	0.11	0.31	0.00	0.00
Black	154,043	0.03	0.18	0.00	0.00
Term (months)	154,043	341.03	52.00	360.00	360.00
Appraisal value (\$1000s)	154,043	315.48	139.07	210.00	400.00
Re nance	154,043	0.23	0.42	0.00	0.00
Interest only	154,043	0.00	0.00	0.00	0.00
Origination revenue (%)	151,558	4.53	2.05	3.18	5.74
Closing costs (%)	151,558	0.90	0.79	0.36	1.24
Secondary marketing income (%)	154,043	3.64	1.94	2.38	4.84

Note: These tables present summary statistics for the 2018 sample (Table A.2a), the subsample of loans originated by banks (Table A.2b), the subsample of loans originated by nonbank-non ntechs (Table A.2c), and the subsample of loans originated by ntechs (Table A.2d). Default indicates 90-day delinquency within 2 years of origination (multiplied by 100). Credit score is the representative credit score, i.e. the minimum of each borrower's representative score, which is either the lower score if there are two scores or the middle score if there are three. Loan-to-value (LTV) is the ratio of the loan amount to the lesser of the appraised value and the selling price. Debt-to-income (DTI) is the ratio of all debt payments to household income. Observable risk is the estimated probability of default based on credit score, LTV, and DTI as described in Section 2.1. Interest rate is the interest rate at origination. G-fee is the total guarantee fee expressed as an annualized rate. Note that the upfront component of the g-fee is converted to an annualized rate using the GSE's respective present value multiplier before being added to the ongoing component of the g-fee. Origination revenue is origination charges plus secondary marketing income (as defined in Section 2.1) divided by the loan amount. Bank is an indicator for depositories. Nonbank-non ntech is an indicator for lenders that are neither banks nor ntechs. Fintech is an indicator for lenders with a mostly online application process based on the designation of ntechs in Fuster et al. (2019)). Income is the gross income of all borrowers. Single female, single male, and > 1 borrower indicate the number of borrowers and, in the case of 1 borrower, the gender. Hispanic and Black refer to the ethnicity and race of the primary borrower. Term is the number of monthly payments from the origination date until the maturity date of the loan specified as of the origination date. Appraisal value is the appraised value of the collateral for the mortgage. Re nance indicates re nance loans. Interest-only indicates loans with an interest-only portion. Origination revenue is a lender's income from originating a loan, expressed as a percentage of the loan amount (see Section 2.1). Closing costs is origination charges as a percentage of the loan amount. Secondary marketing income is the present value of the deviation of a loan's interest rate net of g-fees relative to par. Continuous variables are winsorized at 1%. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out re nance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

(d) Fintechs

	N	Mean	SD	P25	P75
Default (%)	32,740	0.54	7.33	0.00	0.00
Credit score	32,740	738.32	47.91	703.00	778.00
Loan-to-value (%)	32,740	78.35	17.56	69.00	95.00
Debt-to-income (%)	32,740	36.39	9.45	29.41	44.36
Observable risk (%)	32,740	0.88	1.04	0.20	1.18
Interest rate (%)	32,740	4.67	0.44	4.38	4.88
G-fee (%)	32,740	0.45	0.16	0.36	0.51
Income (\$1000s)	32,740	74.81	41.32	44.79	94.85
Single female	32,740	0.21	0.41	0.00	0.00
Single male	32,740	0.33	0.47	0.00	1.00
> 1 borrower	32,740	0.46	0.50	0.00	1.00
Hispanic	32,740	0.07	0.26	0.00	0.00
Black	32,740	0.05	0.21	0.00	0.00
Term (months)	32,740	332.74	58.62	360.00	360.00
Appraisal value (\$1000s)	32,740	282.76	138.69	176.00	365.00
Re finance	32,740	0.40	0.49	0.00	1.00
Interest only	32,740	0.00	0.00	0.00	0.00
Origination revenue (%)	32,739	4.98	2.29	3.37	6.35
Closing costs (%)	32,739	1.37	1.10	0.42	2.20
Secondary marketing income (%)	32,740	3.60	2.24	2.02	4.96

Note: These tables present summary statistics for the 2018 sample (Table A.2a), the subsample of loans originated by banks (Table A.2b), the subsample of loans originated by nonbank-non ntechs (Table A.2c), and the subsample of loans originated by ntechs (Table A.2d). Default indicates 90-day delinquency within 2 years of origination (multiplied by 100). Credit score is the representative credit score, i.e. the minimum of each borrower's representative score, which is either the lower score if there are two scores or the middle score if there are three. Loan-to-value (LTV) is the ratio of the loan amount to the lesser of the appraised value and the selling price. Debt-to-income (DTI) is the ratio of all debt payments to household income. Observable risk is the estimated probability of default based on credit score, LTV, and DTI as described in Section 2.1. Interest rate is the interest rate at origination. G-fees is the total guarantee fee expressed as an annualized rate. Note that the upfront component of the g-fee is converted to an annualized rate using the GSE's respective present value multiplier before being added to the ongoing component of the g-fee. Origination revenue is origination charges plus secondary marketing income (as defined in Section 2.1) divided by the loan amount. Bank is an indicator for depositories. Nonbank-non ntechs is an indicator for lenders that are neither banks nor ntechs. Fintech is an indicator for lenders with a mostly online application process based on the designation of ntechs in Fuster et al. (2019)). Income is the gross income of all borrowers. Single female, single male, and > 1 borrower indicate the number of borrowers and, in the case of 1 borrower, the gender. Hispanic and Black refer to the ethnicity and race of the primary borrower. Term is the number of monthly payments from the origination date until the maturity date of the loan specified as of the origination date. Appraisal value is the appraised value of the collateral for the mortgage. Re finance indicates re finance loans. Interest-only indicates loans with an interest-only portion. Origination revenue is a lender's income from originating a loan, expressed as a percentage of the loan amount (see Section 2.1). Closing costs is origination charges as a percentage of the loan amount. Secondary marketing income is the present value of the deviation of a loan's interest rate net of g-fees relative to par. Continuous variables are winsorized at 1%. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

A.2 Supplemental material for Section 2.3.1

Figure A.1: Origination revenue components and observable risk

This figure presents a binned scatterplot of origination revenue components (closing costs and secondary marketing income as a percentage of the loan amount) on observable risk while controlling for year-month fixed effects. Observable risk is the estimated probability of default based on credit score, the loan-to-value ratio, and the debt-to-income ratio as described in Section 2.1. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2018, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront origination fee deviates from the first table of the origination fee matrix by more than 25 basis points.

Table A.3: Origination revenue and observable risk

	(1)	(2)	(3)	(4)	(5)	(6)
	Orig. rev.	Orig. rev.	Closing costs	Closing costs	Second. income	Second. income
Observable risk	0.597*** (165.62)		0.024*** (16.08)		0.573*** (165.51)	
Credit score		-1.093*** (-153.76)		-0.168*** (-57.00)		-0.926*** (-138.18)
LTV		1.374*** (67.21)		-1.131*** (-138.43)		2.516*** (133.31)
DTI		0.974*** (28.90)		-0.014 (-1.03)		0.992*** (31.42)
Observations	386,219	386,219	386,219	386,219	391,173	391,173
R ²	0.344	0.348	0.207	0.264	0.310	0.336
ZIP Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: Column (1) regresses origination revenue on observable risk while controlling for ZIP code by year-quarter fixed effects. Observable risk is the estimated probability of default based on credit score, the loan-to-value ratio, and the debt-to-income ratio as described in Section 2.1. Column (2) regresses origination revenue on credit score, the loan-to-value (LTV) ratio, and the debt-to-income (DTI) ratio (each divided by 100). Column (3) and column (4) are similar to column (1) and column (2) except that the dependent variable is the closing costs portion of origination revenue. Column (5) and column (6) are similar to column (1) and column (2) except that the dependent variable is secondary marketing income portion of origination revenue. T-statistics computed using robust standard errors are reported in parentheses. * indicates statistical significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2018, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

A.3 Supplemental material for Section 2.3.2

Figure A.2: Origination revenue and default

This figure presents a binned scatterplot of default (multiplied by 100) on origination revenue while controlling for year-month fixed effects and observable risk. Observable risk is the estimated probability of default based on credit score, the loan-to-value ratio, and the debt-to-income ratio as described in Section 2.1. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

Table A.4: Origination revenue and default

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline	Obs. risk	Controls	Safe	Risky	Interact
Orig. rev.	0.150*** (20.78)	0.062*** (8.90)	0.063*** (7.56)	0.022*** (4.43)	0.103*** (7.51)	0.022*** (4.43)
Observable risk		0.557*** (22.93)		0.190** (2.20)	0.575*** (17.64)	0.190** (2.19)
Orig. rev. Risky						0.081*** (5.57)
Observable risk Risky						0.385*** (4.17)
Observations	386,219	386,219	386,200	191,958	176,849	368,807
R ²	0.138	0.144	0.153	0.195	0.197	0.199
ZIP Year-quarter FE	Yes	Yes	Yes	Yes	Yes	No
ZIP Year-quarter Risky FE	No	No	No	No	No	Yes
Controls	No	No	Yes	No	No	No

Note: Column (1) regresses an indicator for default (multiplied by 100) on the origination revenue. Column (2) adds observable risk as a regressor. Column (3) instead includes the following controls: the interaction between 10-point credit score bins (starting at 620, with an additional indicator for all credit scores below 620), 5% loan-to-value bins (starting at 60%, with an additional indicator for all loan-to-value ratios below 60%), and debt-to-income decile indicators (note that this absorbs observable risk); income decile indicators; family type indicators (i.e. single female, single male, or more than 1 borrower); indicators for black and hispanic borrowers; term indicators; appraisal value decile indicators; an indicator for a loan having an interest-only period; an indicator for a refinance loan; loan amount decile indicators; an indicator for self-employed borrowers; an indicator for first-time homebuyers; an indicator for full income documentation; and an indicator for full asset documentation. T-statistics computed using robust standard errors are reported in parentheses. * indicates statistical significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

A.4 Supplemental material for Section 2.3.5

Table A.5: Interest rates, default, and prepayment

	(1)	(2)	(3)
	Baseline	Obs. risk	Controls
Default	0.156*** (54.61)	0.057*** (19.66)	0.043*** (17.12)
Prepayment	0.062*** (81.68)	0.053*** (71.88)	0.052*** (87.75)
Obs. risk		0.093*** (335.06)	
Observations	2,109,041	2,109,041	2,109,029
R ²	0.467	0.497	0.692
ZIP Year-quarter FE	Yes	Yes	Yes
Controls	No	No	Yes

Note: Column (1) regresses the interest rate net of the total g-fee on an indicator for default and an indicator for prepayment. Column (2) adds observable risk as a regressor. Column (3) instead includes the following controls: the interaction between 10-point credit score bins (starting at 620, with an additional indicator for all credit scores below 620), 5% loan-to-value bins (starting at 60%, with an additional indicator for all loan-to-value ratios below 60%), and debt-to-income decile indicators (note that this absorbs observable risk); income decile indicators; family type indicators (i.e. single female, single male, or more than 1 borrower); indicators for black and hispanic borrowers; term indicators; appraisal value decile indicators; an indicator for a loan having an interest-only period; an indicator for a refinance loan; loan amount decile indicators; an indicator for self-employed borrowers; an indicator for first-time homebuyers; an indicator for full income documentation; and an indicator for full asset documentation. T-statistics computed using robust standard errors are reported in parentheses. * indicates statistical significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

Table A.6: Origination revenue, default, and prepayment

	(1)	(2)	(3)
	Baseline	Obs. risk	Controls
Default	1.311*** (24.15)	0.503*** (9.30)	0.372*** (8.20)
Prepayment	0.108*** (13.79)	0.128*** (17.13)	0.271*** (42.89)
Observable risk		0.594*** (164.45)	
Observations	386,219	386,219	386,200
R ²	0.277	0.344	0.548
ZIP	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes
Controls	No	No	Yes

Note: Column (1) regresses the interest rate net of the total g-fee on an indicator for default and an indicator for prepayment. Column (2) adds observable risk as a regressor. Column (3) instead includes the following controls: the interaction between 10-point credit score bins (starting at 620, with an additional indicator for all credit scores below 620), 5% loan-to-value bins (starting at 60%, with an additional indicator for all loan-to-value ratios below 60%), and debt-to-income decile indicators (note that this absorbs observable risk); income decile indicators; family type indicators (i.e. single female, single male, or more than 1 borrower); indicators for black and hispanic borrowers; term indicators; appraisal value decile indicators; an indicator for a loan having an interest-only period; an indicator for a refinance loan; loan amount decile indicators; an indicator for self-employed borrowers; an indicator for first-time homebuyers; an indicator for full income documentation; and an indicator for full asset documentation. T-statistics computed using robust standard errors are reported in parentheses. * indicates statistical significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2018, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

B Calculations and proof details

B.1 Proof of Proposition 1

This section shows using an argument like the one in [Milgrom \(1981\)](#) that there is an equilibrium in which each lender's offered interest rate is $\underline{R}(D(s, s; y, n))$, where $\underline{R}(D)$ introduced in equation (8) is the zero-profits interest rate corresponding to the lender's probability of default D , $D(s, t; y, n)$ introduced in equation (5) is the posterior probability of default conditional on the signals of the supplying lender and the next most competitive lender, and in particular $D(s, s; y, n)$ introduced in equation (6) is the minimum posterior probability of default that a lender could have conditional on winning the auction and updating its risk estimate based on the equilibrium outcome.⁴⁵ By substituting in equation (6), note that $\underline{R}(D(s, s; y, n)) = w l_d + \frac{1}{2}(n - (n + 2)s)y + r$. Without loss of generality, we show that incentive compatibility holds for lender $i = 1$.

Note that a lender's interest rate offer only affects its expected profits insofar as it determines when it wins the auction. That is, conditional on winning the auction, a lender's own interest rate offer has no effect on its expected profits, and similarly in the case where the lender does not win the auction. Therefore, it suffices to check that if a lender wins an auction then it achieves positive expected profits (and therefore cannot profitably deviate by bidding a higher interest rate in order to lose), and if it loses the auction then it cannot profitably deviate by bidding a lower interest rate in order to win.

First, suppose lender 1 wins the auction. Suppose without loss of generality that the equilibrium interest rate is given by lender 2's offered interest rate, or $R_{eq} = w[l_d + \frac{1}{2}(n - (n + 2)s_2)y] + r$. Lender 1 can therefore infer s_2 and update its zero-profits interest rate after learning the information contained within the equilibrium interest rate: $\underline{R}(D(s_1, s_2; y, n)) = w l_d + \frac{1}{2}(n - 2s_1 - ns_2) + r$. Since lender 2's offered interest rate is higher, one can infer from equation (9) that $s_2 < s_1$. Therefore, lender 1's updated zero profits interest rate is less than R_{eq} , so lender 1's offer still achieves positive expected profits. Hence, lender 1 has no profitable deviation.

⁴⁵If lender i wins, then the observation of the equilibrium interest rate will allow it to effectively observe the next most competitive lender, which will lead to an increase in lender i 's estimated posterior risk of default since lender i wins only if it has a lower posterior probability of default conditional on its own signal. Hence, the minimum posterior probability of default that lender i can have conditional on winning and observing the equilibrium interest rate occurs when the next most competitive lender has the same posterior probability of default.

Now, suppose that lender 1 loses the auction. If lender 1 hypothetically knew lender 2's offer, it could infer s_2 and thereby update its zero-profits interest rate after learning the information contained within the equilibrium interest rate: $R(D(s_1, s_2; y, n)) = w l_d + \frac{1}{2}(n - 2s_1 - ns_2) + r$. Since lender 2's offer is lower, one can infer from equation (9) that $s_2 > s_1$. Therefore, lender 1's updated zero profits interest rate is greater than lender 2's offer, so lender 1 has no incentive to deviate by undercutting lender 2. Since this argument holds for all potential values of lender 2's offer, lender 1 can conclude that there is no profitable deviation even if it doesn't observe the equilibrium interest rate.

B.2 Calculation for equation (5)

This section shows

$$D(s, t; y, n) = l_d + \frac{1}{2}(n - 2s - nt)y$$

First, using general results about order statistics, note that the joint distribution of $s_{n:n}$ and $s_{n-1:n}$ for a borrower of type q is given by

$$f(s_{n:n} = s, s_{n-1:n} = tjq) = n(n-1)F(tjq)^{n-2}f(tjq)f(sjq) \quad (30)$$

Then, observe that the predictive distribution for the joint distribution for the two highest signals is given by

$$\begin{aligned} f(s_{n:n} = s, s_{n-1:n} = t) &= l_d f(s_{n:n} = s, s_{n-1:n} = tj_d) + l_r f(s_{n:n} = s, s_{n-1:n} = tj_r) \\ &= l_d n(n-1)F(tj_d)^{n-2}f(tj_d)f(sj_d) + l_r n(n-1)F(tj_r)^{n-2}f(tj_r)f(sj_r) \\ &= l_d n(n-1) \left[t + \frac{1}{2}t^2 \frac{y}{l_d} \right]^{n-2} \left[1 + \frac{1}{2}t \frac{y}{l_d} \right] \left[1 + \frac{1}{2}s \frac{y}{l_d} \right] \\ &\quad + l_r n(n-1) \left[t + \frac{1}{2}t^2 \frac{y}{l_r} \right]^{n-2} \left[1 + t \frac{1}{2} \frac{y}{l_r} \right] \left[1 + s \frac{1}{2} \frac{y}{l_r} \right] \\ &= l_d n(n-1) \left[t^{n-2} + (n-2)t^{n-3} \frac{1}{2}t^2 \frac{y}{l_d} \right] \left[1 + \frac{1}{2}t \frac{y}{l_d} \right] \left[1 + \frac{1}{2}s \frac{y}{l_d} \right] \\ &\quad + l_r n(n-1) \left[t^{n-2} + (n-2)t^{n-3} \frac{1}{2}t^2 \frac{y}{l_r} \right] \left[1 + t \frac{1}{2} \frac{y}{l_r} \right] \left[1 + s \frac{1}{2} \frac{y}{l_r} \right] \\ &= \frac{n(n-1)t^{n-2}}{y} \left\{ \frac{l_d^{n-2} \left[1 + \frac{1}{2}t \frac{y}{l_d} \right] \left[1 + \frac{1}{2}s \frac{y}{l_d} \right] + l_r^{n-2} \left[1 + t \frac{1}{2} \frac{y}{l_r} \right] \left[1 + s \frac{1}{2} \frac{y}{l_r} \right]}{l_d^{n-2} + l_r^{n-2}} \right\} \end{aligned} \quad (31)$$

Then, by Bayesian inference we have

$$\begin{aligned}
 D(s, t; y, n) &= \Pr(d | s_{n:n} = s, s_{n-1:n} = t) \\
 &= \int_d \frac{f(s_{n:n} = s, s_{n-1:n} = t | d)}{f(s_{n:n} = s, s_{n-1:n} = t)} \\
 &= \int_d \frac{n(n-1)F(t|d)^{n-2}f(t|d)f(s|d)}{n(n-1)t^{n-2}} \\
 &= \int_d f(s|d)f(t|d; y) \frac{F(t|d)^{n-2}}{t} \\
 &= \int_d \left(1 + \frac{1}{2} \frac{s}{l_d} \right)^{n-2} \left(1 + \frac{1}{2} \frac{t}{l_d} \right)^{n-2} \left(1 + \frac{1}{2} (1-t) \frac{y}{l_d} \right)^{n-2} \\
 &\quad \int_d \left(1 + \frac{1}{2} \frac{s}{l_d} \right)^{n-2} \left(1 + \frac{1}{2} \frac{t}{l_d} \right)^{n-2} \left(1 + (n-2) \frac{1}{2} (1-t) \frac{y}{l_d} \right) \\
 &\quad \int_d \left(1 + \frac{1}{2} (n-2s-nt) \frac{y}{l_d} \right)
 \end{aligned} \tag{32}$$

B.3 Proof of Proposition 2

This section shows

$$\underline{s} = \frac{n}{n+1} + \frac{2(wl_d + r - A)}{(n+1)wy}$$

The threshold is defined by the point where a lender's expected profits is equal to zero.

To compute this, consider that if the supplying lender's signal is equal to \underline{s} , then, by symmetry, no other lender offers a loan.⁴⁶ Therefore, the supplying lender charges an interest rate A and has an expected zero-profits interest rate of $E_{s_{n-1:n}}[R(D(\underline{s}, s_{n-1:n}; y, n)) | s_{n:n} = \underline{s}]$.

To compute the latter, recall that the predictive distribution of the signal is uniform. Therefore the conditional pdf for the greatest signal among the $n-1$ competing draws

⁴⁶We ignore the zero-probability event where multiple lenders have the same signal.

is given by $f(s_{n-1:n} = t | s_{n:n} = s) = (n-1) \frac{t^{n-2}}{s^{n-1}}$. Therefore

$$\begin{aligned} E_{s_{n-1:n}}[s_{n-1:n} | s_{n:n} = s] &= \int_0^s t (n-1) \frac{t^{n-2}}{s^{n-1}} dt \\ &= \frac{n-1}{n} s \end{aligned} \quad (33)$$

Therefore we have

$$\begin{aligned} E_{s_{n-1:n}}[R(D(s, s_{n-1:n}; y, n)) | s_{n:n} = s] &= w l_d + \frac{1}{2} (n-2) s - n E_{s_{n-1:n}}[s_{n-1:n} | s_{n:n} = s] + r \\ &= w l_d + \frac{1}{2} (n - (n+1)s) y + r \end{aligned} \quad (34)$$

Finally, as mentioned above, \underline{s} is the point where a lender has zero expected profits, which is determined by the condition:

$$\begin{aligned} 0 &= A E_{s_{n-1:n}}[R(D(\underline{s}, s_{n-1:n}; y, n)) | s_{n:n} = \underline{s}] \\ &= A \left(w l_d + \frac{1}{2} (n - (n+1)\underline{s}) y + r \right) \end{aligned} \quad (35)$$

B.4 Calculation for equation (14)

This section computes a lender's profits.

First, recall that the predictive distribution of the signal is uniform. Therefore the conditional pdf for the greatest signal among the $n-1$ competing draws is given by $f(s_{n-1:n} = t | s_{n:n} = s) = (n-1) \frac{t^{n-2}}{s^{n-1}}$. Therefore, modulo the cost of information acquisition, a lender's expected profit conditional on winning with signal s can be written

as

$$\begin{aligned}
 p_L(s) &= E_{s_{n-1:n}} \left(\int_0^h (A - wI_d - r) \underline{R}(D(s, s_{n-1:n}; y, n)) \mathbb{1}_{f_{s=s_{n:n}, s_{n-1:n}} \leq s} \right) \\
 &+ E_{s_{n-1:n}} \left(\underline{R}(D(s_{n-1:n}, s_{n-1:n}; y, n)) - \underline{R}(D(s, s_{n-1:n}; y, n)) \right) \mathbb{1}_{f_{s=s_{n:n}, s_{n-1:n}} \leq s} \\
 &= \int_0^s (A - wI_d - r)(n-1) \frac{t^{n-2}}{s^{n-1}} dt \\
 &\quad + wy \frac{1}{2} \int_0^s (n-2s-nt)(n-1) \frac{t^{n-2}}{s^{n-1}} dt \\
 &\quad + \int_s^z w(s-t)y(n-1) \frac{t^{n-2}}{s^{n-1}} dt \\
 &= (A - wI_d - r) s^{n-1} \frac{1}{s^{n-1}} \\
 &\quad + wy \frac{1}{2} (n-1) s^n + 2s s^{n-1} - n s^{n-1} \frac{1}{s^{n-1}} \\
 &\quad + wy \frac{1}{n} s^n - s s^{n-1} + \frac{n-1}{n} s^n \frac{1}{s^{n-1}}
 \end{aligned} \tag{36}$$

Then, the pdf for the maximum signal is given by $f(s_{n:n}) = ns^{n-1}$. Integrating over potential values of the maximal signal $s_{n:n} \in [s, 1]$ and dividing by n to obtain the profits for a single lender, modulo the cost of information acquisition, results in

$$\begin{aligned}
 p_L &= \frac{1}{n} \int_s^1 p_L(s) ns^{n-1} ds \\
 &= (A - wI_d - r) s^{n-1} (1 - s) \\
 &\quad + wy \frac{1}{2} \left((n+1) s^{n-1} + (2n-1) s^n - n s^{n+1} \right) \\
 &\quad + wy \frac{1}{n(n+1)} \left(\frac{1}{2} s^{n-1} + \frac{n-1}{n} s^n + \frac{n+1}{2(n+1)} s^{n+1} \right)
 \end{aligned} \tag{37}$$

Finally, a lender's total profits is obtained by subtracting out the information acquisition cost $m \frac{(y-z)^2}{2}$.

B.5 Calculation for equation (18)

This section shows

$$D(s_1, s_2; y_1, y_2) = \int_0^1 \int_0^1 \frac{1}{2} s_1 y_1 + \frac{1}{2} s_2 y_2$$

First, since the signals are independently distributed, observe that the pdf of the predictive distribution of the two signals can be written as

$$\begin{aligned} f(s_1, s_2; y_1, y_2) &= \int_d f(s_1|d; y_1) f(s_2|d; y_2) + \int_r f(s_2|r; y_1) f(s_2|r; y_2) \\ &= \int_d \left(1 + \frac{1}{2} s_1 \frac{y_1}{l_d}\right) \left(1 + \frac{1}{2} s_2 \frac{y_2}{l_d}\right) \\ &\quad + \int_r \left(1 + s_1 \frac{1}{2} \frac{y_1}{l_r}\right) \left(1 + s_2 \frac{1}{2} \frac{y_2}{l_r}\right) \\ &\quad \int_0^1 \int_0^1 \frac{1}{2} s_1 \frac{y_1}{l_d} + \frac{1}{2} s_2 \frac{y_2}{l_d} \\ &\quad + \int_r \left(1 + s_1 \frac{1}{2} \frac{y_1}{l_r}\right) + s_2 \frac{1}{2} \frac{y_2}{l_r} \\ &= 1 \end{aligned} \tag{38}$$

Then, by Bayesian inference and independence we have

$$\begin{aligned} D(s_1, s_2; y_1, y_2) &= \Pr(d|s_1, s_2; y_1, y_2) \\ &= \int_d \frac{f(s_1, s_2|d; y_1, y_2)}{f(s_1, s_2; y_1, y_2)} \\ &= \int_d \frac{f(s_1|d; y_1) f(s_2|d; y_2)}{f(s_1, s_2; y_1, y_2)} \\ &= \int_d \left(1 + \frac{1}{2} s_1 \frac{y_1}{l_d}\right) \left(1 + \frac{1}{2} s_2 \frac{y_2}{l_d}\right) \\ &\quad \int_0^1 \int_0^1 \frac{1}{2} s_1 y_1 + \frac{1}{2} s_2 y_2 \end{aligned} \tag{39}$$

B.6 Proof of Proposition 3

This section shows using an argument like the one in Milgrom (1981) that there is an equilibrium in which each lender's offer is $\underline{R}(D(s_1, s_2; y_1, y_2))$, where $\underline{R}(D)$ introduced in equation (8) is the zero-profits interest rate corresponding to the lender's probability of default D , $D(s_1, s_2; y_1, y_2)$ introduced in equation (18) is the posterior probability of default conditional on the signals of both lenders, and in particular $\underline{D}(s_i, s_j; y_i, y_j)$ introduced in equation (6) is the minimum posterior probability of default that a lender could have conditional on winning the auction and updating its risk estimate based on the equilibrium outcome.⁴⁷ By substituting in equation (6), note that $\underline{R}(D(s_i, s_j; y_i, y_j)) = w[l_d + (1 - 2s_i)y_i] + r$. Without loss of generality, we show that incentive compatibility holds for $i = 1$.

Note that a lender's interest rate offer only affects its expected profits insofar as it determines when it wins the auction. That is, conditional on winning the auction, a lender's own interest rate offer has no effect on its expected profits, and similarly in the case where the lender does not win the auction. Therefore, it suffices to check that if a lender wins an auction then it achieves positive expected profits (and therefore cannot profitably deviate by bidding a higher interest rate in order to lose), and if it loses the auction then it cannot profitably deviate by bidding a lower interest rate in order to win.

First, suppose lender 1 wins the auction. Note that the equilibrium interest rate must then be given by lender 2's offer, or $R_{eq} = w[l_d + (1 - 2s_2)y_2] + r$. Lender 1 can therefore infer $\frac{1}{2} s_2 y_2 = \frac{R_{eq} - r}{w} l_d$ and update its zero-profits interest rate after learning the information contained within the equilibrium interest rate: $\underline{R}(D(s_1, s_2; y_1, y_2)) = w[l_d + \frac{1}{2} s_1 y_1 + \frac{1}{2} s_2 y_2] + r$. Since lender 2's offer is higher, one can infer from equation (20) that $(1 - 2s_2)y_2 > (1 - 2s_1)y_1$. Therefore, lender 1's updated zero-profits interest rate is less than R_{eq} , so lender 1's offer still achieves positive expected profits. Hence, lender 1 has no profitable deviation.

Now, suppose that lender 1 loses the auction. If lender 1 hypothetically knew lender

⁴⁷If lender i wins, then the observation of the equilibrium interest rate will allow it to effectively observe the signal of lender i , which will lead to an increase in lender i 's estimated posterior risk of default since lender i wins only if it has a lower posterior probability of default conditional on its own signal. Hence, the minimum posterior probability of default that lender i can have conditional on winning and observing the equilibrium interest rate occurs when lender i has the same posterior probability of default.

2's offer, it could infer $\frac{1}{2} s_2 y_2 = \frac{\text{Req } r}{2} l_d$ and thereby update its zero-profits interest rate after learning the information contained within the equilibrium interest rate: $R(D(s_1, s_2; y_1, y_2)) = w l_d + \frac{1}{2} s_1 y_1 + \frac{1}{2} s_2 y_2 + r$. Since lender 2's offer is lower, one can infer from equation (20) that $(1 - 2s_2)y_2 < (1 - 2s_1)y_1$. Therefore, lender 1's updated zero-profits interest rate is greater than lender 2's offer, so lender 1 has no incentive to deviate by undercutting lender 2. Since this argument holds for all potential values of lender 2's offer, lender 1 can conclude that there is no profitable deviation even if it doesn't observe the equilibrium interest rate.

B.7 Proof of Proposition 4

This section shows

$$\underline{s}_i = \frac{2}{3} \frac{y_1 + y_2}{2y_i} + \frac{w l_d + r - A}{w y_i} \quad (40)$$

The threshold is defined by the point where a lender's expected profits is equal to zero under the assumption that the other lender does not compete. First, consider the case of lender 1.

To compute \underline{s}_1 , consider that, given that lender 2 will not compete, the lender 1 charges an interest rate A and has expected zero-profits interest rate of $E_{s_2}[R(D(s_1, s_2; y_1, y_2)) | s_2 < \underline{s}_2]$.

To compute the latter, recall that the predictive distribution is uniform. Therefore the conditional pdf for s_2 is given by $f(s_2 | s_2 < \underline{s}_2) = \frac{1}{\underline{s}_2}$. Therefore

$$\begin{aligned} E[s_2 | s_2 < \underline{s}_2] &= \int_0^{\underline{s}_2} \frac{s_2}{\underline{s}_2} ds_2 \\ &= \frac{1}{2} \underline{s}_2 \end{aligned} \quad (41)$$

Therefore we have

$$\begin{aligned} E_{s_2}[R(D(s_1, s_2; y_1, y_2)) | s_2 < \underline{s}_2] &= w l_d + \frac{1}{2} \underline{s}_1 y_1 + \frac{1}{2} E_{s_2}[s_2 | s_2 < \underline{s}_2] + r \\ &= w l_d + \frac{1}{2} \underline{s}_1 y_1 + \frac{1}{2} \frac{1}{2} \underline{s}_2 + r \end{aligned} \quad (42)$$

Finally, as mentioned above, \underline{s}_1 is the point where lender 1 has zero expected profits, which is determined by the condition:

$$\begin{aligned} 0 &= A E_{s_2}[R(D(s_1, s_2; y_1, y_2)) | s_2 < \underline{s}_2] \\ &= A \left(w l_d + \frac{1}{2} \underline{s}_1 y_1 + \frac{1}{2} \frac{1}{2} \underline{s}_2 + r \right) \end{aligned} \quad (43)$$

An analogous equation also holds for lender 2. Then the system of equations implies the result.

Note that if $\underline{s}_2 = 1$, then lender 2 never makes an offer, and therefore lender 1 cannot make any inferences about s_2 based on the observation that lender 2 does not make an offer. In that case, \underline{s}_1 is instead defined by

$$\begin{aligned} 0 &= A \left(w l_d + \frac{1}{2} \underline{s}_1 y_1 + r \right) \\ \Rightarrow \underline{s}_1 &= \frac{1}{2} + \frac{w l_d + r}{w y_1} \end{aligned} \quad (44)$$

An analogous argument determines \underline{s}_2 when $\underline{s}_1 = 1$.

B.8 Proof of Proposition 5

This section shows that, if we consider the space of linear bid functions of the form $B_i(s_i) = a_i + b_i s_i$, and, conditional on drawing signal s_i , lender i chooses a_i and b_i in order to maximize the expected profits

$$E_{s_i} \left(B_i(s_i) - R_i(D(s_1, s_2; y)) \right) 1_{a_i + b_i s_i < B_i(s_i)} \quad (45)$$

then there is an equilibrium in which the bid functions are given by

$$B_i(s_i) = w_i l_d + w_i \frac{1}{2} s_i y + r_i \quad (46)$$

Without loss of generality, consider the decision problem of lender 1 conditional on lender 2 playing the corresponding equilibrium bid function. That is, lender 1 chooses a_1 and b_1 while the coefficients for lender 2's bid function are $a_2 = w_2 [l_d + \frac{1}{2}y] + r_2$ and $b_2 = w_2 y$.

Given s_1 , denote the threshold value of s_2 at which lender 1 wins the auction by $\tilde{s}_2(s_1)$, i.e. $B_1(s_1) = B_2(\tilde{s}_2(s_1))$. Note that if $\tilde{s}_2(s_1) < \underline{s}_2$ then marginal changes in lender 1's bid function have no effect on its expected profits since lender 1 always wins and pays the interest rate determined by lender 2's bid. Similarly, if $\tilde{s}_2(s_1) > \underline{s}_2$ (note that the participation threshold \underline{s}_2 is defined in equation (28)) then marginal changes in lender 1's bid function have no effect on its expected profits since it always loses regardless of lender 2's signal. Therefore, consider the case where $\tilde{s}_2(s_1) \in (\underline{s}_2, 1)$. In that case, lender 1's problem is to find a_1 and b_1 to maximize

$$\int_0^{\tilde{s}_2(s_1)} [A - [w_1 [l_d + (1 - s_1 - s_2)y] + r_1]] ds_2 \quad (47)$$

$$+ \int_{\underline{s}_2}^{\tilde{s}_2(s_1)} [(a_2 + b_2 s_2) - [w_1 [l_d + (1 - s_1 - s_2)y] + r_1]] ds_2 \quad (48)$$

Note that a_1 and b_1 only affect the expected profits through $\tilde{s}_2(s_1)$. Therefore, the first order condition for either a_1 or b_1 implies

$$\tilde{s}_2(s_1) = \frac{w_1 [l_d + y] + r_1 - a_2}{b_2 + w_1 y} - \frac{w_1 y}{b_2 + w_1 y} s_1 \quad (49)$$

Then the condition $B_1(s_1) = B_2(\tilde{s}_2(s_1))$ implies

$$\begin{aligned} a_1 + b_1 s_1 &= a_2 + b_2 \tilde{s}_2(s_1) \\ &= a_2 + \frac{b_2 (w_1 [l_d + y] + r_1 - a_2)}{b_2 + w_1 y} - \frac{b_2 w_1 y}{b_2 + w_1 y} s_1 \end{aligned} \quad (50)$$

This implies

$$b_1 = \frac{b_2 w_1 y}{b_2 + w_1 y} - \frac{w_1 y}{y} \quad (51)$$

and

$$\begin{aligned}
 a_1 &= a_2 + \frac{b_2(w_1[l_d + y] + r_1 - a_2)}{b_2 + w_1y} \\
 &= \frac{w_1l_d + w_1y + \frac{w_1}{b_2}(a_2 - w_1l_d - r_1)}{b_2 + w_1y} \\
 &= \frac{w_1l_d + w_1y + \frac{w_1}{w_2}(a_2 - w_1l_d - r_1)}{b_2 + w_1y} \\
 &= \frac{w_1l_d + \frac{1}{2}y + r_1}{w_2[l_d + \frac{1}{2}y] + r_2} \tag{52}
 \end{aligned}$$

Therefore, described bid function is incentive compatible for lender 1.

B.9 Proof of Proposition 6

This section shows

$$\underline{s}_1 = \frac{2}{3} \left(1 + \frac{2w_i - w_i}{w_1w_2} \frac{(w_i l_d + r_i - A)}{y} \right) \tag{53}$$

The threshold is defined by the point where a lender's expected profits is equal to zero under the assumption that the other lender does not compete. First, consider the case of lender 1.

To compute \underline{s}_1 , consider that, given that lender 2 will not compete, lender 1 charges an interest rate A and has expected zero-profits interest rate of $E_{s_2}[R_1(D(s_1, s_2; y)) | s_2 < \underline{s}_2]$.

Recall that the predictive distribution is uniform. Therefore the conditional pdf for s_2 is given by $f(s_2 | s_2 < \underline{s}_2) = \frac{1}{\underline{s}_2}$. Therefore

$$\begin{aligned}
 E[s_2 | s_2 < \underline{s}_2] &= \int_0^{\underline{s}_2} \frac{s_2}{\underline{s}_2} ds_2 \\
 &= \frac{1}{2} \underline{s}_2 \tag{54}
 \end{aligned}$$

Therefore we have

$$\begin{aligned} E_{s_2}[R_1(D(s_1, s_2)) | s_2 < \underline{s}_2] &= w_1 \left[\frac{h}{d+1} \underline{s}_1 + E_{s_2 | s_2 < \underline{s}_2}^i y + r_1 \right] \\ &= w_1 \left[\frac{h}{d+1} \underline{s}_1 + \frac{1}{2} \underline{s}_2 y + r_1 \right] \end{aligned} \quad (55)$$

Finally, as mentioned above, \underline{s}_1 is the point where lender 1 has zero expected profits, which is determined by the condition:

$$\begin{aligned} 0 &= A - E_{s_2}[R_1(D(s_1, s_2)) | s_2 < \underline{s}_2] \\ &= A - w_1 \left[\frac{h}{d+1} \underline{s}_1 + \frac{1}{2} \underline{s}_2 y + r_1 \right] \end{aligned} \quad (56)$$

An analogous equation also holds for lender 2. Then the system of equations implies the result.

Note that if $\underline{s}_2 = 1$, then lender 2 never makes an offer, and therefore lender 1 cannot make any inferences about s_2 based on the observation that lender 2 does not make an offer. In that case, \underline{s}_1 is instead defined by

$$\begin{aligned} 0 &= A - w_1 \left[\frac{h}{d+1} \underline{s}_1 + \frac{1}{2} y + r_1 \right] \\ \Rightarrow \underline{s}_1 &= \frac{1}{2} + \frac{w_1 \left[\frac{h}{d+1} + r_1 \right] - A}{w_1 y} \end{aligned} \quad (57)$$

An analogous argument determines \underline{s}_2 when $\underline{s}_1 = 1$.

Internet appendix

C Additional supplemental material for Section 2

C.1 Supplemental material for Section 2.3.1

Table C.1: Interest rates and observable risk with lock date fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)
	IR	IR	IR	IR	IR	IR
Observable risk	0.159*** (319.62)		0.060*** (51.66)		0.109*** (45.70)	0.040*** (31.36)
Credit score		-0.257*** (-308.29)		-0.067*** (-33.64)		
LTV		0.689*** (287.26)		0.358*** (33.41)		
DTI		0.379*** (103.38)		0.391*** (108.20)		
Observations	914,061	914,061	914,061	914,061	523,713	390,347
R ²	0.421	0.469	0.482	0.489	0.435	0.482
Lock Date FE	Yes	Yes	Yes	Yes	Yes	Yes
G-fee FE	No	No	Yes	Yes	Yes	Yes

Note: Column (1) regresses the interest rate on observable risk while controlling for lock rate fixed effects. Observable risk is the estimated probability of default based on credit score, the loan-to-value ratio, and the debt-to-income ratio as described in Section 2.1 (based on the model estimated using the MLIS sample). Column (2) regresses the interest rate on credit score, the loan-to-value (LTV) ratio, and the debt-to-income (DTI) ratio (each divided by 100). Column (3) and column (4) are similar to column (1) and column (2) except including fixed effects for the upfront g-fee as a function of credit score and LTV based on the first table of the GSEs' g-fee matrix. Column (5) and column (6) are similar to column (3) except restricting to loans with LTV less than or equal to 80% or LTV greater than 80%, respectively. T-statistics computed using robust standard errors are reported in parentheses. * indicates statistical significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level. Source: Optimal Blue, 2016-2017, restricting to conforming, fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

Table C.2: Interest rates and observable risk with lock date fixed effects for government-insured loans

	(1)	(2)
	IR	IR
Credit score	-0.279*** (-323.07)	-0.115*** (-30.17)
LTV	0.277*** (42.81)	-0.401*** (-25.05)
DTI	0.160*** (38.57)	0.170*** (41.49)
Observations	640,733	640,733
R ²	0.387	0.417
Lock Date FE	Yes	Yes
Credit score-LTV FE	No	Yes

Note: Column (1) regresses the interest rate on credit score, the loan-to-value (LTV) ratio, and the debt-to-income (DTI) ratio (each divided by 100). Column (2) is similar except including fixed effects for grid cells in credit score and LTV corresponding to the first table of the GSEs' g-fee matrix. Note that column (2) is only for purposes of comparison, as the GSEs' g-fees do not apply to loans insured by government agencies. Source: Optimal Blue, 2016-2017, restricting to fixed rate, purchase or no cash-out refinance loans insured by FHA, VA, or USDA for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

C.2 Supplemental material for Section 2.3.4

Table C.3: Denial reasons for loans accepted by GSE AUS

Denial reason	Count	Percent
Debt-to-income ratio	15,5683	12.98
Employment history	3,283	2.71
Credit history	9,019	7.46
Collateral	30,0490	25.24
Insufficient cash (downpayment, closing costs)	6,885	5.69
Unverifiable information	8,386	6.94
Credit application incomplete	37,7562	31.09
Mortgage insurance denied	131	.10
Other	9,355	7.74
Exempt	3	.00

Source: mortgage application data, 2018, restricting to applications accepted by the GSE automated underwriting systems for conventional, purchase or no cash-out refinance, first lien loan applications for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit.

C.3 Supplemental material for Section 2.4.1

Figure C.1: Distributions of observable risk

These figures present the distribution of observable risk for banks, nonbank-non-ntechs, and ntechs. Observable risk is the estimated probability of default based on credit score, the loan-to-value ratio, and the debt-to-income ratio as described in Section 2.1. Figure C.1a shows the kernel density, which is computed using the Epanechnikov kernel with a bandwidth of 0.1. Figure C.1b shows the cumulative distribution function. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit and loans with subordinate financing.

(a) Density

(b) Cumulative distribution function

Figure C.2: Distributions of risk characteristics

These figures present the distribution of credit score, the loan-to-value ratio, the debt-to-income ratio, and observable risk for loans originated by banks and nonbanks. Observable risk is the estimated probability of default based on credit score, the loan-to-value ratio, and the debt-to-income ratio as described in Section 2.1. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit and loans with subordinate financing.

(a) Credit score

(b) Loan-to-value ratio

(c) Debt-to-income

(d) Observable risk

C.4 Supplemental material for Section 2.4.2

Table C.4: Default, observable risk, and lender type: distinguish by ntech

(a) Nonbank versus bank				
	(1)	(2)	(3)	(4)
	Base	Obs. risk	Controls	+ IR
Nonbank non- ntech	0.275*** (22.69)	0.193*** (16.16)	0.177*** (14.77)	0.157*** (12.81)
Observable risk		0.952*** (64.10)		
IR - g-fee				0.336*** (7.11)
Observations	1,962,173	1,962,173	1,962,162	1,962,162
R ²	0.093	0.103	0.106	0.106
ZIP Year-quarter FE	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes

(b) Fintech vs non- ntech				
	(1)	(2)	(3)	(4)
	Base	Obs. risk	Controls	+ IR
Fintech	0.262*** (10.84)	0.193*** (8.02)	0.204*** (8.41)	0.181*** (7.35)
Observable risk		0.835*** (48.93)		
IR - g-fee				0.268*** (6.68)
Observations	1,351,868	1,351,868	1,351,856	1,351,856
R ²	0.118	0.125	0.129	0.129
ZIP Year-quarter FE	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes

Note: In the first table, ntechs are excluded. Column (1) regresses an indicator for default (multiplied by 100) on an indicator for nonbank-non ntechs. Column (2) adds observable risk as a regressor. Column (3) instead includes the following controls: the interaction between 10-point credit score bins (starting at 620, with an additional indicator for all credit scores below 620), 5% loan-to-value bins (starting at 60%, with an additional indicator for all loan-to-value ratios below 60%), and debt-to-income decile indicators (note that this absorbs observable risk); income decile indicators; family type indicators (i.e. single female, single male, or more than 1 borrower); indicators for black and hispanic borrowers; term indicators; appraisal value decile indicators; an indicator for a loan having an interest-only period; an indicator for a refinance loan; loan amount decile indicators; an indicator for self-employed borrowers; an indicator for first-time homebuyers; an indicator for full income documentation; and an indicator for full asset documentation. Column (4) additionally adds the interest rate net of the total g-fee. The second table is similar except that it uses an indicator for ntechs and excludes nonbank-non ntechs. T-statistics computed using robust standard errors are reported in parentheses. * indicates statistical significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

C.5 Supplemental material for Section 2.4.3

Table C.5: Interest rates, observable risk, and lender type: distinguish by ntech

(a) Nonbank versus bank

	(1) Base	(2) Obs. risk	(3) Controls
Nonbank non- ntech	0.084*** (153.04)	0.075*** (141.63)	0.060*** (135.31)
Observable risk		0.096*** (323.00)	
Observations	1,962,173	1,962,173	1,962,162
R ²	0.456	0.485	0.661
ZIP Year-quarter FE	Yes	Yes	Yes
Controls	No	No	Yes

(b) Fintech vs bank

	(1) Base	(2) Obs. risk	(3) Controls
Fintech	0.086*** (78.77)	0.079*** (72.65)	0.085*** (91.82)
Observable risk		0.091*** (247.92)	
Observations	1,351,868	1,351,868	1,351,856
R ²	0.479	0.504	0.677
ZIP Year-quarter FE	Yes	Yes	Yes
Controls	No	No	Yes

Note: In the first table, ntechs are excluded. Column (1) regresses the interest rate net of the total g-fee on an indicator for nonbank-non ntechs. Column (2) adds observable risk as a regressor. Column (3) instead includes the following controls: the interaction between 10-point credit score bins (starting at 620, with an additional indicator for all credit scores below 620), 5% loan-to-value bins (starting at 60%, with an additional indicator for all loan-to-value ratios below 60%), and debt-to-income decile indicators (note that this absorbs observable risk); income decile indicators; family type indicators (i.e. single female, single male, or more than 1 borrower); indicators for black and hispanic borrowers; term indicators; appraisal value decile indicators; an indicator for a loan having an interest-only period; an indicator for a refinance loan; loan amount decile indicators; an indicator for self-employed borrowers; an indicator for first-time homebuyers; an indicator for full income documentation; and an indicator for full asset documentation. Column (4) additionally adds the interest rate net of the total g-fee. The second table is similar except that it uses an indicator for ntechs and excludes nonbank-non ntechs. T-statistics computed using robust standard errors are reported in parentheses. * indicates statistical significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2016-2017, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, loans with subordinate financing, and loans where the upfront g-fee is deviates from the first table of the g-fee matrix by more than 25 basis points.

C.6 Supplemental material for Section 2.4.4

Figure C.3: Interest rates, default, observable risk, lender type, and fintech market share

Figure C.3a presents a binned scatterplot of the default rate, split by banks versus nonbanks and counties in the top versus bottom tercile of fintech market share, during 2011-2012. Figure C.3b is similar except that the sample period is 2016-2017. Figure C.3c and Figure C.3d are respectively analogous except showing the interest rate net of the total g-fee and including year-month fixed effects. Observable risk is the estimated probability of default based on credit score, the loan-to-value ratio, and the debt-to-income ratio as described in Section 2.1. Source: Mortgage Loan Integrated System (Fannie Mae and Freddie Mac), 2011-2017, restricting to fixed rate, purchase or no cash-out refinance loans for one-unit, owner-occupied, single-family detached houses and excluding high balance loans exceeding the base conforming loan limit, and loans with subordinate financing.

(a) Default (2011-2012)

(b) Default (2016-2017)

(c) Interest rate (2011-2012)

(d) Interest rate (2016-2017)

D Additional supplemental material for Section 3

Figure D.1: Number of offers

This figure shows the probability that a borrower receives 2, 1, or 0 offers in the model for active intermediation (the baseline model in which lenders screen the applicant, approve or deny the application, and engage in imperfect competition to determine the interest rate).

Figure D.2: Probability of a single offer

This figure shows the probability that a borrower receives exactly 1 offer conditional on receiving an offer for the baseline model for active intermediation (the baseline model in which lenders screen the applicant, approve or deny the application, and engage in imperfect competition to determine the interest rate), for a low loss given default w and a high loss given default.

Figure D.3: Probability of credit outcomes with heterogeneous γ

These figures show the probability of various credit outcomes in the version of the model with 2 lenders with exogenous and different information levels γ (described in Section 5.1). Figure D.3a shows the probability that a borrower receives credit from the lender with high γ , receives credit from the lender with low γ , and does not receive credit. Figure D.3b shows the probability that a borrower receives 2, 1, or 0 offers. Figure D.3c shows the probabilities corresponding to each lender type and number of offers.

(a) By lender type

(b) By offer count

(c) By lender type and offer count

Figure D.4: Probability of credit outcomes with heterogeneous w

These figures show the probability of various credit outcomes in the version of the model with 2 lenders with exogenous and the same information levels y but different losses given default w (described in Section 5.2). Figure D.4a shows the probability that a borrower receives credit from the lender with high y , receives credit from the lender with low y , and does not receive credit. Figure D.4b shows the probability that a borrower receives 2, 1, or 0 offers. Figure D.4c shows the probabilities corresponding to each lender type and number of offers.

(a) By lender type

(b) By offer count

(c) By lender type and offer count

Figure D.5: Association between interest rate and default

Figure D.5a shows the association between the average interest rate and the default rate across l_d in the model for active intermediation (the baseline model in which lenders screen the applicant, approve or deny the application, and engage in imperfect competition to determine the interest rate). Figure D.5b shows the association with a decomposition of the interest rate between the risk spread and the markup.

(a) Default rate on interest rate

(b) Interest rate on default rate

Figure D.6: Association between interest rate and default with heterogeneous γ

This figure shows the association between the average interest rate and the default rate across l_d in the version of the model with 2 lenders with exogenous and different information levels γ (described in Section 5.1).

Figure D.7: Association between interest rate and default with heterogeneous w

This figure shows the association between the average interest rate and the default rate across l_d in the version of the model with 2 lenders with exogenous γ but different losses given default w (described in Section 5.2).

Figure D.8: Participation thresholds with heterogeneous w

This figure shows the participation thresholds s_i (as in equation (28)) across l_d in the version of the model with 2 lenders with exogenous γ and different losses given default w (described in Section 5.2).

Figure D.9: Active intermediation with heterogeneous and homogeneous w

These figures show various features of the version of the model with 2 lenders with exogenous information levels and different loss given default (described in Section 5.2) or 2 lender with the higher loss given default, w_1 . The probability of receiving credit is the probability that at least one lender approves the application. The default rate is the fraction of approved applications that consist of defaulting borrowers. The average interest rate is the average interest payment divided by the probability of receiving credit. The average origination cost is the average zero-profits interest rate of the supplying lender conditional on its own signal and inferring from the equilibrium the signal of the next most competitive lender. The average markup is a lender's total expected profits (average interest rate - average origination cost). $\Pr(1 \text{ offer} | \text{receiving credit})$ is the probability that the consumer receives only one offer conditional on receiving an offer. Parameters: $r = 0$, $n = 2$, $A = .0026$, $w = .065$ and $.04$ for the heterogeneous case (or just $w = .065$ for the homogeneous case), $m = .1$, $y = .02$.

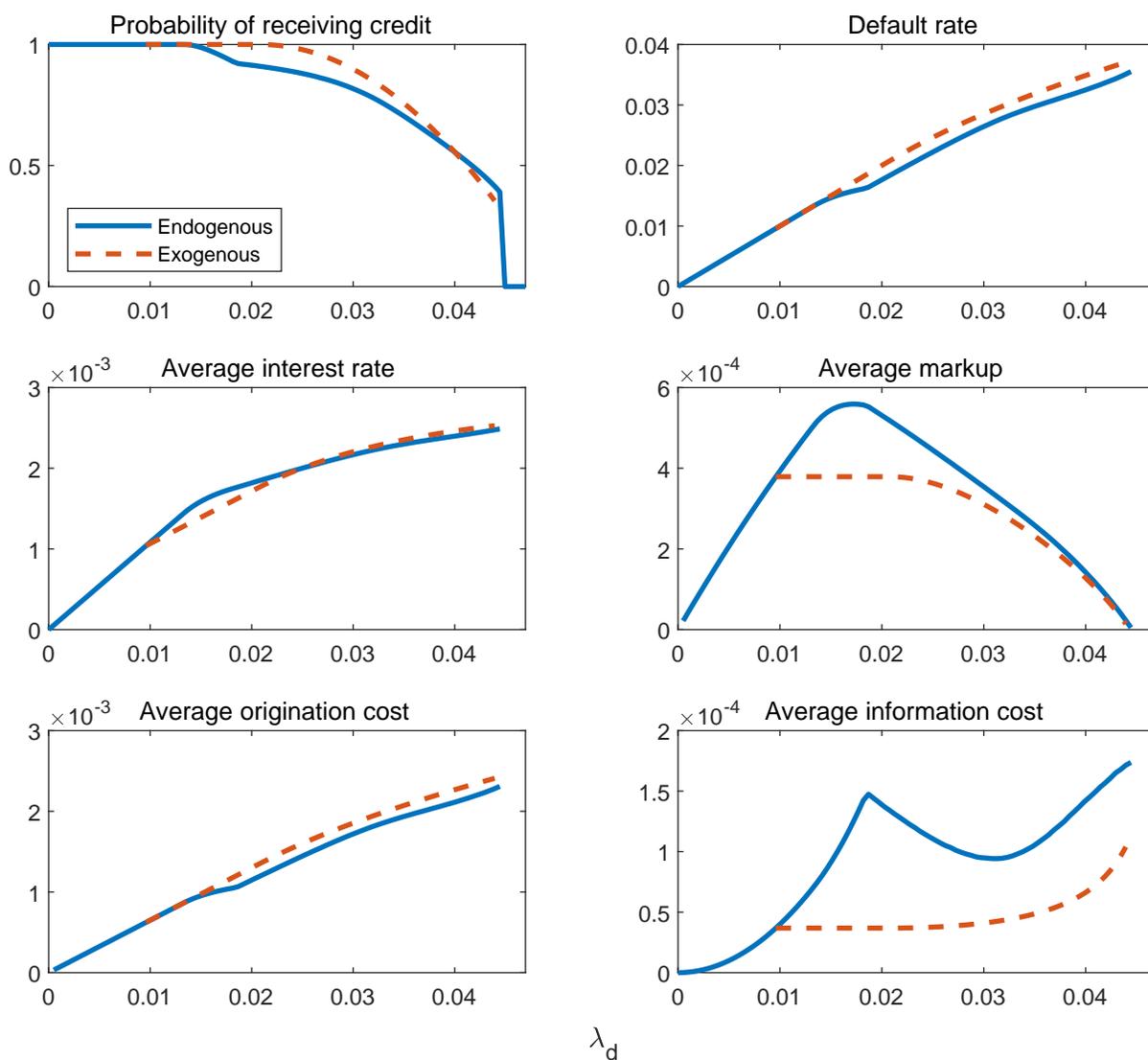
D.1 Endogenous versus exogenous screening

To illustrate the role of investment in screening technology, Figure D.10 compares active intermediation with endogenous information acquisition, as in the original model, versus active intermediation with active intermediation with information at a fixed level corresponding to the endogenously chosen information level at one particular value of I_d .⁴⁸ As I_d increases from this point, lenders endogenously would improve their screening technology, resulting in a higher markup but also a lower risk spread compared to holding their screening technology constant. Endogenous improvements in screening technology can actually lead to a higher interest rate in some cases by increasing the markup, but it is more likely to lead to a lower interest rate for sufficiently risky borrowers, whose interest rate is to a greater extent determined by origination costs.

⁴⁸Note that, in the case where the information level is exogenous, the outcomes are only computed for I_d sufficiently large due to the constraint $y \geq 2I_d$ from Section 3.3.1.

Figure D.10: Active intermediation with endogenous and exogenous information level

These figures show various features of the model in the case where the information level y corresponding to the quality of lender screening is endogenous or exogenous. The *probability of receiving credit* is the probability that at least one lender approves the application. The *default rate* is the fraction of approved applications that consist of defaulting borrowers. The *average interest rate* is the average interest payment divided by the probability of receiving credit. The *average origination cost* is the average zero-profits interest rate of the supplying lender conditional on its own signal and inferring from the equilibrium the signal of the next most competitive lender. The *information cost* is the cost associated with the parameter y corresponding to the quality of screening. The *average markup* is a lender's total expected profits (average interest rate - average origination cost - information cost). The *average markup* is a lender's total expected profits (average interest rate - average origination cost).



D.2 Generality of the distribution system

Up to a first order approximation in y , the distribution system defined by equations (2) and (3) can be assumed without loss of generality conditional on the following set of intuitive properties: the predictive distribution does not depend on the information level, the conditional distributions converge to the predictive distribution when the information level is equal to zero, and the first order effect of information on the conditional pdf for a good signal is given by the probability of receiving as high a signal under the predictive distribution.

To show this, consider a distribution system with conditional pdfs $f(s|q; y)$ and predictive distribution

$$f(s) = I_d f(s|d; y) + I_r f(s|r; y) \quad (58)$$

Since the conditional distributions converge to the predictive distribution when the information level is equal to zero, a first order approximation obtains

$$f(s|q; y) = f(s) + y \frac{\eta f(s|q; y)}{\eta y}$$

Differentiating (58) obtains

$$I_d \frac{\eta f(s|d; y)}{\eta y} = -I_r \frac{\eta f(s|r; y)}{\eta y}$$

Let

$$\begin{aligned} h(s) &= I_r \frac{\eta f(s|r; y)}{\eta y} f(s)^{-1} \\ &= -I_d \frac{\eta f(s|d; y)}{\eta y} f(s)^{-1} \end{aligned}$$

Then the conditional distributions can be written

$$\begin{aligned} f(s|d; y) &= f(s) \left(1 - h(s) \frac{y}{I_d} \right) \\ f(s|r; y) &= f(s) \left(1 + h(s) \frac{y}{I_r} \right) \end{aligned}$$

The assumption that the first order effect of information on the conditional pdf for a

good signal is given by the probability of receiving as high a signal under the predictive distribution is captured by choosing

$$h(s) = F(s) - \frac{1}{2}$$

Intuitively, $h(s)$ is equal to $F(s)$ plus a translation by $-\frac{1}{2}$ to ensure that the pdf integrates to 1. Finally, note that the implied distribution system

$$f(s|d; y) = f(s) \left[1 - \frac{1}{2} \frac{y}{l_d} \right]$$

$$f(s|r; y) = f(s) \left[1 + \frac{1}{2} \frac{y}{l_r} \right]$$

yields the same results as the distribution system with a uniform predictive distribution since all the relevant quantities are computed using integrals that are equivalent via a change in variables.