Internet appendix

This Appendix contains the following supplemental material:

1. Appendix D provides supplementary material for the empirical observations (Section 2)

    (a) Table D.1 shows summary statistics for the 2016-2017 sample of originations from the Mortgage Loan Information System (provided by Fannie Mae and Freddie Mac).

    (b) Table D.2 shows summary statistics for the 2018 sample of originations from the Mortgage Loan Information System (provided by Fannie Mae and Freddie Mac).

    (c) Figure D.1 shows origination revenue and its components as a function of observable risk.

    (d) Table D.3 estimates the association between origination revenue (and its components) and observable risk.

    (e) Table D.4 estimates the association between interest rates and observable risk when using data from Optimal Blue and controlling for lock date fixed effects.

    (f) Table D.5 estimates the association between interest rates and observable risk when using data from Optimal Blue and controlling for lock date fixed effects while restricting to government-insured loans.

    (g) Figure D.2 shows the association between default and origination revenue with and without controlling for observable risk.

    (h) Table D.6 estimates the association between default and origination revenue with and without controlling for observable risk.

    (i) Table D.7 shows the fraction of denials attributable to various reasons.

    (j) Table D.8 estimates the association between interest rates and observable risk while controlling for observable prepayment risk.

    (k) Table D.9 estimates the association between interest rates and default while controlling for observable risk and observable prepayment risk.

    (l) Figure D.3 shows the kernel density and cumulative distribution function of observable risk for different types of lender.

    (m) Figure D.4 shows the distributions of observable risk and constituent underwriting characteristics for different types of lenders.

    (n) Table D.10 estimates the association between default and either nonbank non-fintechs or fintechs relative to banks.

    (o) Table D.11 estimates the association between default and nonbanks over time.

    (p) Table D.12 estimates the association between interest rates and either nonbank non-fintechs or fintechs relative to banks.

(q) Figure D.5 shows the association between either defaults or interest rates and observable risk by lender type and fintech market share.

2. Appendix E provides supplementary material for the model (Section 3)

   (a) Subsection E.1 describes the generality of the signal distribution system in the model.

   (b) Subsection E.2 computes a lender's expected profits.

   (c) Subsection E.3 computes a consumer's expected surplus.

3. Appendix F provides supplementary material for the baseline model simulation (Section 4)

   (a) Figure F.1 shows the direct association between interest rates and default as functions of λ_d.

   (b) Figure F.2 shows a detailed decomposition of the number of offers received by a borrower as a function of λ_d.

   (c) Figure F.3 shows the probability of a borrower receiving exactly one offer conditional on receiving at least one offer as a function of λ_d.

   (d) Figure F.4 compares outcomes in the model when the information level is either endogenous or exogenous.

4. Appendix G provides supplementary material for the model simulation with heterogeneous lenders (Section 5)

   (a) Figure G.1 shows a decomposition of the number of offers received by a borrower as a function of λ_d in the version of the model with heterogeneous ψ.

   (b) Figure G.2 shows the association between interest rates and default as functions of λ_d in the version of the model with heterogeneous ψ.

   (c) Figure G.3 shows a decomposition of the number of offers received by a borrower as a function of λ_d in the version of the model with heterogeneous ω.

   (d) Figure G.4 shows the association between interest rates and default as functions of λ_d in the version of the model with heterogeneous ω.

   (e) Figure G.5 compares the case of 2 lenders with different ω to the case of 2 homogeneous lenders with the same ω.

   (f) Figure G.6 shows the lender participation thresholds for the signal in the version of the model with heterogeneous ω.

   (g) Figure G.7 compares model outcomes with different levels of ω and adjusting ρ according to equation (27).

Note that proofs are contained in Section C of the non-Internet Appendix.

D   Supplemental material for the observations (Section 2)

D.1   Supplemental material for the data description (Section 2.2)

Table D.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-nonfintech	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
Refinance	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 D.1a), the subsample of loans originated by banks (Table D.1b), the subsample of loans originated by nonbank-nonfintechs (Table D.1c), and the subsample of loans originated by fintechs (Table D.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-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-nonfintech* is an indicator for lenders that are neither banks nor fintechs. *Fintech* is an indicator for lenders with a mostly online application process based on the designation of fintechs 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. *Refinance* indicates refinance loans. *Interest-only* indicates loans with an interest-only portion. Continuous variables are winsorized at 1%. Source: Mortgage Loan Information 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.

Table D.1: Summary statistics (continued)

(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
Refinance	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 D.1a), the subsample of loans originated by banks (Table D.1b), the subsample of loans originated by nonbank-nonfintechs (Table D.1c), and the subsample of loans originated by fintechs (Table D.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-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-nonfintech* is an indicator for lenders that are neither banks nor fintechs. *Fintech* is an indicator for lenders with a mostly online application process based on the designation of fintechs 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. *Refinance* indicates refinance loans. *Interest-only* indicates loans with an interest-only portion. Continuous variables are winsorized at 1%. Source: Mortgage Loan Information 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.

Table D.1: Summary statistics (continued)

(c) Nonbank-nonfintechs

	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
Refinance	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 D.1a), the subsample of loans originated by banks (Table D.1b), the subsample of loans originated by nonbank-nonfintechs (Table D.1c), and the subsample of loans originated by fintechs (Table D.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-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-nonfintech* is an indicator for lenders that are neither banks nor fintechs. *Fintech* is an indicator for lenders with a mostly online application process based on the designation of fintechs 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. *Refinance* indicates refinance loans. *Interest-only* indicates loans with an interest-only portion. Continuous variables are winsorized at 1%. Source: Mortgage Loan Information 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.

Table D.1: Summary statistics (continued)

(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
Refinance	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 D.1a), the subsample of loans originated by banks (Table D.1b), the subsample of loans originated by nonbank-nonfintechs (Table D.1c), and the subsample of loans originated by fintechs (Table D.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-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-nonfintech* is an indicator for lenders that are neither banks nor fintechs. *Fintech* is an indicator for lenders with a mostly online application process based on the designation of fintechs 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. *Refinance* indicates refinance loans. *Interest-only* indicates loans with an interest-only portion. Continuous variables are winsorized at 1%. Source: Mortgage Loan Information 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.

Table D.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-nonfintech	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
Refinance	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 D.2a), the subsample of loans originated by banks (Table D.2b), the subsample of loans originated by nonbank-nonfintechs (Table D.2c), and the subsample of loans originated by fintechs (Table D.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 market income (as defined in Section 2.1) divided by the loan amount. *Bank* is an indicator for depositories. *Nonbank-nonfintech* is an indicator for lenders that are neither banks nor fintechs. *Fintech* is an indicator for lenders with a mostly online application process based on the designation of fintechs 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. *Refinance* indicates refinance 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 market 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 Information 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.

Table D.2: Summary statistics for the 2018 sample (continued)

(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
Refinance	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 D.2a), the subsample of loans originated by banks (Table D.2b), the subsample of loans originated by nonbank-nonfintechs (Table D.2c), and the subsample of loans originated by fintechs (Table D.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 market income (as defined in Section 2.1) divided by the loan amount. *Bank* is an indicator for depositories. *Nonbank-nonfintech* is an indicator for lenders that are neither banks nor fintechs. *Fintech* is an indicator for lenders with a mostly online application process based on the designation of fintechs 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. *Refinance* indicates refinance 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 market 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 Information 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.

Table D.2: Summary statistics for the 2018 sample (continued)

(c) Nonbank-nonfintechs

	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
Refinance	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 D.2a), the subsample of loans originated by banks (Table D.2b), the subsample of loans originated by nonbank-nonfintechs (Table D.2c), and the subsample of loans originated by fintechs (Table D.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 market income (as defined in Section 2.1) divided by the loan amount. *Bank* is an indicator for depositories. *Nonbank-nonfintech* is an indicator for lenders that are neither banks nor fintechs. *Fintech* is an indicator for lenders with a mostly online application process based on the designation of fintechs 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. *Refinance* indicates refinance 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 market 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 Information 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.

Table D.2: Summary statistics for the 2018 sample (continued)

(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
Refinance	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 D.2a), the subsample of loans originated by banks (Table D.2b), the subsample of loans originated by nonbank-nonfintechs (Table D.2c), and the subsample of loans originated by fintechs (Table D.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 market income (as defined in Section 2.1) divided by the loan amount. *Bank* is an indicator for depositories. *Nonbank-nonfintech* is an indicator for lenders that are neither banks nor fintechs. *Fintech* is an indicator for lenders with a mostly online application process based on the designation of fintechs 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. *Refinance* indicates refinance 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 market 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 Information 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.

D.2 Supplemental material for Observation 1 (Section 2.3.1)

Figure D.1: Origination revenue components and observable risk

This figure presents a binned scatterplot of origination revenue components (closing costs and secondary market 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 Information 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 deviates from the first table of the g-fee matrix by more than 25 basis points.
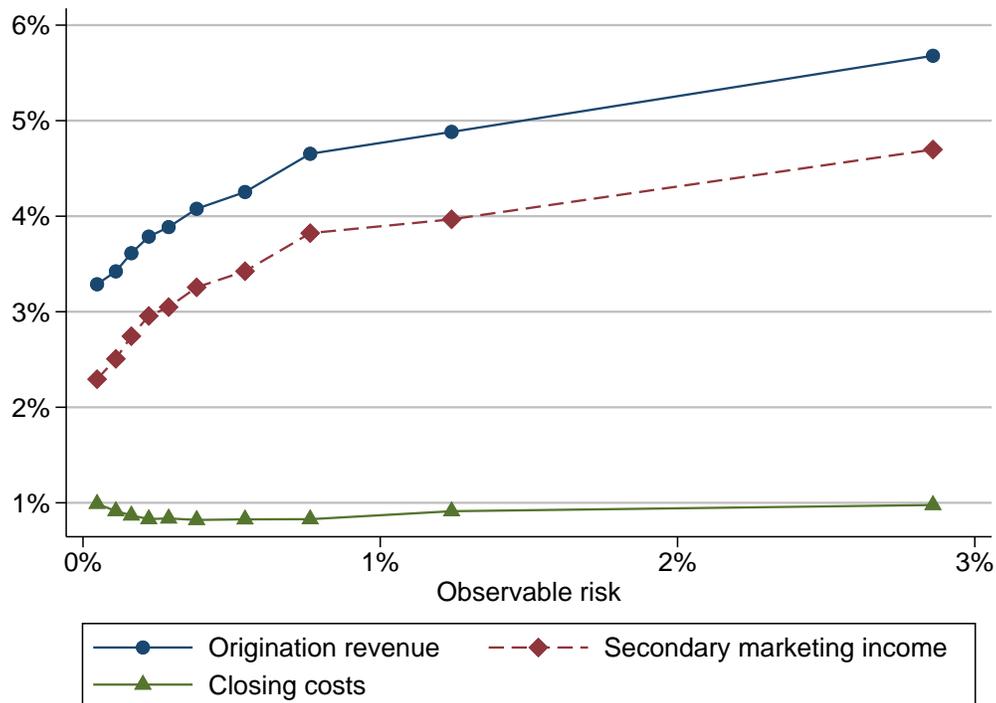
Table D.3: Origination revenue and observable risk

	(1) Orig. rev.	(2) Orig. rev.	(3) Closing costs	(4) Closing costs	(5) Second. income	(6) Second. income
Observable risk	0.597***		0.024***		0.573***	
	(165.62)		(16.08)		(165.51)	
Credit score		-1.093***		-0.168***		-0.926***
		(-153.76)		(-57.00)		(-138.18)
LTV		1.374***		-1.131***		2.516***
		(67.21)		(-138.43)		(133.31)
DTI		0.974***		-0.014		0.992***
		(28.90)		(-1.03)		(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 the secondary market 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 Information 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 deviates from the first table of the g-fee matrix by more than 25 basis points.

Table D.4: Interest rates and observable risk with lock date fixed effects

	(1) IR	(2) IR	(3) IR	(4) IR	(5) IR	(6) IR
Observable risk	0.159***		0.060***		0.109***	0.040***
	(319.62)		(51.66)		(45.70)	(31.36)
Credit score		-0.257***		-0.067***		
		(-308.29)		(-33.64)		
LTV		0.689***		0.358***		
		(287.26)		(33.41)		
DTI		0.379***		0.391***		
		(103.38)		(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 the 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 deviates from the first table of the g-fee matrix by more than 25 basis points.
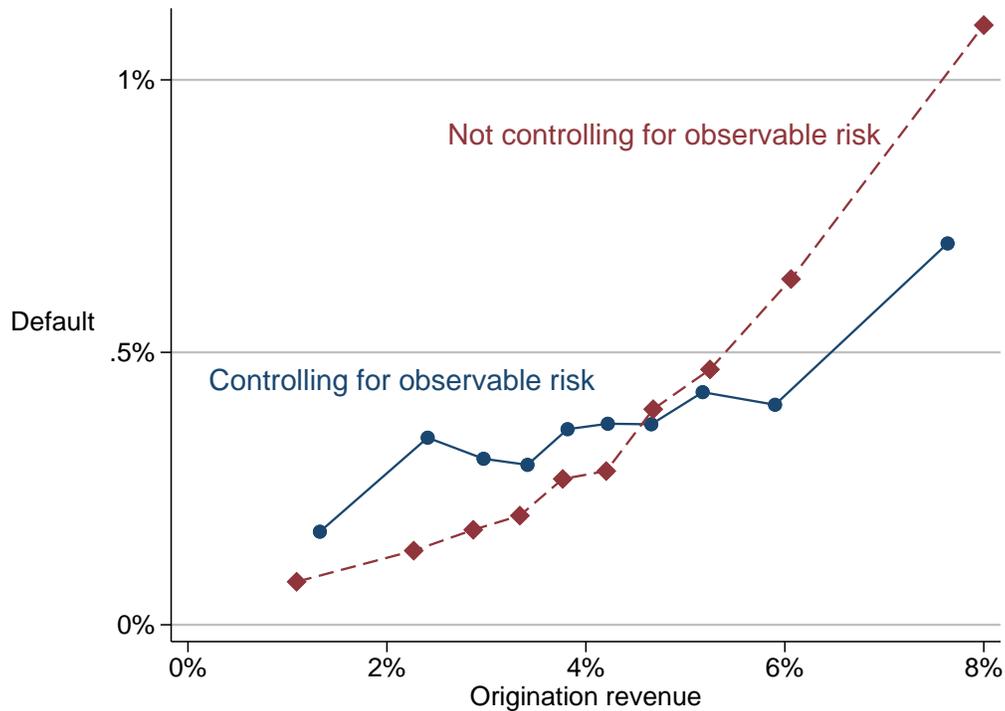
Table D.5: Interest rates and observable risk with lock date fixed effects for government-insured loans

	(1) IR	(2) IR
Credit score	-0.279***	-0.115***
	(-323.07)	(-30.17)
LTV	0.277***	-0.401***
	(42.81)	(-25.05)
DTI	0.160***	0.170***
	(38.57)	(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 deviates from the first table of the g-fee matrix by more than 25 basis points.

D.3 Supplemental material for Observation 2 (Section 2.3.2)

Figure D.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 Information 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.

Table D.6: Origination revenue and default

	(1) Baseline	(2) Obs. risk	(3) Controls	(4) Safe	(5) Risky	(6) Interact
Orig. rev.	0.150***	0.062***	0.063***	0.022***	0.103***	0.022***
	(20.78)	(8.90)	(7.56)	(4.43)	(7.51)	(4.43)
Observable risk		0.557***		0.190**	0.575***	0.190**
		(22.93)		(2.20)	(17.64)	(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 Information 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.

D.4 Supplemental material for Observation 4 (Section 2.3.4)

Table D.7: Denial reasons for loans accepted by GSE AUS

Denial reason	Count	Percent
Debt-to-income ratio	12,2261	11.40
Employment history	2,334	2.17
Credit history	7,264	6.75
Collateral	27,7128	25.23
Insufficient cash (downpayment, closing costs)	5,603	5.21
Unverifiable information	7,169	6.66
Credit application incomplete	36,6729	34.16
Mortgage insurance denied	70	.06
Other	8,957	8.33
Exempt	3	.00

Source: mortgage application data, 2018, restricting to applications accepted by the GSE automated underwriting systems but denied by the lender 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. For simplicity, we also restrict to loans within this sample that report only one denial reason in order to have non-intersecting categories, which includes about 89% of observations.

D.5   Supplemental material for Observations 1-4 overview (Section 2.3.5)

Table D.8: Interest rates, observable risk, and observable prepayment risk

	(1) IR	(2) IR - g-fee	(3) IR - g-fee
Observable risk	0.237***	0.129***	0.117***
	(550.16)	(344.19)	(331.79)
Obs. prepayment risk	-0.024***	-0.027***	-0.025***
	(-125.76)	(-155.45)	(-154.15)
Observations	2,109,041	2,109,041	2,108,976
R^2	0.496	0.502	0.558
ZIP × Year-quarter FE	Yes	Yes	Yes
Intermediary FE	No	No	Yes

Note: Column (1) regresses the interest rate on observable risk and observable prepayment 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. Observable prepayment risk is defined analogously but based on prepayment within 2 years. Column (2) is similar to column (1) except that the dependent variable is the interest rate net of the total g-fee. Column (3) is similar to column (2) except also controlling for intermediary fixed effects. 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 Information 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.

Table D.9: Interest rates, default, and prepayment risk

	(1) All	(2) Safe	(3) Risky
IR - g-fee	0.345***	0.130***	0.680***
	(18.38)	(9.84)	(17.31)
Obs. risk	0.930***	0.718***	0.931***
	(53.38)	(11.72)	(44.77)
Obs. prepay risk	-0.002	0.002	0.010
	(-0.35)	(0.35)	(1.37)
Observations	2,109,041	1,030,232	1,040,160
R^2	0.100	0.137	0.137
ZIP $\times$ Year-quarter FE	Yes	Yes	Yes

Note: Column (1) regresses an indicator for default (multiplied by 100) on the interest rate net of the total g-fee, observable risk, and observable prepayment 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. Observable prepayment risk is defined analogously but based on prepayment within 2 years. Column (2) estimates the specification in column (1) except restricting to relatively safe borrowers with observable risk below the median. Column (3) estimates the specification in column (1) except restricting to relatively risky borrowers with observable risk above the median. 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 Information 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 deviates from the first table of the g-fee matrix by more than 25 basis points.

D.6 Supplemental material for Observation 5 (Section 2.4.1)

Figure D.3: Distributions of observable risk

These figures present the distribution of observable risk for banks, nonbank-nonfintechs, and fintechs. 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 D.3a shows the kernel density, which is computed using the Epanechnikov kernel with a bandwidth of 0.1. Figure D.3b shows the cumulative distribution function. Source: Mortgage Loan Information 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
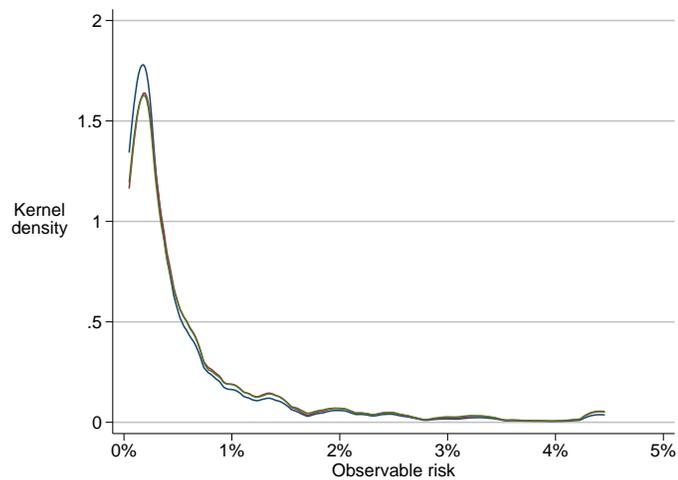
(b) Cumulative distribution function
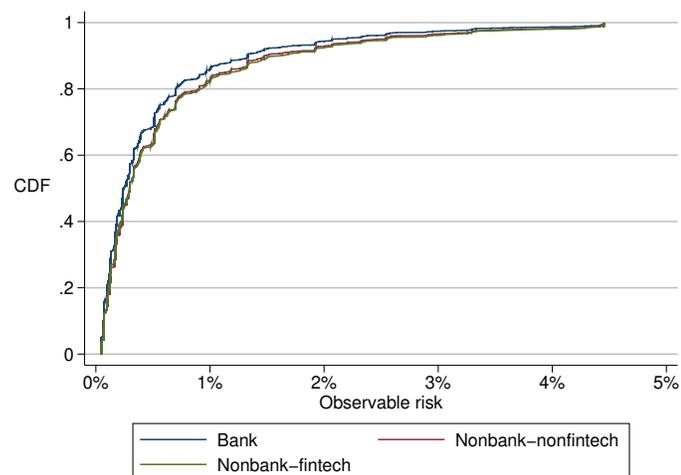
Figure D.4: Distributions of risk characteristics

These figure 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 Information 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
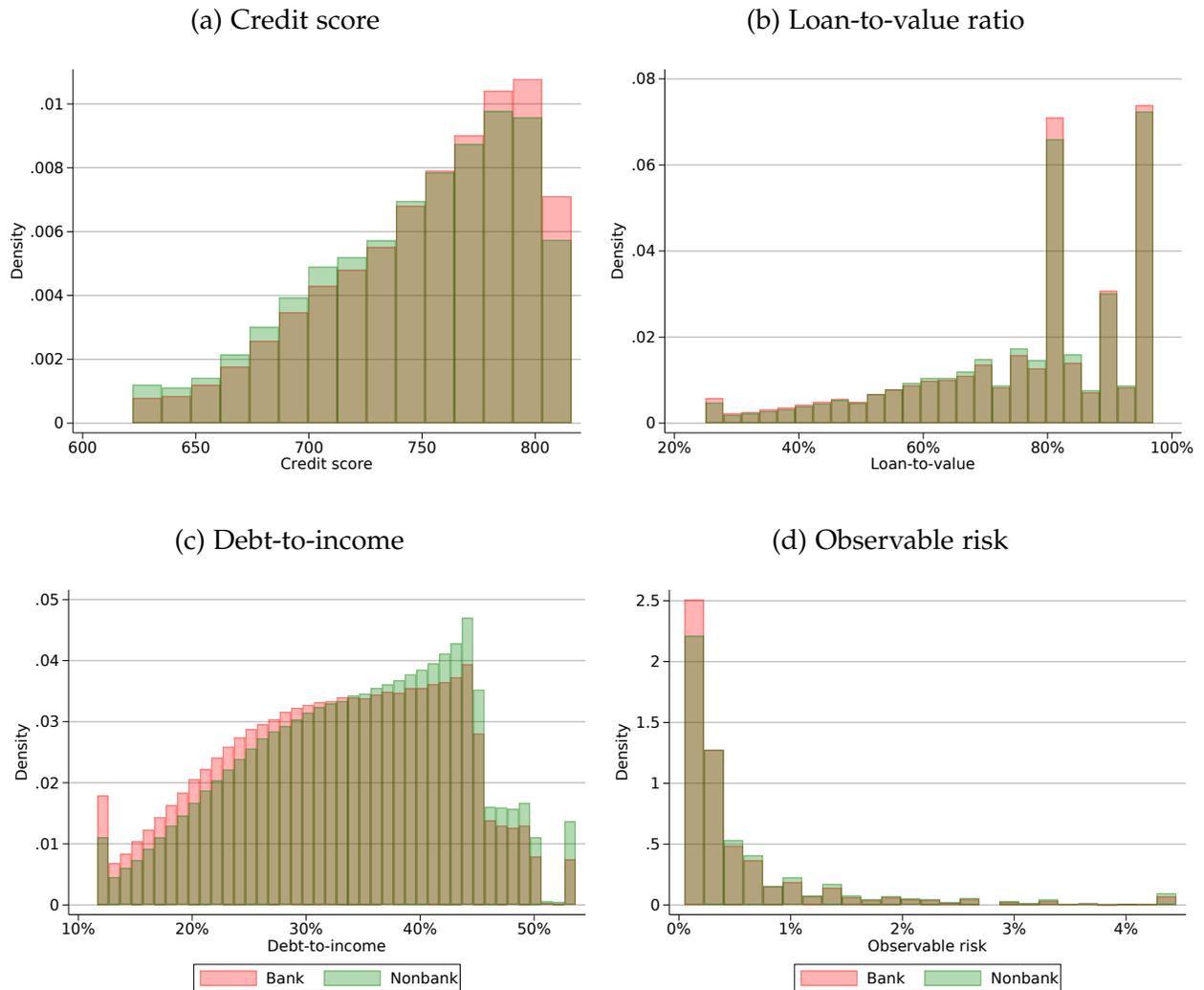
(d) Observable risk

D.7 Supplemental material for Observation 6 (Section 2.4.2)

Table D.10: Default, observable risk, and lender type: distinguish by fintech

(a) Nonbank versus bank

	(1) Base	(2) Obs. risk	(3) Controls	(4) + IR
Nonbank non-fintech	0.275***	0.193***	0.177***	0.157***
	(22.69)	(16.16)	(14.77)	(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-fintech

	(1) Base	(2) Obs. risk	(3) Controls	(4) + IR
Fintech	0.262***	0.193***	0.204***	0.181***
	(10.84)	(8.02)	(8.41)	(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, fintechs are excluded. Column (1) regresses an indicator for default (multiplied by 100) on an indicator for nonbank-nonfintechs. 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 fintechs and excludes nonbank-nonfintechs. 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 Information 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.

Table D.11: Default, observable risk, and lender type over time

	(1) 2011	(2) 2012	(3) 2013	(4) 2014	(5) 2015	(6) 2016	(7) 2017
Nonbank	-0.051***	0.034***	0.085***	0.121***	0.156***	0.198***	0.255***
	(-3.44)	(2.85)	(6.44)	(7.28)	(11.41)	(13.97)	(14.32)
Observable risk	1.094***	1.015***	1.003***	1.029***	1.014***	0.965***	0.998***
	(56.26)	(67.31)	(64.17)	(49.43)	(50.19)	(59.53)	(60.70)
Observations	1,453,258	2,251,273	1,775,251	1,012,013	1,226,247	1,437,160	1,130,299
R^2	0.081	0.060	0.068	0.092	0.085	0.090	0.093
ZIP × Year-quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Each column regresses an indicator for default (multiplied by 100) on an indicator for nonbanks while controlling for ZIP code by year-quarter fixed effects for a particular year. T-statistics computed using robust standard errors are reported in parentheses. Note that 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 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. * indicates statistical significance at the 10% level, ** indicates significance at the 5% level, and *** indicates significance at the 1% level. Source: Mortgage Loan Information 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.

D.8 Supplemental material for Observation 7 (Section 2.4.3)

Table D.12: Interest rates, observable risk, and lender type: distinguish by fintech

(a) Nonbank versus bank

	(1) Base	(2) Obs. risk	(3) Controls
Nonbank non-fintech	0.084***	0.075***	0.060***
	(153.04)	(141.63)	(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***	0.079***	0.085***
	(78.77)	(72.65)	(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, fintechs are excluded. Column (1) regresses the interest rate net of the total g-fee on an indicator for nonbank-nonfintechs. 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 fintechs and excludes nonbank-nonfintechs. 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 Information 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.

D.9 Supplemental material for Observation 8 (Section 2.4.4)

Figure D.5: Interest rates, default, observable risk, lender type, and fintech market share

Figure D.5a 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 D.5b is similar except that the sample period is 2016-2017. Figure D.5c and Figure D.5b 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 Information 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)

85
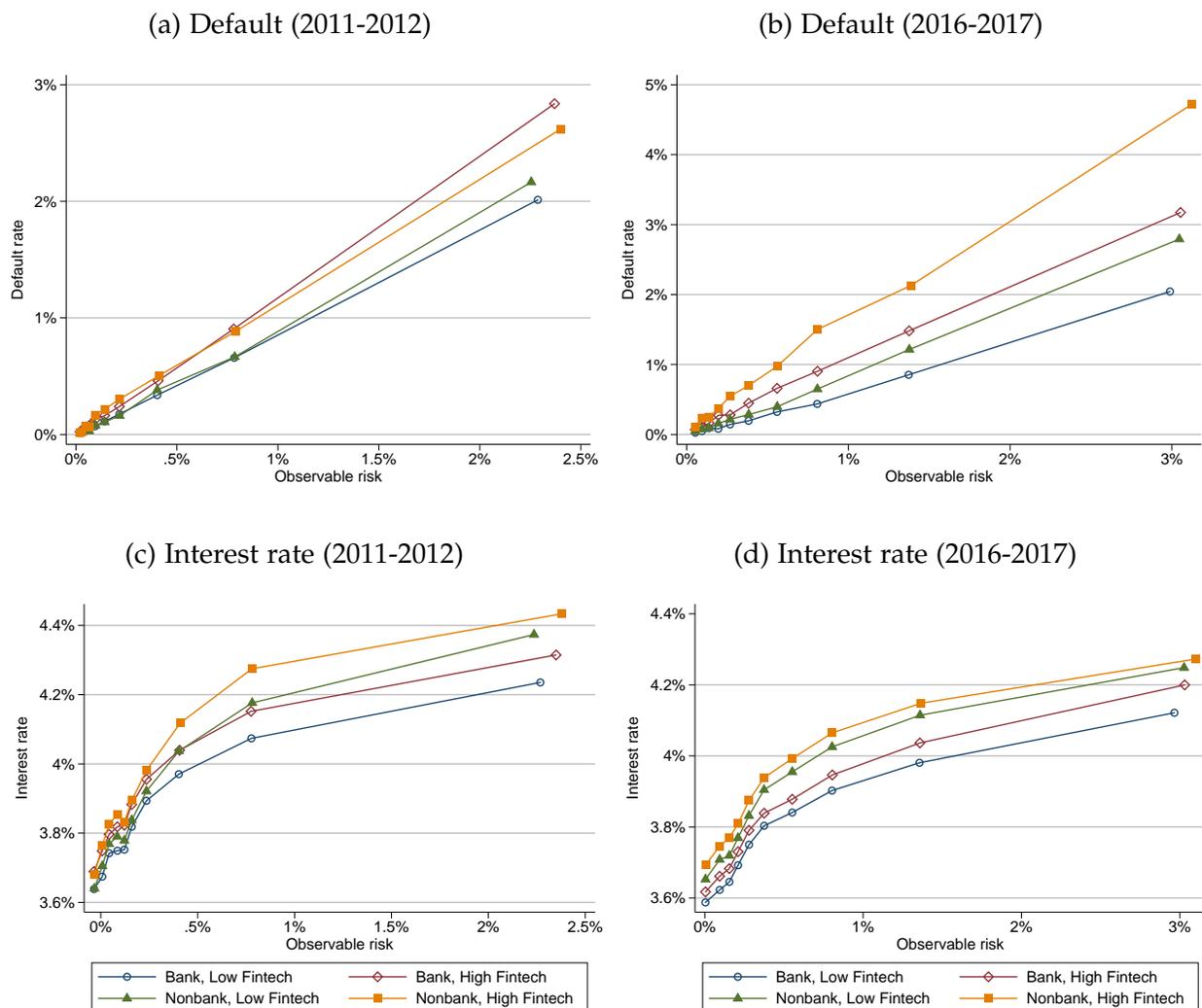
E   Supplemental material for the model (Section 3)

E.1   Generality of the distribution system

Up to a first-order approximation in ψ, 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 $f(s|\theta;\psi)$ and predictive distribution

$$f(s) = \lambda_d f(s|d;\psi) + \lambda_r f(s|r;\psi) \tag{63}$$

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

$$f(s|\theta;\psi) = f(s) + \psi \frac{\partial f(s|\theta;\psi)}{\partial \psi}$$

Differentiating (63) obtains

$$\lambda_d \frac{\partial f(s|d;\psi)}{\partial \psi} = -\lambda_r \frac{\partial f(s|r;\psi)}{\partial \psi}$$

Let

$$h(s) = \lambda_r \frac{\partial f(s|r;\psi)}{\partial \psi} f(s)^{-1}$$
$$= -\lambda_d \frac{\partial f(s|d;\psi)}{\partial \psi} f(s)^{-1}$$

Then the conditional distributions can be written

$$f(s|d;\psi) = f(s)\left(1 - h(s)\frac{\psi}{\lambda_d}\right)$$
$$f(s|r;\psi) = f(s)\left(1 + h(s)\frac{\psi}{\lambda_r}\right)$$

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

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

86

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

$$f(s|d;\psi) = f(s)\left(1 - \left(F(s) - \frac{1}{2}\right)\frac{\psi}{\lambda_d}\right)$$

$$f(s|r;\psi) = f(s)\left(1 + \left(F(s) - \frac{1}{2}\right)\frac{\psi}{\lambda_r}\right)$$

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.

E.2   Calculation for equation (14)

This section computes lender 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

$$
\begin{aligned}
\pi_L(s) &= E_{s_{n-1:n}}\left[(A - \underline{R}(D(s, s_{n-1:n}; \psi, n)))\mathbb{1}_{\{s = s_{n:n}, s_{n-1:n} \leq \underline{s}\}}\right] \\
&\quad + E_{s_{n-1:n}}\left[(\underline{R}(D(s_{n-1:n}, s_{n-1:n}; \psi, n)) - \underline{R}(D(s, s_{n-1:n}; \psi, n)))\mathbb{1}_{\{s = s_{n:n}, s_{n-1:n} \geq \underline{s}\}}\right] \\
&= \int_0^{\underline{s}} (A - \omega\lambda_d - \rho)(n - 1)\frac{t^{n-2}}{s^{n-1}}dt \\
&\quad - \omega\psi\frac{1}{2}\int_0^{\underline{s}} (n - 2s - nt)(n - 1)\frac{t^{n-2}}{s^{n-1}}dt \\
&\quad + \int_{\underline{s}}^{s} \omega(s - t)\psi(n - 1)\frac{t^{n-2}}{s^{n-1}}dt \\
&= (A - \omega\lambda_d - \rho)\underline{s}^{n-1}\frac{1}{s^{n-1}} \\
&\quad + \omega\psi\frac{1}{2}\left[(n - 1)\underline{s}^n + 2s\underline{s}^{n-1} - n\underline{s}^{n-1}\right]\frac{1}{s^{n-1}} \\
&\quad + \omega\psi\left[\frac{1}{n}s^n - s\underline{s}^{n-1} + \frac{n-1}{n}\underline{s}^n\right]\frac{1}{s^{n-1}}
\end{aligned}
\tag{64}
$$

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 $s_{n:n} \in [\underline{s}, 1]$ and dividing by n to obtain the profits

for a single lender, modulo the cost of information acquisition, results in

$$\pi_L = \frac{1}{n} \int_{\underline{s}}^{1} \pi_L(s) n s^{n-1} ds$$

$$= (A - \omega \lambda_d - \rho) \underline{s}^{n-1}(1 - \underline{s})$$

$$+ \omega \psi \frac{1}{2} \left[(-n+1)\underline{s}^{n-1} + (2n-1)\underline{s}^n - n\underline{s}^{n+1} \right]$$

$$+ \omega \psi \left[\frac{1}{n(n+1)} - \frac{1}{2}\underline{s}^{n-1} + \frac{n-1}{n}\underline{s}^n + \frac{-n+1}{2(n+1)}\underline{s}^{n+1} \right] \tag{65}$$

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

E.3 Calculation for equation (18)

This section computes consumer surplus.

First, observe that a consumer only achieves a positive surplus if $s_{n-1:n} \geq \underline{s}$. In particular, if $s_{n-1:n} < \underline{s}$ then either the consumer doesn't obtain a loan and receives the outside option of zero (if $s_{n:n} < \underline{s}$) or it obtains a loan with an interest rate equal to its willingness to pay and therefore also achieves a net surplus of zero (if $s_{n:n} \geq \underline{s}$). Therefore, the consumer surplus is given by

$$\pi_C = E_{s_{n-1:n}} \left[\{ A - \underline{R}(D(s_{n-1:n}, s_{n-1:n}; \psi, n)) \} 1_{\{s_{n-1:n} \geq \underline{s}\}} \right]$$

$$= E_{s_{n-1:n}} \left[\left\{ A - \left(\omega \left[\lambda_d + \frac{1}{2}(n - (n+2)s_{n-1:n}) \psi \right] + \rho \right) \right\} 1_{\{s_{n-1:n} \geq \underline{s}\}} \right]$$

$$= (A - \omega \lambda_d - \rho) E_{s_{n-1:n}} \left[1_{\{s_{n-1:n} \geq \underline{s}\}} \right]$$

$$- E_{s_{n-1:n}} \left[\frac{1}{2}\omega (n - (n+2)s_{n-1:n}) \psi 1_{\{s_{n-1:n} \geq \underline{s}\}} \right] \tag{66}$$

Recall that the predictive distribution of the signal is uniform. Therefore, the conditional pdf for $n-1$ order statistic among the n signals is given by $f(s_{n-1:n} = t) = n(n-1)(1-t)t^{n-2}$. Then computing the expectations determines

$$\pi_C = (A - \omega \lambda_d - \rho)(1 - n\underline{s}^{n-1} + (n-1)\underline{s}^n)$$

$$- \frac{1}{2}\omega \psi \left[\frac{2}{n+1} - n^2 \underline{s}^{n-1} + 2(n-1)(n+1)\underline{s}^n - \frac{n(n-1)(n+2)}{n+1}\underline{s}^{n+1} \right] \tag{67}$$

F   Supplemental material for the baseline model simulation (Section 4)

Figure F.1: Association between interest rate and default

Figure F.1a shows the association between the average interest rate and the default rate as functions of λ_d in the model of 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 F.1b 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
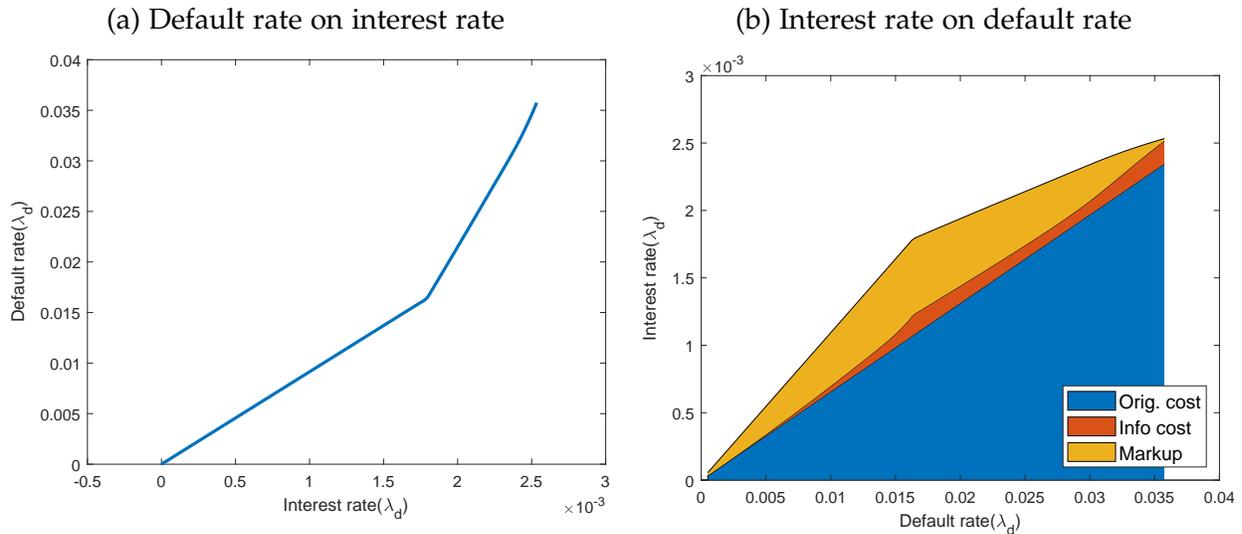
Figure F.2: Number of offers

This figure shows the probability that a borrower receives 2, 1, or 0 offers in the model of 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).
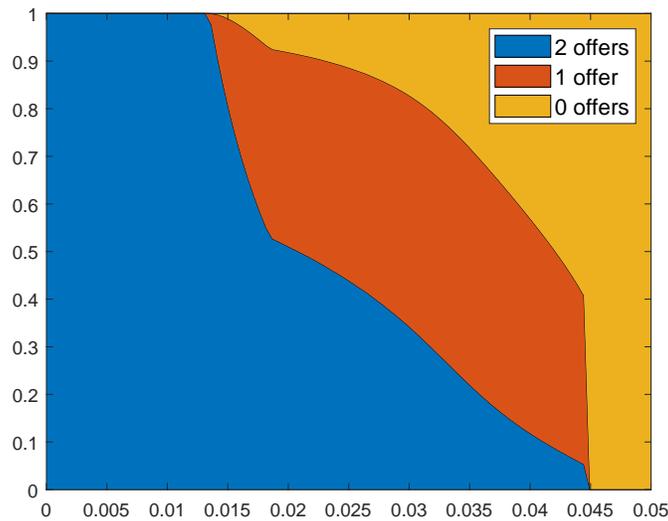
Figure F.3: Probability of a single offer

This figure shows the probability that a borrower receives exactly 1 offer conditional on receiving an offer for the baseline model of 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 ω and a high loss given default.
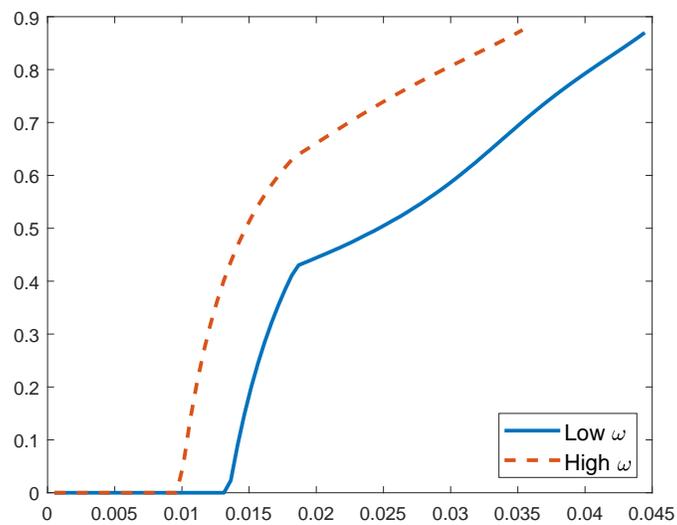
F.1 Endogenous versus exogenous screening

To illustrate the role of investment in screening technology, Figure F.4 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 λ_d.[58] As λ_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.

---

[58]Note that, in the case where the information level is exogenous, the outcomes are only computed for λ_d sufficiently large due to the constraint $\psi \leq 2\lambda_d$ from Section 3.3.1.

Figure F.4: Active intermediation with endogenous and exogenous information level

These figures show various features of the model in the case where the information level ψ 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 ψ 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).
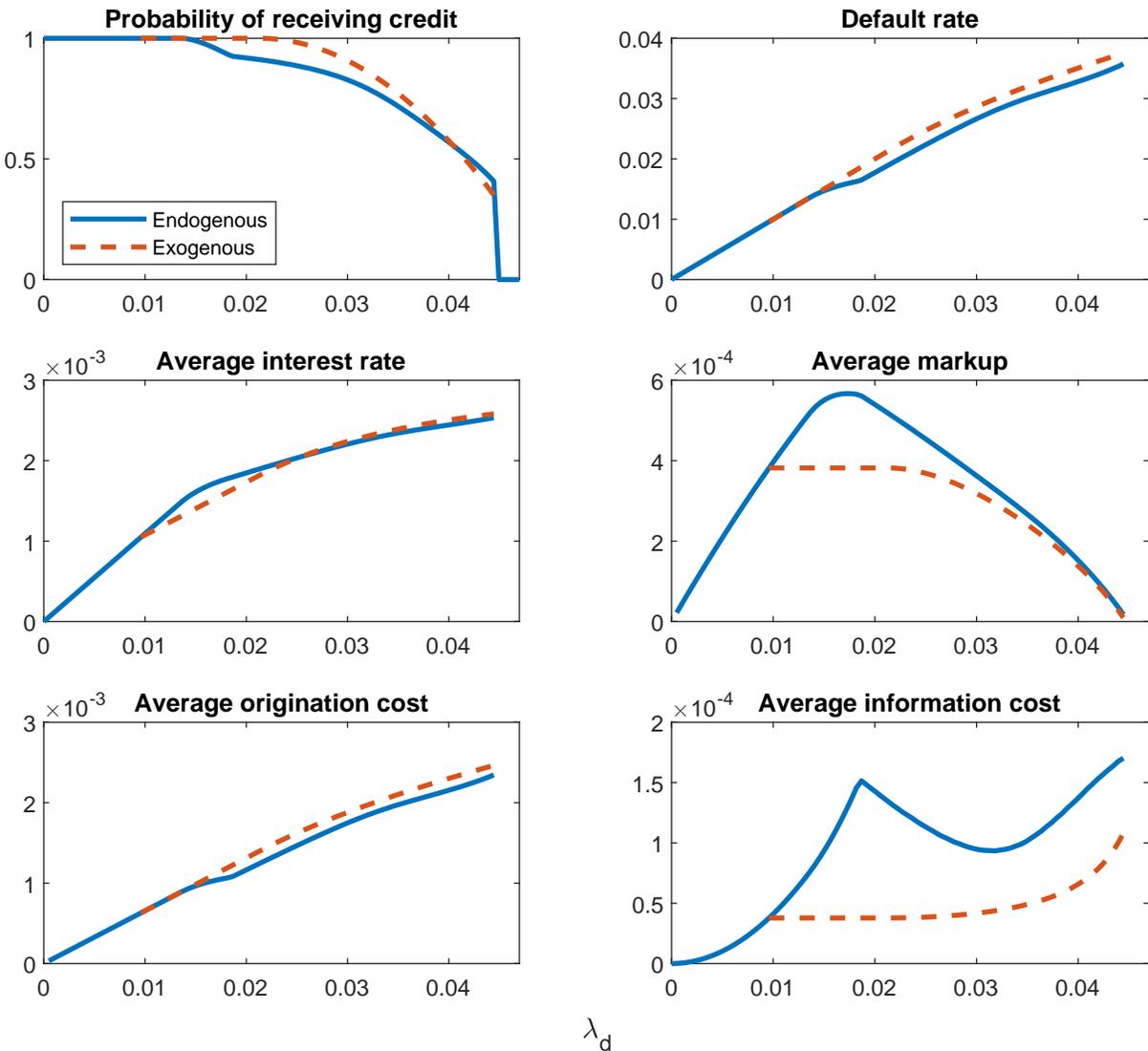
G Supplemental material for the model simulation with heterogeneous lenders (Section 5)

G.1 Supplemental material for heterogeneous ψ (Section 5.1)

Figure G.1: Probability of credit outcomes with heterogeneous ψ

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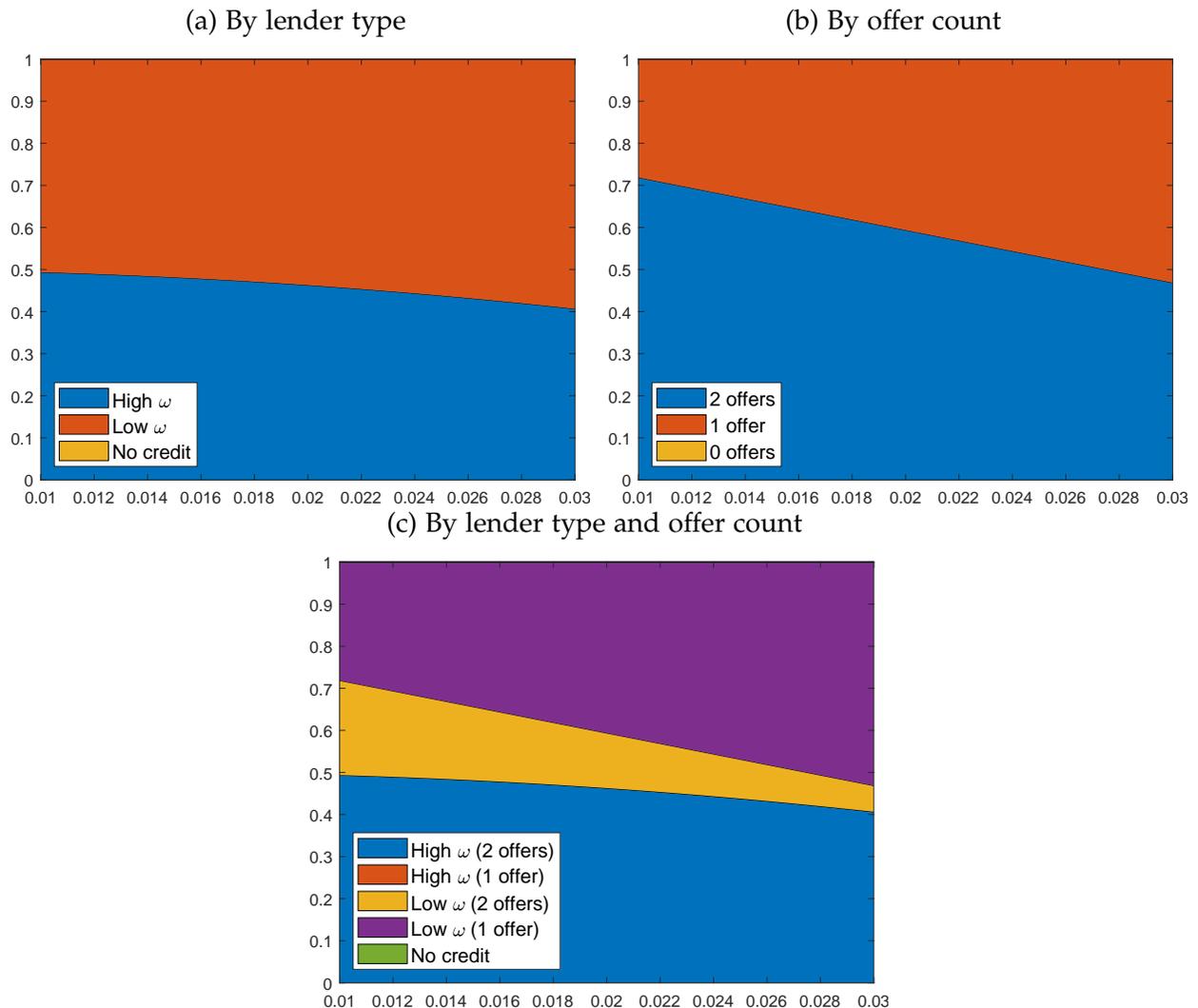

(a) By lender type

(b) By offer count

(c) By lender type and offer count
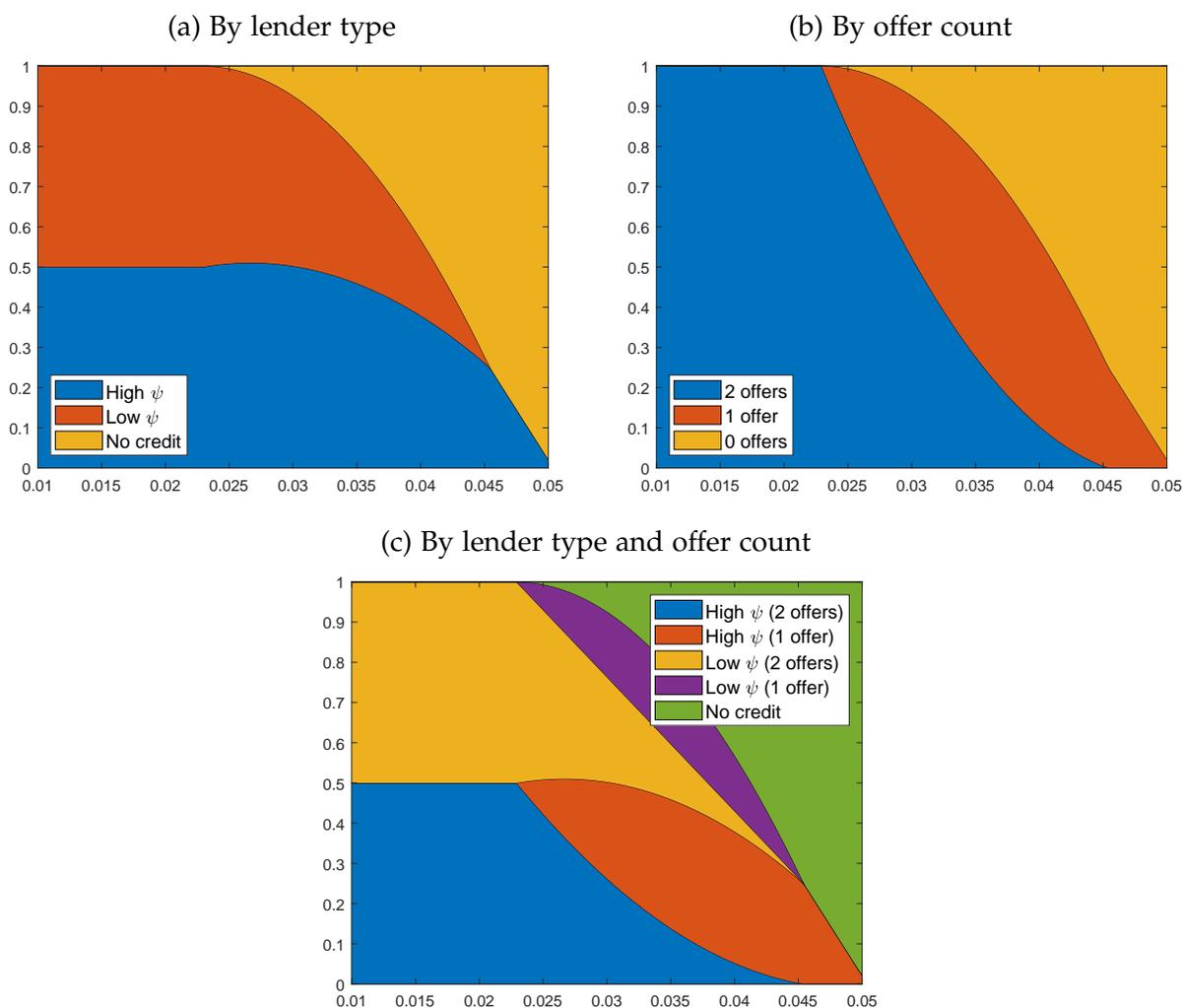
Figure G.2: Association between interest rate and default with heterogeneous ψ

This figure shows the association between the average interest rate and the default rate as functions of λ_d in the version of the model with 2 lenders with exogenous and different information levels ψ (described in Section 5.1).
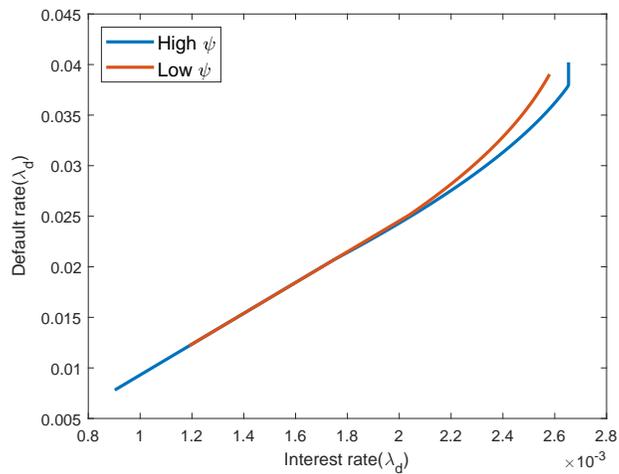
G.2 Supplemental material for heterogeneous ω (Section 5.2)

Figure G.3: Probability of credit outcomes with heterogeneous ω

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 ψ but different loss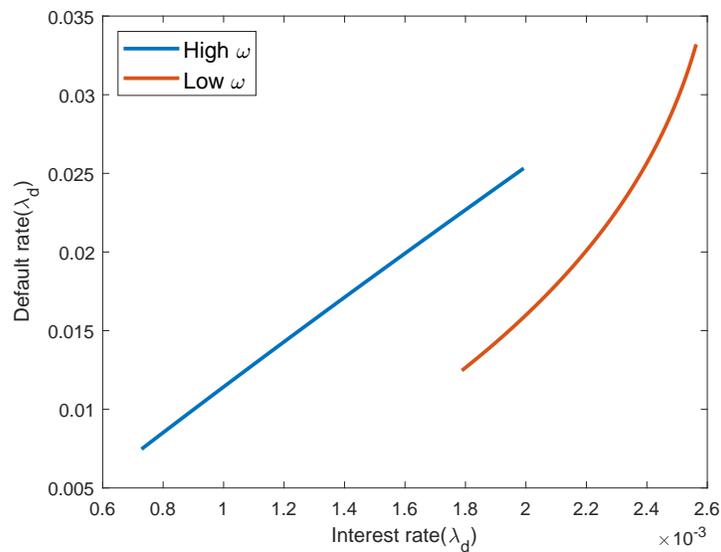es given default ω (described in Section 5.2). Figure G.3a shows the probability that a borrower receives credit from the lender with high ψ, receives credit from the lender with low ψ, and does not receive credit. Figure G.3b shows the probability that a borrower receives 2, 1, or 0 offers. Figure G.3c 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 G.4: Association between interest rate and default with heterogeneous ω

This figure shows the association between the average interest rate and the default rate across λ_d in the version of the model with 2 lenders with exogenous and the same information levels ψ but different losses given default ω (described in Section 5.2).

Figure G.5: Active intermediation with heterogeneous and homogeneous ω

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, ω_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 offer | receiving credit)* is the probability that the consumer receives only one offer conditional on receiving an offer. Parameters: $\rho = 0$, $n = 2$, $A = .0027$, $\omega = .066$ and $.04$ for the heterogeneous case (or just $\omega = .066$ for the homogeneous case), $\mu = .1$, $\psi = .02$.
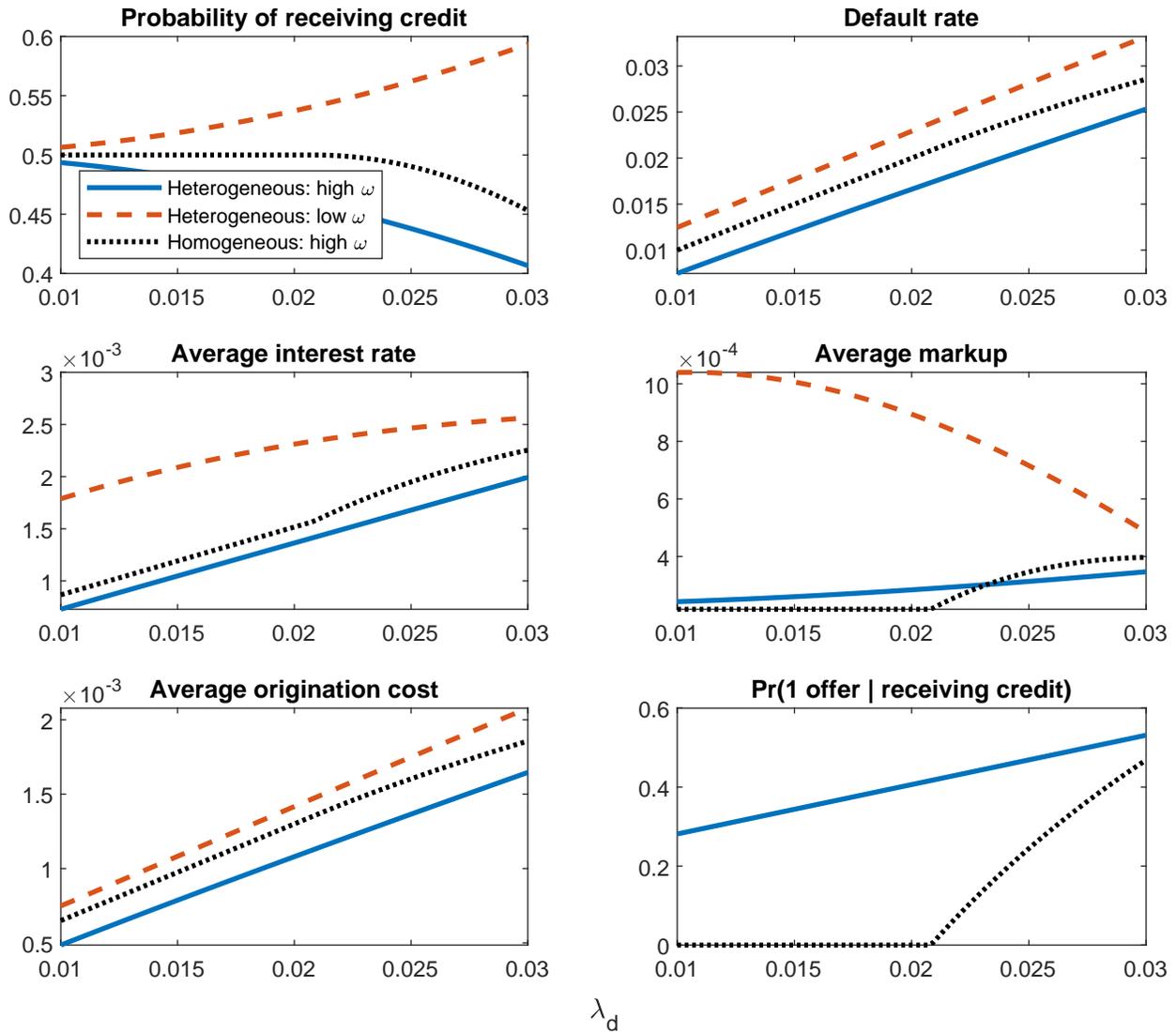
Figure G.6: Participation thresholds with heterogeneous ω

This figure shows the participation thresholds s_i (as in equation (33)) across λ_d in the version of the model with 2 lenders with exogenous ψ and different losses given default ω (described in Section 5.2).
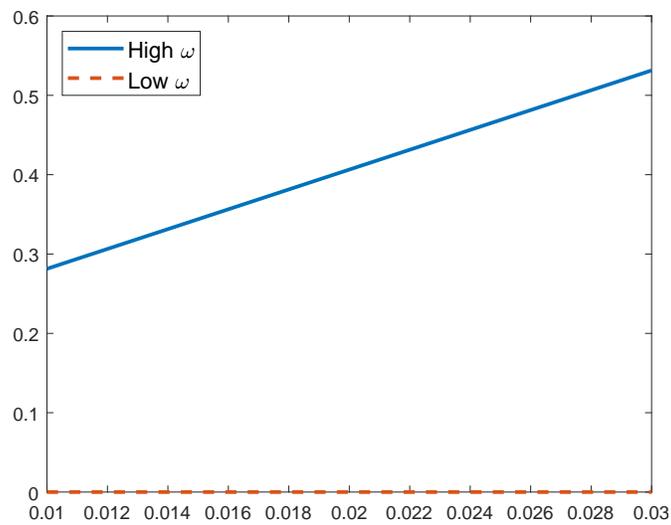
Figure G.7: Active intermediation as ω varies while also adjusting ρ

These figures show various features of the model for a low loss given default ω and a high loss given default. The cost of funding varies with the loss given default according to equation 27. 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 ψ 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: $\omega = .066$ or $.09$, $\rho = \lambda_b * (.09 - .066)$ for $\omega = .066$ or $\rho = 0$ for $\omega = .09$, $n = 2$, $A = .0027$, $\mu = .1$.