How Subprime Borrowers and Mortgage Brokers Shared the Pie*

Antje Berndt† Burton Hollifield‡ Patrik Sandås§

October 2013

Abstract

We develop an equilibrium model for origination fees charged by mortgage brokers and show how the equilibrium distribution of fees depends on borrowers' valuation for their loans and on their level of informedness. We use non-crossing quantile regressions and data from, formerly, one of the largest subprime lenders to estimate conditional broker fee distributions and—subject to the level of informedness in the borrower population—the implied distributions of borrower valuations. Our findings are consistent with higher conditional fees, and hence higher valuations or a lower level of informedness, for larger borrowers and for borrowers in more racially diverse neighborhoods. While fully informed borrowers pocket the spread between the value they assign to the loan and the broker's reservation value for the fees, uninformed borrowers are left empty-handed. For partially informed borrowers we quantify the fraction of the overall surplus from the mortgage that goes to the broker, and show how that fraction decreases as the borrower becomes marginally more informed.

JEL Classifications: G21

Keywords: Mortgage broker compensation; Borrower valuation; Borrower informedness

*We are grateful for financial support from the McIntire Center for Financial Innovation. We thank Michael Gage of IPRecovery for help with the New Century database.

†Associate Professor of Finance, Poole College of Management, North Carolina State University, Raleigh, NC, 27695. Phone: 919-515-4539. Email: aberndt@ncsu.edu.

‡PNC Professor of Finance, Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, 15213. Phone: 412-268-6505. Email: burtonh@andrew.cmu.edu.

§Associate Professor of Finance, McIntire School of Commerce, University of Virginia, Charlottesville, VA, 22904. Phone: 434-243-2289. Email: patriks@virginia.edu.

1. Introduction

Borrowers in the residential mortgage market take out a loan to purchase a new home or to refinance an existing mortgage. The value borrowers assign to a mortgage is the value of the benefits they expect to draw from owning the home in excess of the expected present value of the mortgage payments and the outside option of not obtaining the loan. If the loan is originated through an independent mortgage broker, the broker charges the borrower a fee for the origination services provided. Neither the borrower's valuation for the loan nor the broker's reservation value for the fees can be observed directly. The goal of this paper is to quantify these values as well as the fraction of the spread between them that the broker is able to capture through the fees.

We focus on subprime borrowers and use a simple model of bargaining between the borrower and broker where the broker learns the borrower's reservation value for the fees and has all the bargaining power. The broker sets her fees equal to the borrower's reservation value, meaning that the surplus associated with the interaction between borrower and broker goes entirely to the broker. Although the borrower's net benefit from his interaction with the broker is zero, it does not necessarily imply that the broker can capture all of the value that the borrower assigns to the mortgage. Indeed, if the borrower's reservation value for the fees is lower than his valuation for the underlying house purchase or mortgage refinance, then the borrower's surplus from taking out the loan is positive. The overall surplus from the mortgage is equal to the amount by which the borrower's valuation for the loan exceeds the broker's reservation value for the fees.

To quantify the borrower's valuation for the loan, the broker's reservation value for the fees, and the fraction of the overall surplus from the mortgage that goes to the broker, we propose a model where borrowers form their reservation value for the fees in one of three ways: (i) *Fully informed borrowers* learn the broker's reservation value, possibly by shopping from different brokers, and set their reservation value for the fees equal to the broker's; (ii) *Partially informed borrowers* consult with a number of friends about the fees they paid in a similar recent mortgage transaction and set their reservation value equal to the minimum of the fees their friends paid, as long as that does not exceed their valuation for the loan; (iii) *Uninformed borrowers* do not consult with friends or shop from different brokers, and set their reservation value equal to their valuation for the loan. We assume that only a small fraction of the borrower population is fully informed, and that most borrowers are either partially informed or uninformed.

The notion of borrowers consulting with friends goes back to Woodward and Hall (2009), who distinguish between borrowers who are fully informed and borrowers who consult with friends. In their paper, borrowers who consult with friends also gather information from the broker and set their reservation value equal to the minimum of the fees their friends paid plus some upward effect that the broker has on the reservation value. Woodward and Hall do not cap the borrower's reservation value at his valuation for the loan. Moreover, they do not consider a scenario where the borrower neither learns the broker's reservation value nor consults with friends about the fees they paid. Allowing for borrowers to be uninformed is particularly important in light of the survey evidence in Lacko and Pappalardo (2007) and ICF Macro (2009), which shows that a large fraction of the borrower population does not actively shop for comparison quotes.

We are interested in the equilibrium distribution of borrower reservation values which is given when the borrower reservation value distribution coincides with the broker fee distribution. We characterize the equilibrium distribution of borrower reservation values—and hence of broker fees—and describe how it depends on borrowers' valuation for their loans and on their level of informedness. We invert the relationship between broker fees and borrower valuations and informedness to express the distribution of borrower valuations as a function of the equilibrium distribution of fees, subject to the level of informedness in the borrower population. We show that as the fraction of uninformed borrowers increases or the number of friends that partially informed borrowers consult with decreases, the implied distribution of borrower valuations shifts towards smaller values.

Estimates for the broker fee distribution, as a function of loan and borrower characteristics, are obtained using non-crossing quantile regressions. Classical quantile regression analysis estimates each quantile function separately. The resulting regression functions may cross in finite samples, which is particularly problematic for our purposes since it would prevent us from obtaining a well-defined implied distribution of borrower valuations. Non-crossing quantile regression analysis, on the other hand, estimates the quantile functions simultaneously and under an explicit non-crossing constraint. Our data contain detailed records of more than 70,000 brokered loans originated in California, Florida, New York or Illinois between 2004 and 2006 and funded by New Century Financial Corporation. In the years prior to the subprime crisis, New Century was one for the top three subprime lenders in the U.S., with a loan pool that was representative of the broader subprime market.

Equipped with estimates of the conditional broker fee distribution, we find that—for 2/28 purchase loans with full documentation originated in California and a benchmark specification of other conditioning variables—expected fees are about \$3.6K per loan. If three quarters of the borrower population are uninformed and 50% or even 100% of these uninformed borrowers became partially informed, expected fees would decrease by \$400 or \$800 per loan, respectively. All else the same, expected fees are estimated to be higher for larger loans than for smaller loans, and for loans originated in more diverse neighborhoods with a lower fraction of white population than for loans originated in less diverse neighborhoods with a higher fraction of white population. Higher fees for borrowers that take out larger loans or loans in more diverse neighborhoods point to higher valuations or a lower level of informedness for such borrowers.

Depending on the borrower's level of informedness, fees range from the broker's reservation value for fully informed borrowers to the borrower's valuation for the loan for uninformed borrowers. We set the broker's reservation value for the fees equal to a low quantile of the conditional fee distribution and—for benchmark 2/28 purchase loans with full documentation originated in California—estimate it to be \$700. Assuming that three quarters of the borrower population are uninformed and that partially informed borrowers consult with one friend, the expected value of borrower valuations is \$4.0K.

We are interested in the expected fraction of the overall surplus from the mortgage that goes to the broker. While fully informed borrowers capture the entire spread between the borrower's valuation for the loan and the broker's reservation value for the fees, uninformed borrowers are left empty-handed. Partially informed borrowers, on the other hand, split the overall surplus from the mortgage with the broker. Consider benchmark 2/28 purchase loans with full documentation originated in California and assume that partially informed borrowers consult with one friend. If three quarters of the borrower population are uninformed, any given partially informed borrower leaves, in expectation, 75% of the overall surplus from the mortgage on the table. If only one quarter of the borrower population is uninformed, partially informed borrowers are expected to leave 66% of the overall surplus to the broker. All else the same, if a partially informed borrower were to consult with two rather than just one friend, these estimates would drop by 13-14 percentage points.

Our results are robust across different strata of the data. Independent of the type of the loan, the purpose of the loan, the documentation level and which state the loan is originated in, we find that partially informed borrowers that consult with just one friend leave a large fraction of the overall surplus from the mortgage on the table. Another robust feature of the data is that expected broker fees are higher for larger than for smaller loans, and for loans originated in more diverse rather than less diverse neighborhoods. In addition, we confirm the validity of our empirical findings for a wide range of co-dependence between the borrower's valuation for the loan and his level of informedness.

2. Equilibrium Model of Broker Fees

We develop a model of the mortgage origination process to understand how broker origination fees are determined and what they may reveal about the borrower's valuation for the loan. We focus on loans originated in the wholesale market, where independent mortgage brokers act as financial intermediaries matching borrowers with lenders. Brokers assist borrowers in the selection of the loan and in completing the loan application, and provide services to wholesale lenders by generating business and helping them complete the paperwork.

Consider a borrower who arrives at a broker requesting a mortgage.[1] The broker evaluates

[1] The borrower is matched with the broker either by chance, following a recommendation of a real estate

the borrower's and the property's characteristics and, based on that information, provides the borrower with one or more financing options. A financing option is a specification of the terms of the loan, such as the loan type, loan purpose, level of income documentation, loan amount and mortgage rate. It also outlines the fees the broker will charge the borrower.

To compile the list of financing options, the broker reviews wholesale rate sheets distributed by potential lenders. These rate sheets state the minimum rate at which a given lender is willing to finance a loan, as a function of loan and borrower characteristics. We refer to this rate as the lender's base rate. Rate sheets also inform the broker about the yield spread premium (YSP), if any, that the lender pays to the broker for originating the loan at a rate higher than the base rate. The borrower and the broker bargain over the terms of the loan and the fees. Once they reach an agreement, the broker submits a funding request to one or more lenders. The lender reviews the application material and responds with a decision to fund the loan or not.[2] If the loan is funded, the broker receives the fees and YSP at the loan closing.

The borrower's valuation for a loan—that is, his perceived benefit from taking out a loan—is the dollar value of the benefits the borrower expects to draw from purchasing the house or refinancing the existing mortgage in excess of the expected present value of the mortgage payments and the perceived dollar value of the outside option of not obtaining the loan. Clearly, the borrower will not agree to pay a broker fee f that is higher than his valuation for the loan. Hence the borrower's reservation value for the fees, $\overline{f}$, cannot exceed ν: $\overline{f} \leq \nu$. And if the borrower perceives to have outside options other than not obtaining the mortgage, perhaps by shopping from other brokers or consulting with recent mortgagors about the fees they paid, then the borrower's reservation value for the fees may be strictly lower than his valuation for the loan: $\overline{f} < \nu$. In any case, the borrower's net benefit from interacting with the broker is $\overline{f} - f$.

---

broker or someone else, or as a result of marketing efforts by the broker. We do not model borrower-broker interactions prior to the time that a deal is made.

[2]We suppose that a lender will fund the loan as long as the broker collects and transfers the requested application materials and secures a rate at or above the lender's base rate. Since the broker is paid only if the loan is made, she will only offer fundable proposals to the borrower and will ensure that the application materials are presented to the lender in a timely fashion.

Let y denote the YSP paid by the lender and c denote the broker's cost of originating the loan. Broker costs are the costs the broker expects to incur between the time she strikes a deal with the borrower and the time the loan closes. They include the broker's time costs of dealing with the borrower as well as any administrative costs incurred by the broker for intermediating the mortgage. The broker's reservation value for the fees, $\underline{f}$, is equal to

$$\underline{f} \;=\; c - y, \tag{1}$$

and the broker's net benefit from originating the loan is $f - \underline{f}$. The borrower's and broker's joint gains from trade is the sum of their respective benefits, $(\overline{f} - f) + (f - \underline{f}) = \overline{f} - \underline{f}$.

We consider a simple model of bargaining between the borrower and broker where the broker learns the borrower's reservation value $\overline{f}$ and has all the bargaining power. Then the broker sets the fee equal to the borrower's reservation value,

$$f \;=\; \overline{f},$$

and, as long as $\underline{f} \leq \overline{f}$, a deal is made. The surplus associated with the interaction between borrower and broker goes entirely to the broker. The broker sets the terms of the loan so as to maximize the joint gains from trade, subject only to the constraint $\underline{f} \leq \overline{f}$.

Even though the borrower's net benefit from his interaction with the broker is zero, it does not necessarily imply that the broker can capture all of the value ν that the borrower assigns to the mortgage. Indeed, the borrower's surplus from obtaining the mortgage, $\nu - f = \nu - \overline{f}$, is positive as long as the borrower's valuation for the loan exceeds his reservation value for the fees. The overall surplus from the mortgage is defined as the sum of the borrower's surplus from obtaining the mortgage and the broker's surplus from originating the mortgage. It is equal to the dollar amount by which the borrower's valuation for the loan exceeds the broker's reservation value for the fees: $(\nu - \overline{f}) + (\overline{f} - \underline{f}) = \nu - \underline{f}$. The fraction of the overall surplus from the mortgage that goes to the broker is

$$\gamma \;=\; \frac{\overline{f} - \underline{f}}{\nu - \underline{f}}, \tag{2}$$

and the fraction that goes to the borrower is $1 - \gamma$.

Our goal is to quantify ν, $\underline{f}$ and γ. To do so we propose a model for how borrowers form their reservation values $\overline{f}$.[3] In particular, we assume that borrowers fall into one of three categories:

- **Fully informed borrower:** The borrower is an expert in the mortgage market. He learns the broker's reservation value for the fees, possibly by shopping from different brokers, and sets his reservation value $\overline{f}$ equal to $\underline{f}$.
- **Partially informed borrower:** The borrower consults N friends who are entirely honest about the broker fees f_i they paid in a similar recent mortgage transaction. He forms his reservation value as $\overline{f} = \min\{f_1, \ldots, f_N, \nu\}$.
- **Uninformed borrower:** The borrower neither consults with friends nor shops from different brokers. His reservation value $\overline{f}$ equals ν.

Going forward, we condition on observable loan and borrower characteristics that include the terms of the loan and YSP, but exclude broker fees. The fraction of borrowers in the borrower population that are fully informed, partially informed and uninformed is denoted by π_0, π_1 and π_2, respectively. We are mainly interested in scenarios where only a small fraction of the borrower population is fully informed, and where most borrowers are either partially informed or uninformed.[4] We assume that the friends of a partially informed borrower are random draws from the population of recent borrowers with the same characteristics, and

[3] Our model for the borrower reservation value $\overline{f}$ is an extension of the framework in Woodward and Hall (2009), which does not allow for uniformed borrowers and does not ensure that the reservation value of borrowers that consult with friends is capped at ν.

[4] Two borrower surveys, Lacko and Pappalardo (2007) and ICF Macro (2009), find that many borrowers shop from just one broker and that only few borrowers shop from a large number of brokers. This motivates us to keep π_0 small. And although these surveys do not directly poll borrowers about the number of friends they consult with, they are consistent with a substantial fraction of borrowers being uninformed.

that no change in mortgage-market conditions has occurred since they borrowed. We also assume that the borrower's valuation for the loan is independent of his level of informedness.[5] In addition, we abstract from any unobserved heterogeneity in broker costs c.

Suppose that the distribution of borrower valuations ν across the population of borrowers has support on $[\underline{f}, \overline{\nu}]$, for some $\overline{\nu} > \underline{f}$. Using F_v to denote the cumulative distribution function of a random variable v, we have

$$\begin{aligned} F_{\overline{f}}(a) &= \pi_0 + \pi_1 Prob(\min\{f_1, \ldots, f_N, \nu\} \leq a) + \pi_2 F_\nu(a) \\ &= \pi_0 + \pi_1 \left(1 - Prob(\nu > a) \prod_{i=1}^{N} Prob(f_i > a) \right) + \pi_2 F_\nu(a) \\ &= \pi_0 + \pi_1 \left\{ 1 - (1 - F_\nu(a))(1 - F_f(a))^N \right\} + \pi_2 F_\nu(a), \end{aligned} \tag{3}$$

for any $a \in [\underline{f}, \overline{\nu}]$.

We are interested in the equilibrium distribution of borrower reservation values or, equivalently, the equilibrium distribution of broker fees. An equilibrium distribution is one where $F_{\overline{f}} = F_f$. In equilibrium, Equation (3) can be rewritten as

$$F_f(a) + \kappa_1(a)(1 - F_f(a))^N - \kappa_0(a) \;=\; 0, \tag{4}$$

where $a \in [\underline{f}, \overline{\nu}]$, $\kappa_1(a) = \pi_1(1 - F_\nu(a))$ and $\kappa_0(a) = \pi_0 + \pi_1 + \pi_2 F_\nu(a)$. In Appendix A we prove the following proposition:

Proposition 1 *Let F_ν be some distribution function with support on $[\underline{f}, \overline{\nu}]$. For $N \geq 1$, $\pi_0, \pi_1, \pi_2 > 0$ and $\pi_1 N \leq 1$, the equilibrium distribution of broker fees is well-defined by Equation (4), has support on $[\underline{f}, \overline{\nu}]$ and is unique.*

Figure 1 shows the equilibrium distribution of broker fees implied by a flat distribution of borrower valuations. Proposition 2 states that independent of $\pi = (\pi_0, \pi_1, \pi_2)$, the distribution of f lies to the left of that of ν.

[5]This restriction is lifted in Section 6.2.

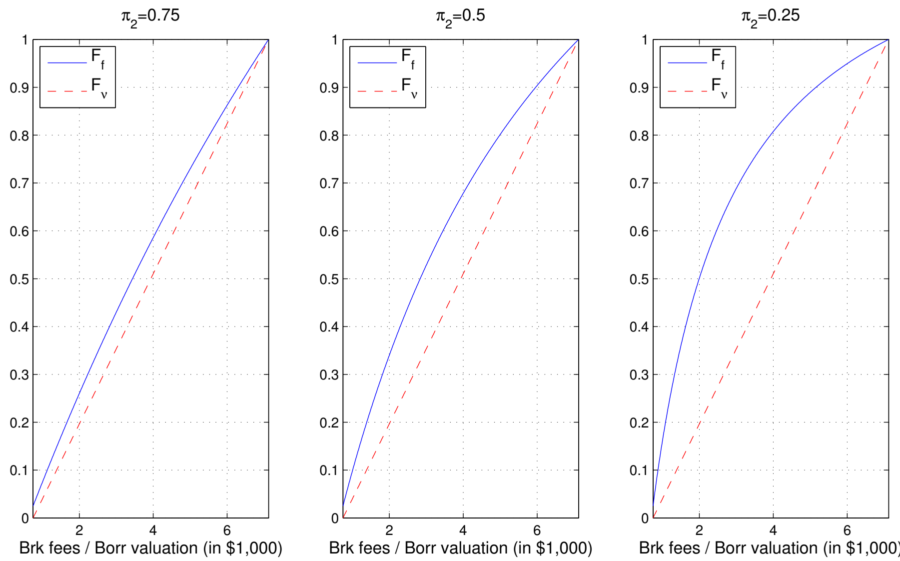

Figure 1: **Equilibrium distribution of broker fees** The figure shows the equilibrium distribution of broker fees implied by Equation (4), assuming a uniform distribution for ν and $N = 1$. Given a choice for π_2, π_0 and π_1 are chosen such that $\pi_0 + \pi_1 + \pi_2 = 1$ and $F_f(\underline{f}) = 0.025$. The left plot is for $\pi = (0.02, 0.23, 0.75)$, the middle plot for $\pi = (0.01, 0.49, 0.50)$, and the right plot for $\pi = (0.01, 0.74, 0.25)$.

Proposition 2 *For $N \geq 1$, $\pi_0, \pi_1, \pi_2 > 0$ and $\pi_1 N \leq 1$, the equilibrium distribution of broker fees satisfies $F_f(a) \geq F_\nu(a)$ for all $a \in [\underline{f}, \overline{\nu}]$.*

Since we observe broker fees but not borrower valuations, we invert Equation (4) and express F_ν as a function of the equilibrium distribution F_f:

$$F_\nu(a) \quad = \quad 1 - \frac{1 - F_f(a)}{\pi_1(1 - F_f(a))^N + \pi_2}, \qquad \text{for } a \in [\underline{f}, \overline{\nu}], \tag{5}$$

and $F_\nu(a) = 0$ for $a < \underline{f}$. Proposition 3 states the conditions under which F_ν defined in Equation (5) is a well-behaved distribution function:

Proposition 3 *Let F_f be some distribution function with support on $[\underline{f}, \overline{\nu}]$. For $N \geq 1$, $\pi_0, \pi_1, \pi_2 > 0$ and $\pi_1 N \leq 1$, F_ν given by Equation (5) is a well-defined distribution function as long as $F_f(\underline{f}) \geq x$, where $x \in (0, 1)$ solves $x + \pi_1(1 - x)^N - (1 - \pi_2) = 0$. F_ν has support on $[\underline{f}, \overline{\nu}]$ and is unique.*

Next we describe our data and discuss our approach to estimating the broker fee distribution F_f as a function of observable loan and borrower characteristics. Given F_f, we obtain estimates for F_ν (via Equation (5)) and for $\underline{f}$. Equipped with F_f, F_ν and $\underline{f}$, we sample from the fee and borrower valuation distributions to compute what fraction of the overall surplus from the mortgage is expected to go to the broker.

3. Our Data

Our dataset is obtained from IPRecovery, Inc. and contains detailed records of all broker-originated loans funded by New Century Financial Corporation. New Century was founded in the mid-1990s in California and quickly grew into a nationwide lender that originated, retained, sold and serviced residential mortgages designed for subprime borrowers. Berndt, Hollifield, and Sandas (2012) compare New Century's origination, broker compensation and loan performance statistics to those for the overall subprime market. They document that prior to the subprime crisis, the New Century loan pool was representative of the broader subprime market.[6] We therefore believe that our empirical analysis is not simply a case study of New Century, and that our findings shed light on the pre-crisis subprime market in general.

Table 1 defines the variables used in our empirical analysis and Table 2 reports descriptive statistics for them. We focus on free-standing first-lien loans for single-unit primary residences in California, Florida, New York or Illinois.[7] The loan types considered are 30-year fixed-rate mortgages (FRMs) and 2/28 loans. The latter are 30-year hybrid loans that have a two-year fixed interest rate period after which the interest rate on the mortgage begins to float based on LIBOR plus a margin. The sample is comprised of more than 70,000 loans.

About 19% of the loans in our sample are purchase loans as compared to refinance

[6] Following its bankruptcy filing in 2007, New Century received widespread attention in the popular press, mainly because it was the largest subprime lender to default by that date. By 2009, however, virtually all of New Century's main competitors had either declared bankruptcy, had been absorbed into other lenders, or had otherwise unwound their lending activities.

[7] The five states with the largest share of loan originations are California, Florida, Texas, New York and Illinois. Together they account for more than 50% of the brokered loans funded by New Century. Texas is unique in that cash-out refinance mortgages—which account for over 50% of the Texas loans in our sample—are subject to what is called the "Texas Home Equity 3% Fee Cap" which limits broker fees to 3% of the loan amount (see Section 50(a)(6)(E) of the Texas Constitution). Since a binding exogenous constraint on the conditional fee distribution would invalidate the link between broker fees and borrower valuation established in Equation (3), we exclude Texas from our analysis.

Table 1: **List of loan and borrower characteristics**

Variable	Description
Property location	We consider loans originated in California, Florida, New York or Illinois
Loan purpose	Purchase or refinance loan
Loan type	2/28 or 30-year fixed-rate mortgage
Level of documentation	We consider full and low (i.e., limited or stated) documentation loans
Loan amount	Amount borrowed
Loan-to-value ratio (LTV)	Value of the loan divided by that of the house
Prepay penalty	Indicator for a loan with a prepayment penalty
Initial rate	Initial mortgage rate
Rate margin for 2/28 loan	Rate margin that is added to LIBOR to determine the floating rate
NC charges	Upfront charges by New Century as percentage of loan amount
FICO	Fair, Isaac and Company's borrower credit score at origination
Debt-to-income ratio (DTI)	All monthly debt payments divided by monthly gross income
Race	Fraction white population in zip code, based on 2000 census data
Broker competition	For a given month and zip code, broker competition is the number of brokers who submitted loan applications to New Century divided by the number of housing units (in thousands)

Table 2: **Descriptive statistics** The table reports average statistics for brokered loans funded by New Century. Our data include 70,679 free-standing first-lien loans for single-unit primary residences originated in California, Florida, New York or Illinois between 2004 and 2006.

	CA	FL	NY	IL	All
No of loans (×1,000)	35.7	21.4	7.7	6.0	70.7
No of brokers (×1,000)	7.0	4.2	1.1	1.3	13.0
	Loan and borrower characteristics				
Purchase loans (%)	14	23	26	25	19
2/28 loans (%)	66	63	56	83	66
Full documentation (%)	53	57	53	66	55
Loan amount (×$1,000)	258	156	246	176	219
LTV (%)	74	78	78	82	76
Prepay penalty (%)	96	97	48	20	85
Initial rate for 2/28 loan (%)	7.10	7.95	7.59	7.80	7.47
Rate margin for 2/28 loan (%)	5.75	5.86	5.85	5.70	5.78
Initial rate for 30-year FRMs (%)	6.54	7.53	6.95	7.96	6.98
NC charges (% of loan amount)	0.13	0.16	0.12	0.04	0.13
FICO	607	591	610	598	602
DTI (%)	41	40	41	39	40
Race (%)	56	74	69	66	64
Broker competition	1.22	0.66	0.63	0.48	0.93

mortgages. 66% of the loans in our sample are 2/28 loans, and 55% of the loans are full documentation loans. The average loan amount is $219K, with loans originated in California and New York generally being larger than those originated in Florida or Illinois. The average loan-to-value (LTV) ratio is about 76%, the average FICO score is just above 600, and the average debt-to-income (DTI) ratio is 40%. The majority of the loan contracts in California and Florida include a prepayment penalty clause, whereas a large fraction of New York and Illinois loans are made without such a clause.[8] The average percentage of whites in the zip code is 64%. Our broker competition measure—computed as the number of brokers who submit loan applications to New Century in the month and zip code where the loan is originated divided by the number of housing units (in thousands) in the zip code—has a sample average of 0.9.

During our sample period, more than 13,000 different brokerage firms did business with New Century. They earned revenues from two sources: a direct fee paid by the borrower and an indirect fee—the yield spread premium—paid by the lender. Direct fees include all compensation associated with the mortgage origination paid by the borrower directly to the broker, including finance charges such as appraisal and credit report fees. The YSP rewards the broker for originating loans with higher mortgage rates, holding other things equal.[9]

Table 3 shows that total broker revenues per loan, as a percentage of loan amount, average to about 3.0% for the loans in our sample.[10] The majority of the revenues stems from fees, which on average amount to 2.2% of the size of the loan. The remaining 0.8% are due to YSP. Average dollar revenues, fees and YSP per loan are about $6,000, $4,300 and $1.700, respectively. While percentage revenues are generally lower for loans originated in California and New York than for loans originated in Florida, average dollar revenues are higher in

[8] New York Bank Law Section 393 provides that a prepayment penalty may be imposed only during the first year. If prepayment is made within one year, the penalty may not exceed interest for a period of three months on the principal prepaid or interest for the remaining months of the first year on the principal prepaid. Illinois Compiled Statutes Section 5-8 stipulates that brokers may not make a loan with a prepayment penalty unless the broker also offers the borrower a loan without a prepayment penalty and discloses the discount in rate received in consideration for the prepay penalty. If the borrower declines the no-prepay-penalty offer the loan contract may include a prepayment penalty. The penalty may not extend beyond three years or the first rate-adjustment date, and is capped at 3%, 2% or 1% depending if less than one, two or three years have passed since origination.

[9] New loan originator compensation rules went in effect April 1, 2011 as part of the Federal Reserve Board's Regulation Z. They prohibit mortgage broker compensation to vary based on loan terms other than principal. In particular, brokers can no longer receive yield spread premia from the lender.

[10] Ambrose and Conklin (2012) use the New Century data to study a different aspect of mortgage broker compensation. They focus on refinance loans and report the sum of percentage broker revenues and percentage lender fees as 3.8% for their sample.

California and New York due to the larger size of the loans in those two states.

Table 3: **Broker charges** The table reports average broker revenues, fees and YSP per loan, by state of origination. Our data include 70,679 stand-along first-lien brokered New Century loans for single-unit primary residences originated in California, Florida, New York or Illinois between 2004 and 2006.

	CA	FL	NY	IL	All	CA	FL	NY	IL	All
	Percent of loan amount					*Dollar per loan (×1,000)*				
Revenue	2.89	3.41	2.70	2.58	3.00	6.860	4.894	6.420	4.201	5.993
Fee	2.17	2.48	2.05	1.54	2.20	5.028	3.484	4.936	2.294	4.319
YSP	0.71	0.92	0.65	1.04	0.80	1.832	1.410	1.484	1.907	1.673

The top panel of Figure 2 shows the unconditional distribution of broker revenues and of its two components.[11] All three distributions are disperse and skewed to the right: some very large fees and yield spread premia were paid out to brokers. We find the right skewness in the revenue distribution to be a robust feature across different strata of our data, although the skewness is less pronounced after conditioning on the loan amount, as can be seen in the bottom panel of Figure 2.

4. Non-Crossing Quantile Regressions

The crucial first step in our empirical analysis is the estimation of the broker fee distribution $F_{f|X}$ as a function of conditioning variables X. A subset of the conditioning variables in X are used to stratify the data.[12] The remaining conditioning variables in X form the state vector x. For each stratum of the data, the qth quantile of the conditional fee distribution $F_{f|X}$ is given by $\beta_0(q) + x\beta(q)$, defined via $Prob(f \leq \beta_0(q) + x\beta(q)|X) = q$. Keep in mind that there is a different set of $\beta_0(q)$ and $\beta(q)$ coefficients for each stratum.

[11] About 29% of the YSP entries in our data are left blank. All else the same, loans with lower FICO scores, lower risk grades and less documentation are more likely to have a missing YSP entry. Such loans usually have high base rates, leaving less room for brokers to convince borrowers to pay rates in excess of the base rate. Moreover, while an increase in YSP is usually associated with a decrease in direct broker fees, we find no statistical significance for a missing-YSP dummy when regressing broker fees on YSP and other observable covariates. With this in mind, we interpret missing-YSP entries as zero YSP, which brings the percentage of zero-YSP loans in our data to 34%. Our findings are robust, however, to excluding missing-YSP loans from the sample.

[12] In our empirical analysis, we stratify the data along the property's location, the loan purpose, the product type and the documentation level.

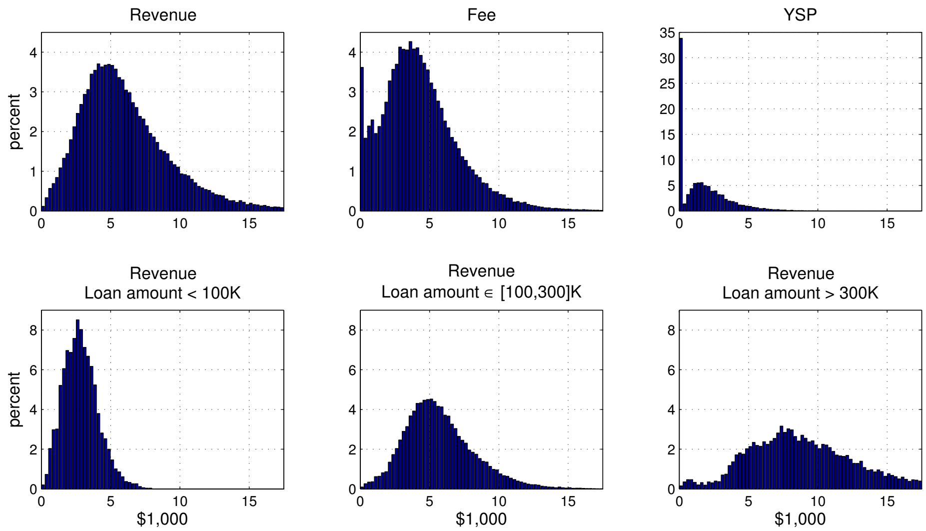

Figure 2: **Broker revenues, fees and YSP** The top panel shows the unconditional distribution of dollar broker revenues, fees and yield spread premia. The bottom panel plots the distribution of dollar broker revenues conditional on the loan amount. Our data include 70,679 free-standing first-lien loans for single-unit primary residences originated in California, Florida, New York or Illinois between 2004 and 2006.

The classical estimator of the regression coefficients for the quantile function is given by

$$\widehat{\beta}(q) \quad = \quad \text{argmin}_{\beta_0,\beta} \left(\sum_{f_i \geq \beta_0 + x_i\beta} q(f_i - \beta_0 - x_i\beta) - \sum_{f_i < \beta_0 + x_i\beta} (1-q)(f_i - \beta_0 - x_i\beta) \right).$$

This optimization problem can be rewritten as a standard linear programming problem:

$$\widehat{\beta}(q) \quad = \quad \text{argmin}_{\beta_0,\beta,u^+,u^-} \left(q\, u^+ + (1-q)\, u^- \right), \tag{6}$$

subject to

$$f - \beta_0 - x\beta \quad = \quad u^+ - u^-, \quad u^+, u^- \geq 0 \text{ and only one is non-zero.} \tag{7}$$

We are interested in many different quantile levels q. A typical quantile regression analysis will solve (6) separately for each of the q. The resulting regression functions may cross in finite samples, meaning that the associated conditional quantile curves are not necessarily monotonically increasing in q. This problem is illustrated in Figure 3, which shows estimated conditional quantile curves for 2/28 purchase loans with full documentation originated in California. We identify specifications for x for which $\beta_0(q)+x\beta(q)$ is monotonically increasing in q, and specifications for which it is not.

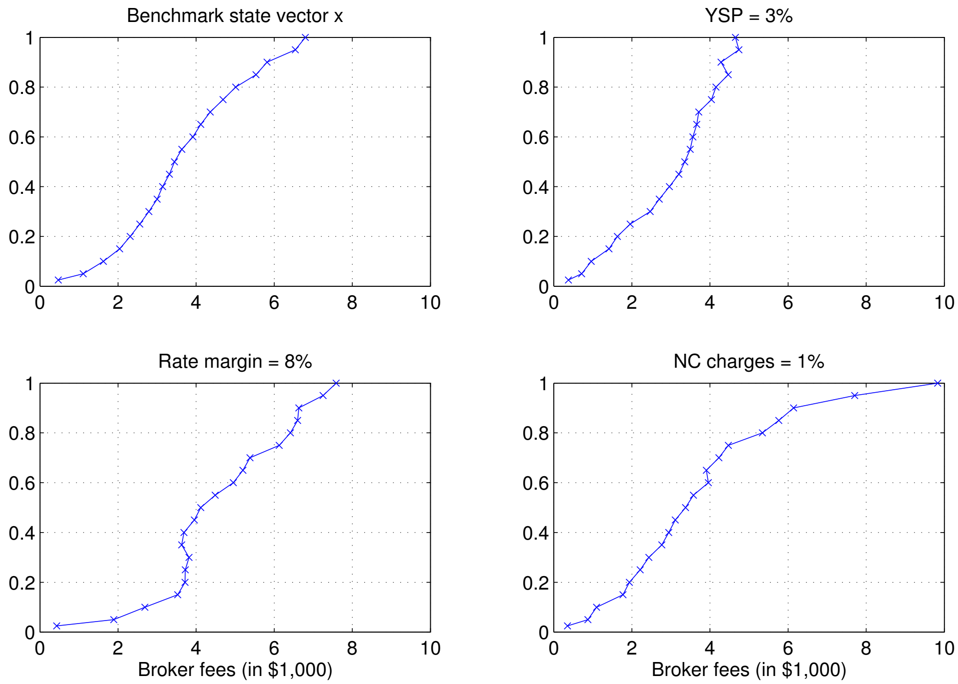

Figure 3: **Conditional quantile curves without non-crossing constraint** The figure shows the results of a classical quantile regression (6), for $q = 0.025, 0.05, 0.1, \ldots, 0.95, 0.975$. The data include all 2/28 purchase loans with full documentation originated in California. The benchmark state vector x sets the loan amount to \$200K, percentage YSP to 1%, LTV to 75%, prepay dummy to one, initial rate to 7.5%, rate margin to 6%, NC charges to zero, FICO score to 600, DTI to 40%, race to 65% and broker competition to 1. In the top right plot, we change the percentage YSP to 3%, in the bottom left plot the rate margin is raised to 8%, and in the bottom right plot NC charges are set to 1%.

Non-monotonicity of F_f is especially problematic since it implies non-monotonicity of F_ν (per Equation (5)), effectively preventing us from obtaining a well-defined empirical distribution for the borrower valuations. We therefore propose to estimate the quantiles of the conditional fee distribution simultaneously under a non-crossing constraint.

Suppose that the state vector consists of m variables, each of which has bounded support, and that $\mathcal{D} \subset \mathcal{R}^m$ is a closed convex polytope such that $x \in \mathcal{D}$ for all x. For $0 < q_1 < \ldots <$

$q_k < 1$, we replace the optimization problem in (6) and (7) with

$$\widehat{\beta} \;=\; \operatorname{argmin}_{\beta_0(q_j),\beta(q_j),u_i^+(q_j),u_i^-(q_j)} \sum_{j=1}^{k}\sum_{i=1}^{n}\left\{q_j\, u_i^+(q_j) + (1-q_j)\, u_i^-(q_j)\right\},$$

where n is the number of observations, subject to

$$f - \beta_0(q_j) - x_i\beta(q_j) \;=\; u_i^+(q_j) - u_i^-(q_j), \quad u_i^+(q_j), u_i^-(q_j) \geq 0,$$

and the non-crossing restriction

$$\beta_0(q_{j-1}) + x\beta(q_{j-1}) \;\leq\; \beta_0(q_j) + x\beta(q_j), \tag{8}$$

for all $x \in \mathcal{D}$ and $j = 2, \ldots, k$.

Since the control variables have bounded support, we can represent $\mathcal{D}$ as the convex hull of 2^m points. It therefore suffices to enforce the non-crossing restriction (8) at each of the 2^m vertices of $\mathcal{D}$. Without loss of generality, suppose that $\mathcal{D} = [0,1]^m$.[13] For $\eta_0(q_1) = \beta_0(q_1)$ and $\eta(q_1) = \beta(q_1)$, and $\eta_0(q_j) = \beta_0(q_j) - \beta_0(q_{j-1})$ and $\eta(q_j) = \beta(q_j) - \beta(q_{j-1})$, the constraint in (8) is equivalent to

$$\eta_0(q_j) + x\eta(q_j) \;\geq\; 0, \quad \text{for all } x \in \mathcal{D} \text{ and } j = 2, \ldots, k. \tag{9}$$

For $\eta(q) = (\eta_1(q), \ldots, \eta_m(q))'$, break each component $\eta_l(q)$ into its positive and negative parts: $\eta_l(q) = \eta_l^+(q) - \eta_l^-(q)$, where both $\eta_l^+(q)$ and $\eta_l^-(q)$ are non-negative and only one is non-zero. Using this parameterization, the "worst case scenario" in (9) obtains for $x^{wc} = (x_1^{wc}, \ldots, x_m^{wc})$, where $x_l^{wc} = 1$ when $\eta_l(q) < 0$ and $x_l^{wc} = 0$ when $\eta_l(q) > 0$. Hence the

[13] As long as each of the control variables in x has bounded support, there exists an invertible affine transformation that maps $\mathcal{D}$ to $[0,1]^m$. The transformation is performed before the estimation, and then transformed back after the estimation, while retaining the non-crossing property.

constraint in (9) becomes

$$\eta_0(q_j) - \sum_{l=1}^{m} \eta_l^-(q_j) \quad \geq \quad 0, \quad \text{for all } j = 2, \ldots, k. \tag{10}$$

As a result, only a total of $k-1$ additional constraints are needed to ensure non-crossing, rather than the $n \times (k-1)$ constraints in (8). After reparameterization, the problem is thus a straightforward linear programming problem, which can be solved via standard software. For our implementation we use the R code provided by Bondell, Reich, and Wang (2010).[14]

5. Empirical Analysis and Results

In our empirical analysis, we first use non-crossing quantile regressions to estimate the distribution of broker fees, $F_{f|X}$, as a function of conditioning variables X. We set $\underline{f}(X)$ and $\overline{\nu}(X)$ equal to the lowest and highest quantile estimated for $F_{f|X}$. For a given specification of X, $\pi = (\pi_0, \pi_1, \pi_2)$ and N, we then use Equation (5) to obtain the implied distribution of borrower valuations, $F_{\nu|X,\pi,N}$. Equipped with estimates for $F_{f|X}$, $\underline{f}(X)$ and $F_{\nu|X,\pi,N}$, we compute (i) expected borrower valuations, brokers' reservation values for the fees and expected fees, (ii) the expected fraction of the overall surplus from the mortgage that goes to the broker, (iii) the expected decrease in fees and γ if borrowers become more informed, and (iv) the ex-post distribution of the borrower's informedness given observed broker fees.

We stratify the data along property location (CA, FL, NY and IL), loan purpose (purchase versus refinance), product type (2/28 versus 30-year FRM), and documentation level (full versus low). For each stratum, we perform a non-crossing quantile regression for quantiles $q \in \{0.025, 0.05, 0.1, \ldots, 0.95, 0.975\}$. The state vector x includes the loan amount, the percentage YSP, the loan-to-value ratio, a prepay-penalty indicator, the initial interest rate, the rate margin for 2/28 loans, NC charges, the borrower's FICO score and debt-to-income ratio, the percentage of whites in the zip code and our broker competition measure. Table 4 describes the specifications of x that we will use to illustrate our findings. The vector of

[14] As noted by Bondell, Reich, and Wang (2010), the linear program is extremely sparse and thus the use of a sparse matrix representation via the SparseM package of Koenker and Ng (2003) is recommended.

conditioning variables X is given as

$$X = (\text{property location, loan purpose, product type, documentation level}, x),$$

Table 4: **Specifications of state vector x**

	benchmark (bm)	large loan (large)	small loan (small)	80% white (80% wt)	50% white (50% wt)
Loan amount (in $1,000)	200	250	150	200	200
LTV (%)	75	75	75	75	75
Prepay penalty	1	1	1	1	1
Initial rate for 2/28 loans (%)	7.5	7.5	7.5	7.5	7.5
Rate margin for 2/28 loans (%)	6	6	6	6	6
Initial rate for 30-year FRMs (%)	7.0	7.0	7.0	7.0	7.0
NC charges (%)	0.0	0.0	0.0	0.0	0.0
FICO	600	600	600	600	600
DTI (%)	40	40	40	40	40
Race (%)	65	65	65	80	50
Broker competition	1	1	1	1	1

The results reported in this section are for loans originated in California. In Section 6, the empirical analysis is extended to Florida, New York and Illinois. Table 5 shows the non-crossing quantile regression results for one particular stratum of the data: 2/28 purchase loans with full documentation.[15] We find that the size of the loan and the neighborhood's racial composition have a significant impact on median broker fees. A marginal increase in the loan amount by $100K is associated with an increase in median fees per loan by almost $800. And a marginal decrease in the percentage of white population in the zip code by 10% is associated with an increase in median fees by about $150. The remaining variables that exhibit statistical significance—including the prepay penalty dummy and the rate margin—are closely tied to other conditioning variables.[16] Analyzing their marginal impact on fees is therefore less informative.

When computing $F_{\nu|X,\pi,N}$ via Equation (5), we impose the condition $F_{\nu|X,\pi,N}(\underline{f}(X)) = 0$.

[15]The non-crossing quantile regression results for other strata are qualitatively similar. See also Table 11.
[16]For example, initial rates depend on the rate margin and on whether or not there is a prepay penalty.

Table 5: **Non-crossing quantile regression results** This table reports the results from non-crossing quantile regressions of broker fees (in \$1,000) on the state vector x, for select quantiles. The sample includes 1,355 2/28 purchase loans with full documentation originated in California.

	Quantiles (in percent)				
	2.5	25	50	75	97.5
Constant	-3.413	-2.762	-2.307	-0.942	2.666
	(2.003)	(1.772)	(1.476)	(1.647)	(3.143)
Loan amt (\$100,000)	-0.119	0.306	0.781	0.841	1.683
	(0.049)	(0.099)	(0.054)	(0.075)	(0.230)
YSP (%)	-0.191	-0.218	-0.141	-0.314	-0.713
	(0.110)	(0.124)	(0.105)	(0.126)	(0.189)
LTV (%)	0.011	0.011	0.008	0.008	-0.010
	(0.007)	(0.007)	(0.006)	(0.007)	(0.016)
Prepay penalty	0.787	1.293	0.980	0.789	-0.203
	(0.289)	(0.393)	(0.366)	(0.414)	(1.300)
Initial rate (%)	0.037	0.084	0.087	0.130	0.130
	(0.114)	(0.118)	(0.099)	(0.119)	(0.214)
Rate margin (%)	0.474	0.474	0.474	0.474	0.474
	(0.233)	(0.205)	(0.188)	(0.207)	(0.360)
NC charges (%)	-0.315	-0.315	-0.162	-0.146	4.437
	(0.352)	(0.365)	(0.386)	(0.630)	(2.670)
FICO (in 100's)	-0.011	0.013	0.013	0.013	0.013
	(0.167)	(0.168)	(0.131)	(0.144)	(0.270)
DTI (%)	0.011	0.006	0.006	0.014	0.014
	(0.009)	(0.009)	(0.007)	(0.009)	(0.015)
Race (%)	-0.008	-0.015	-0.015	-0.025	-0.026
	(0.003)	(0.004)	(0.003)	(0.004)	(0.008)
Broker competition	0.009	0.009	0.009	0.009	0.009
	(0.040)	(0.043)	(0.034)	(0.069)	(0.214)

Together with $\pi_0 = 1 - \pi_1 - \pi_2$, it implies

$$\pi_1 \quad = \quad \frac{1 - F_{f|X}(\underline{f}(X)) - \pi_2}{(1 - F_{f|X}(\underline{f}(X)))^N} \tag{11}$$

In other words, the condition $F_{\nu|X,\pi,N}(\underline{f}(X)) = 0$ implies that the distribution of borrower informedness is fully specified by the choice of π_2. For $N = 1$ and $F_{f|X}(\underline{f}(X))$=0.025, a choice of $\pi_2 = 0.75$ corresponds to $\pi = (0.02, 0.23, 0.75)$, a choice of $\pi_2 = 0.5$ corresponds to $\pi = (0.01, 0.49, 0.50)$, and a choice of $\pi_2 = 0.25$ corresponds to $\pi = (0.01, 0.74, 0.25)$. Note that for all three choices, the fraction of fully informed borrowers is rather small, which is in line with our belief that only few borrowers are fully informed about minimal broker charges.

Figure 4 shows the estimates for $F_{f|X}$ and $F_{\nu|X,\pi,N}$ for different choices of π. The data include all 2/28 purchase loans with full documentation, and the state vector x is set equal to the benchmark specification in Table 4. The figure confirms that as the fraction of uniformed borrowers, π_2, increases, the implied distribution of borrower valuations shifts to the left, meaning towards smaller values. Figure 5 displays the estimates for $F_{f|X}$ and $F_{\nu|X,\pi,N}$ for different numbers of friends N that partially informed borrowers consult with. As N increases, the implied distribution of borrower valuations shifts towards higher outcomes.

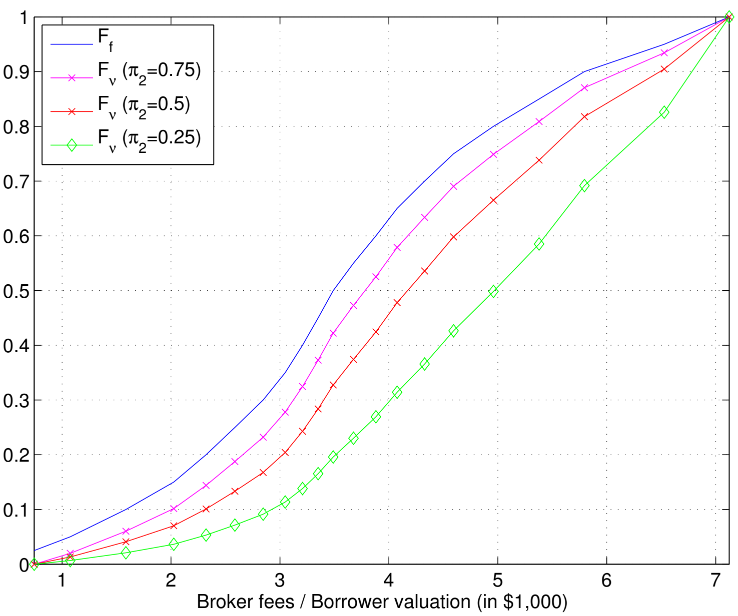

Figure 4: **Broker fees and borrower valuation for different** π The figure shows the conditional distributions of broker fees and of borrower valuations, for different choices of π_2. The data consist of 2/28 purchase loans with full documentation originated in California. The state vector x is set equal to the benchmark specification in Table 4 and $N = 1$.

5.1 Borrower valuations, broker reservation values and broker fees

For a given specification of X, π and N, we use the inverse CDF method (see Glasserman (2004)) to draw independent random samples from $F_{f|X}$ and $F_{\nu|X,\pi,N}$ and compute Monte Carlo estimates of expected broker fees and borrower valuations. For the benchmark specification of the state vector x, Table 6 reports expected fees of \$3.2K–\$3.7K for purchase loans and \$4.0K–\$4.7K for refinance loans.

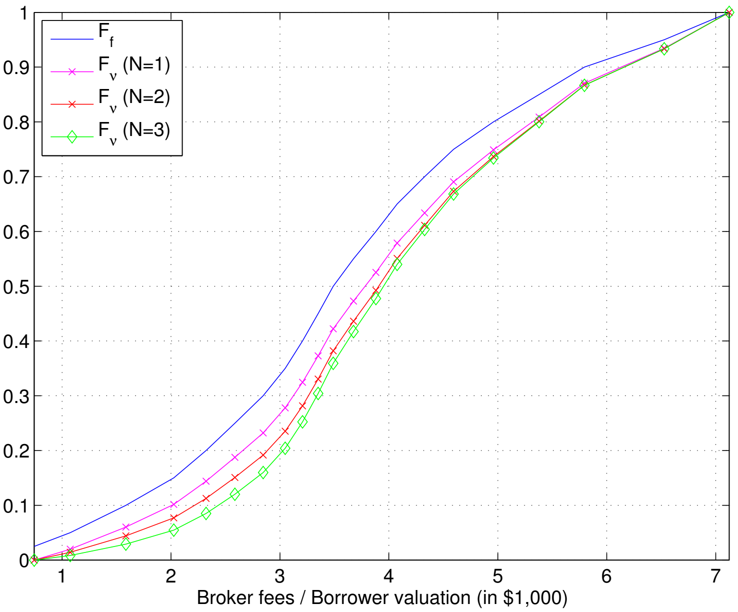

Figure 5: **Broker fees and borrower valuation for different** N The figure shows the conditional distributions of broker fees and of borrower valuations, as a function of the number of friends N. The data consist of 2/28 purchase loans with full documentation originated in California. The state vector x is set equal to the benchmark specification in Table 4 and $\pi_2 = 0.75$.

Table 6: **Expected broker fees** The table reports expected broker fees (in \$1,000). The data include all 2/28 loans and 30-year FRMs originated in California and are stratified by loan type, loan purpose and documentation level. The state vector x is specified as in Table 4 and $N = 1$. The MC estimates are based on 100,000 scenarios.

	Purchase					Refinance				
Docs	bm	large	small	80% wt	50% wt	bm	large	small	80% wt	50% wt
					2/28 loans					
Full	3.7	4.0	3.4	3.4	4.0	4.4	4.9	3.9	4.2	4.6
Low	3.6	4.0	3.3	3.3	3.9	4.7	5.1	4.2	4.4	4.9
					30-year FRMs					
Full	3.2	3.5	2.8	3.0	3.4	4.2	4.6	3.7	4.0	4.3
Low	3.7	4.1	3.2	3.4	3.9	4.0	4.5	3.6	3.8	4.3

Conditional on the borrower's level of informedness, fees range from the broker's reservation value $\underline{f}(X)$ for fully informed borrowers, to $\min\{f_1(X), \ldots, f_N(X), \nu(X, \pi, N)\}$ for partially informed borrowers, to $\nu(X, \pi, N)$ for uninformed borrowers. Expected fees, as a function of the borrower's level of informedness, are reported in Table 7 (for 2/28 loans) and in Table B.1 in the appendix (for 30-year FRMs). For benchmark 2/28 loans, the estimate of the broker's reservation value ranges from \$0.5K to \$1.1K, depending on loan purpose and documentation level.[17] All else the same, broker reservation values are estimated to be higher for refinance loans than for purchase loans.

Table 7: **Expected broker fees for 2/28 loans given borrower informedness** The table reports expected broker fees (in \$1,000) as a function of the borrower's level of informedness. The data include all 2/28 loans originated in California and are stratified by loan purpose and documentation level. The state vector x is specified as in Table 4 and $N = 1$. The MC estimates are based on 100,000 scenarios.

		Purchase					Refinance				
Docs	Info	bm	large	small	80% wt	50% wt	bm	large	small	80% wt	50% wt
						$\pi_2 = 0.75$					
	full	0.7	0.7	0.8	0.6	0.9	1.0	0.9	1.1	0.8	1.2
Full	partial	3.0	3.2	2.7	2.7	3.2	3.6	3.9	3.3	3.4	3.8
	none	4.0	4.3	3.6	3.7	4.3	4.8	5.4	4.2	4.6	4.9
	full	0.5	0.5	0.5	0.4	0.6	1.1	1.0	1.2	0.9	1.3
Low	partial	2.9	3.1	2.6	2.6	3.1	3.8	4.1	3.5	3.6	4.1
	none	3.9	4.3	3.6	3.6	4.3	5.0	5.6	4.5	4.8	5.3
						$\pi_2 = 0.5$					
	full	0.7	0.7	0.8	0.6	0.9	1.0	0.9	1.1	0.8	1.2
Full	partial	3.1	3.3	2.9	2.8	3.3	3.8	4.1	3.4	3.6	3.9
	none	4.3	4.7	3.9	4.0	4.6	5.2	5.8	4.5	5.0	5.3
	full	0.5	0.5	0.5	0.4	0.6	1.1	1.0	1.2	0.9	1.3
Low	partial	3.0	3.2	2.8	2.7	3.3	4.0	4.3	3.7	3.8	4.2
	none	4.3	4.8	3.9	4.0	4.6	5.4	6.0	4.8	5.2	5.7
						$\pi_2 = 0.25$					
	full	0.7	0.7	0.8	0.6	0.9	1.0	0.9	1.1	0.8	1.2
Full	partial	3.3	3.6	3.0	3.0	3.6	4.0	4.4	3.6	3.8	4.2
	none	4.9	5.4	4.4	4.6	5.3	5.9	6.7	5.0	5.7	6.0
	full	0.5	0.5	0.5	0.4	0.6	1.1	1.0	1.2	0.9	1.3
Low	partial	3.2	3.5	2.9	2.9	3.5	4.2	4.6	3.9	4.0	4.5
	none	4.9	5.5	4.4	4.6	5.3	6.1	6.9	5.4	5.9	6.3

For benchmark 2/28 loans and $N = 1$, expected borrower valuations—which are equal to expected fees for uninformed borrowers—range from \$3.9K to \$4.9K for purchase loans and from \$4.8K to \$6.1K for refinance loans. For partially informed borrowers, expected fees fall between those for fully informed and uninformed borrowers. They range between \$2.9K and \$3.3K for purchase loans and between \$3.6K and \$4.2K for refinance loans. Note that for partially informed or uninformed borrowers, expected fees decrease as the level of

[17] The broker's cost of originating the loan is equal to the sum of the broker's reservation value and YSP (see Equation (1)). For benchmark 2/28 loans, cost estimates range from \$2.5K to \$3.1K per loan, depending on loan purpose and documentation level.

informedness in the borrower population decreases. This is intuitive since the observed fee distribution is a mixture of $\min\{f_1, \ldots, f_N, \nu\}$ and ν (and, to a small extent, $\underline{f}$). As more and more weight is shifted towards ν, the implied distribution for ν, and hence that of $\min\{f_1, \ldots, f_N, \nu\}$, shifts towards smaller values.

5.1.1 Size effect

All else the same, expected broker fees are estimated to be higher for larger loans than for smaller loans. We refer to this finding as the *size effect*. For 2/28 loans, Table 6 reports expected fees of \$4.0K for large purchase loans versus \$3.3K–\$3.4K for small purchase loans, and of \$4.9K–\$5.1K for large refinance loans versus \$3.9–\$4.2K for small refinance loans.[18] Assuming that, all else the same, the distribution of informedness is the same among borrowers that take out loans of \$250K versus those that take out loans of \$150K, higher expected fees for larger loans imply higher expected borrower valuations for larger loans. This impact of the size of the loan on borrower valuations is reflected in the results in Table 7. It is also visualized in Figure 6, which shows the conditional distribution function $F_{\nu|X,\pi,N}$ for large versus small loans.

5.1.2 Race effect

In addition to the size effect, we also find a *race effect*: All else the same, expected broker fees are estimated to be higher for loans originated in more diverse neighborhoods with a lower fraction of white population compared to loans originated in less diverse neighborhoods with a higher fraction of white population. For 2/28 loans, Table 6 reports expected fees of \$3.9K–\$4.0K for purchase loans originated in more diverse neighborhoods versus \$3.3K–\$3.4K for purchase loans originated in less diverse neighborhoods, and of \$4.6K–\$4.9K for refinance loans originated in more diverse neighborhoods versus \$4.2K–\$4.4K for refinance loans originated in less diverse neighborhoods.[19]

[18] For 30-years FRMs, expected fees are \$3.5K–\$4.1K for large purchase loans versus \$2.8K–\$3.2K for small purchase loans, and \$4.5K–\$4.6K for large refinance loans versus \$3.6–\$3.7K for small refinance loans.

[19] For 30-years FRMs, expected fees are \$3.4K–\$3.9K for "50% white" purchase loans versus \$3.0K–\$3.4K for "80% white" purchase loans, and \$4.3K for "50% white" refinance loans versus \$3.8K–\$4.0K for "80% white" refinance loans.

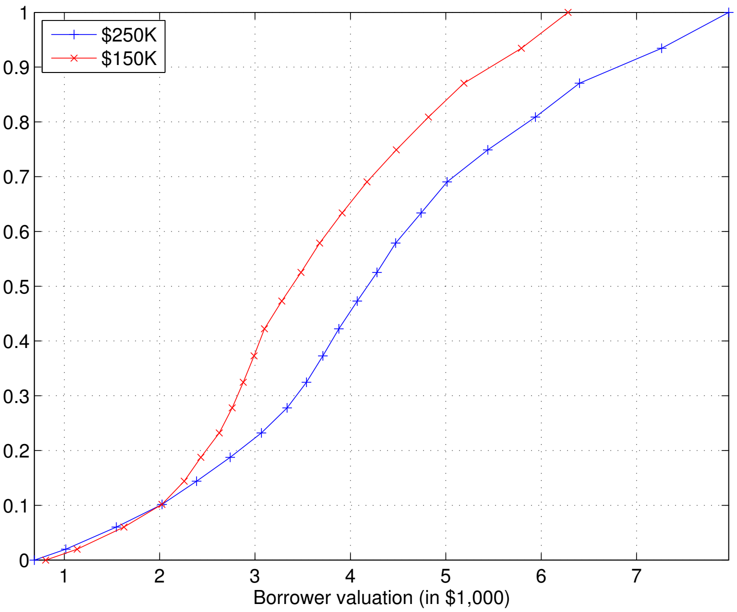

Figure 6: **Borrower valuations by loan amount** The figure shows the conditional distribution of borrower valuations, as a function of the loan amount. The data consist of 2/28 purchase loans with full documentation originated in California. The state vector x for the plot marked "$250K" is set equal to the "large loan" specification in Table 4, and x for the plot marked "$150K" is set equal to the "small loan" specification. We set $\pi_2 = 0.75$.

If the same distribution of borrower informedness is applied across neighborhoods, higher expected fees for loans originated in more diverse neighborhoods imply higher borrower valuations for these loans (see Table 7). A potentially more realistic description of borrower behavior, however, may be a scenario where conditional on X, borrowers draw from similar distributions of loan valuations, independent of the racial diversity within their zip code. Holding the fraction of uninformed borrowers in more diverse neighborhoods constant at 75%, we determine that fraction of uninformed borrowers in less diverse neighborhoods for which the implied borrower valuation distribution is as close as possible to that for more diverse neighborhoods.[20] Figure 7 reveals that the resulting fraction of uninformed borrowers in less diverse neighborhoods is 38.4%. In this sense, the data are consistent with the population of borrowers in less diverse zip codes being more informed than the population of borrowers

[20]We measure the "closeness" of the implied distributions as $\max_{a \in [\underline{f}, \overline{\nu})} \left| F_{\nu|X(50\%\text{wt}),\pi(50\%\text{wt}),N}(a) - F_{\nu|X(80\%\text{wt}),\pi(80\%\text{wt}),N}(a) \right|$.

in more diverse zip codes.

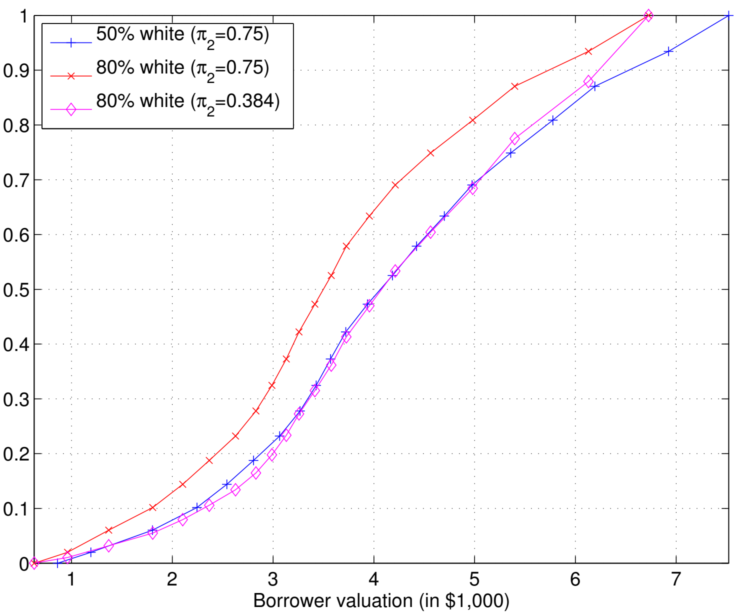

Figure 7: **Borrower valuations by race** The figure shows the conditional distribution of borrower valuations, as a function of the percentage of whites in the zip code. The data consist of 2/28 purchase loans with full documentation originated in California. The state vector x for the plot marked "50% white" is set equal to the "50% white" specification in Table 4, and x for the plots marked "80% white" is set equal to the "80% white" specification. We set $N = 1$.

5.1.3 Increase in broker fees from doing business with a less informed borrower

The results in Tables 7 and B.1 allow us to compute the expected increase in broker fees from doing business with a less informed borrower. Consider, for example, benchmark 2/28 purchase loans and $N = 1$. For $\pi_2 = 0.75$, expected fees are \$2.3K–\$2.4K higher when the loan is originated from a partially rather than a fully informed borrower, and \$1K higher when the loan is originated from an uninformed rather than a partially informed borrower. In addition, we find that as the level of informedness in the borrower population increases, the brokers' benefits from interacting with less informed borrowers increase. For $\pi_2 = 0.25$, expected fees are \$2.6K–\$2.7K higher for partially informed than for fully informed borrowers, and \$1.6K–\$1.7K higher for uninformed than for partially informed borrowers.

Moreover, the increase in expected fees as the borrower's informedness decreases is greater for larger loans, and for loans originated in more diverse zip codes. For benchmark 2/28 purchase loans and $N = 1$, the increase in expected fees per loan from doing business with a partially rather than a fully informed borrower is \$400-\$700 higher for large than for small loans, and \$200-\$400 higher for loans in more versus less diverse neighborhoods. The increase in expected fees from doing business with an uninformed rather than a partially informed borrower is \$200-\$500 higher for large than for small loans, and up to \$200 higher for loans in more versus less diverse zip codes.

5.2 Fraction of the overall surplus from the mortgage that goes to the broker

The fraction $\gamma = (\overline{f} - \underline{f})/(\nu - \underline{f})$ of the overall surplus from the mortgage that goes to the broker is zero when the borrower is fully informed and one when the borrower is uninformed. Conditional on X, π and N, the expected value of γ when the borrower is partially informed is shown in Table 8 (for 2/28 loans) and in Table B.1 (for 30-year FRMs). For benchmark 2/28 loans and $N = 1$, the expected value of γ varies between 75-76% when $\pi_2 = 0.75$, between 71-73% when $\pi_2 = 0.5$ and between 66-67% when $\pi_2 = 0.25$. As the level of informedness in the borrower population increases, the implied borrower valuation ν shifts towards higher values (see Section 5.1), and hence γ shifts towards smaller values.

Table 8: **Fraction of surplus from mortgage to partially informed borrower that goes to broker** The table reports the expected value of γ in Equation (2) when the borrower is partially informed. The data include all 2/28 loans originated in California and are stratified by loan purpose and documentation level. The state vector x is specified as in Table 4 and $N = 1$. The MC estimates are based on 100,000 scenarios.

	Purchase					Refinance				
Docs	bm	large	small	80% wt	50% wt	bm	large	small	80% wt	50% wt
					$\pi_2 = 0.75$					
Full	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.76	0.75	0.75
Low	0.75	0.74	0.75	0.74	0.75	0.76	0.75	0.77	0.76	0.76
					$\pi_2 = 0.5$					
Full	0.72	0.71	0.72	0.72	0.72	0.72	0.71	0.73	0.72	0.72
Low	0.71	0.71	0.72	0.71	0.72	0.73	0.72	0.73	0.72	0.73
					$\pi_2 = 0.25$					
Full	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.67	0.66	0.66
Low	0.66	0.65	0.67	0.65	0.67	0.67	0.67	0.68	0.67	0.68

5.3 Reduction in broker fees if borrowers become more informed

We are interested in the reduction of broker fees and γ if partially informed borrowers were to consult one extra friend, or if the borrower population became more informed.

5.3.1 Reduction in broker fees and γ if partially informed borrowers consult one extra friend

Given $F_{f|X}$ and $F_{\nu|X,\pi,N=1}$, we quantify how the fees that a broker extracts from a partially informed borrower would change in expectation if that borrower were to consult one extra friend. In particular, we compute the difference between the expected value of $\min\{f_1, f_2, \nu\}$ and the expected value of $\min\{f_1, \nu\}$, where f_1 and f_2 are independent draws from $F_{f|X}$ that are independent of $\nu \sim F_{\nu|X,\pi,N=1}$. This difference is clearly negative. Its absolute value is reported in Table 9. We find that a partially informed borrower that takes out a benchmark 2/28 loan saves an average of \$500-\$700 when consulting with two instead of just one friend.

We also compute how the expected fraction of the overall surplus from the mortgage that goes to the broker changes if a partially informed borrower consults with two rather than one friend. In particular, we compute the difference between the expected value of $(\min\{f_1, f_2, \nu\} - \underline{f})/(\nu - \underline{f})$ and the expected value of $(\min\{f_1, \nu\} - \underline{f})/(\nu - \underline{f})$, where f_1 and f_2 are independent draws from $F_{f|X}$ that are independent of $\nu \sim F_{\nu|X,\pi,N=1}$. For benchmark 2/28 loans, our results in Table 9 show that the expected value of γ drops by 13-15 percentage points.[21]

5.3.2 Reduction in broker fees if borrower population becomes more informed

Given $F_{f|X}$ and $F_{\nu|X,\pi,N}$, we are also in a position to quantify the changes in expected broker fees as the borrower population becomes more informed. Holding the implied borrower valuation distribution fixed, we compute

$$E(f|\pi^*; X, \pi, N) \quad = \quad \pi_0^* \, \underline{f}(X) + \pi_1^* \, E \min(f_1, \ldots, f_N, \nu) + \pi_2^* \, E\nu \qquad (12)$$

for a new distribution of borrower informedness, $\pi^* = (\pi_0^*, \pi_1^*, \pi_2^*)$.

[21] For benchmark 30-year FRMs, expected fees drop by \$400-\$700 per loan, and the expected value of γ drops by 12-15 percentage points.

Table 9: **Reduction in broker fees and γ if partially informed borrowers consult one extra friend** The table reports the decrease in expected broker fees (in \$1,000) and in the expected value of γ in Equation (2) if partially informed borrowers were to consult one extra friend. The data include all 2/28 loans originated in California and are stratified by loan purpose and documentation level. The distribution of borrower valuations is estimated based on the benchmark state vector x in Table 4 and $N = 1$. The MC estimates are based on 100,000 scenarios.

	Purchase					Refinance				
Docs	bm	large	small	80% wt	50% wt	bm	large	small	80% wt	50% wt
					Reduction in broker fees (in \$1,000)					
					$\pi_2 = 0.75$					
Full	0.5	0.5	0.4	0.4	0.5	0.5	0.6	0.4	0.5	0.5
Low	0.5	0.6	0.4	0.5	0.5	0.6	0.6	0.5	0.6	0.5
					$\pi_2 = 0.5$					
Full	0.5	0.6	0.4	0.5	0.6	0.6	0.7	0.5	0.6	0.6
Low	0.6	0.7	0.5	0.6	0.6	0.6	0.7	0.5	0.6	0.6
					$\pi_2 = 0.25$					
Full	0.6	0.7	0.5	0.6	0.7	0.7	0.9	0.6	0.7	0.7
Low	0.7	0.8	0.6	0.6	0.7	0.7	0.9	0.6	0.7	0.7
					Reduction in γ					
					$\pi_2 = 0.75$					
Full	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
Low	0.14	0.14	0.13	0.14	0.13	0.13	0.13	0.12	0.13	0.12
					$\pi_2 = 0.5$					
Full	0.14	0.14	0.13	0.14	0.14	0.14	0.14	0.13	0.14	0.14
Low	0.14	0.14	0.14	0.14	0.14	0.13	0.13	0.13	0.13	0.13
					$\pi_2 = 0.25$					
Full	0.14	0.15	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14
Low	0.15	0.15	0.14	0.15	0.15	0.14	0.14	0.14	0.14	0.14

Table 10 reports results for $\pi^* = (\pi_0, \pi_1 + \delta\pi_2, (1 - \delta)\pi_2)$, where $\delta = 0, 0.5, 1$. For $\delta = 0$, borrower informedness remains unchanged. We use $\delta = 0.5$ to describe the scenario where 50% of previously uninformed borrowers become partially informed, and $\delta = 1$ to describe the scenario where all previously uninformed borrowers become partially informed. For benchmark 2/28 loans and $N = 1$, we find that expected fees decrease by \$200-\$500 per loan if $\delta = 0.5$, and by \$400-\$900 if $\delta = 1$.[22] Not surprisingly, the reduction in expected fees is larger for larger values of π_2. Conditional on π, the absolute value of $E(f|\pi^*; X, \pi, N) - E(f|\pi; X, \pi, N)$ is somewhat higher for large than for small loans and for loans originated in more rather than less diverse neighborhoods.

[22]For benchmark 30-year FRMs, expected fees decrease by \$200-\$500 if $\delta = 0.5$, and by \$400-\$900 if $\delta = 1$.

Table 10: **Expected broker fees as borrowers become more informed** The table reports expected broker fees $E(f|\pi^*; X, \pi, N)$ defined in Equation (12) (in \$1,000) for different π^*. We consider the specifications $\pi^* = (\pi_0, \pi_1 + \delta\pi_2, (1-\delta)\pi_2)$, for $\delta = 0, 0.5, 1$. The data include all 2/28 loans originated in California and are stratified by loan purpose and documentation level. The distribution of borrower valuations is estimated based on the benchmark state vector x in Table 4 and $N = 1$. The MC estimates are based on 100,000 scenarios.

		Purchase					Refinance				
Docs	δ	bm	large	small	80% wt	50% wt	bm	large	small	80% wt	50% wt
						$\pi_2 = 0.75$					
	0	3.7	4.0	3.4	3.4	4.0	4.4	4.9	3.9	4.2	4.6
Full	0.5	3.3	3.6	3.0	3.0	3.6	4.0	4.4	3.6	3.8	4.2
	1	2.9	3.1	2.7	2.7	3.2	3.5	3.9	3.2	3.4	3.7
	0	3.6	4.0	3.3	3.3	3.9	4.7	5.1	4.2	4.4	4.9
Low	0.5	3.2	3.5	2.9	2.9	3.5	4.2	4.6	3.9	4.0	4.5
	1	2.8	3.0	2.6	2.5	3.1	3.8	4.1	3.5	3.5	4.0
						$\pi_2 = 0.5$					
	0	3.7	4.0	3.4	3.4	4.0	4.4	4.9	3.9	4.2	4.6
Full	0.5	3.4	3.7	3.1	3.1	3.6	4.1	4.5	3.6	3.9	4.2
	1	3.1	3.3	2.8	2.8	3.3	3.7	4.1	3.3	3.5	3.9
	0	3.6	4.0	3.3	3.3	3.9	4.7	5.1	4.2	4.4	4.9
Low	0.5	3.3	3.6	3.0	3.0	3.6	4.3	4.7	3.9	4.1	4.6
	1	3.0	3.2	2.7	2.7	3.2	4.0	4.3	3.6	3.7	4.2
						$\pi_2 = 0.25$					
	0	3.7	4.0	3.4	3.4	4.0	4.4	4.9	3.9	4.2	4.6
Full	0.5	3.5	3.8	3.2	3.2	3.8	4.2	4.7	3.7	4.0	4.4
	1	3.3	3.5	3.0	3.0	3.5	4.0	4.4	3.5	3.8	4.1
	0	3.6	4.0	3.3	3.3	3.9	4.7	5.1	4.2	4.4	4.9
Low	0.5	3.4	3.7	3.1	3.1	3.7	4.4	4.9	4.0	4.2	4.7
	1	3.2	3.5	2.9	2.9	3.5	4.2	4.6	3.8	4.0	4.5

5.4 *Borrower informedness conditional on broker fees*

For a given X, π and N, the ex-post distribution of borrower informedness is given by

$$\begin{aligned}
\pi_0^{ex,-}(a) &= Prob(\text{borrower is fully informed}|f \le a) = \frac{\pi_0}{F_{f|X}(a)} \\
\pi_1^{ex,-}(a) &= Prob(\text{borrower is part. informed }|f \le a) = \frac{F_{f|X}(a) - \pi_0 - \pi_2 F_{\nu|X,\pi,N}(a)}{F_{f|X}(a)} \\
\pi_2^{ex,-}(a) &= Prob(\text{borrower is uninformed }|f \le a) = \frac{\pi_2 F_{\nu|X,\pi,N}(a)}{F_{f|X}(a)},
\end{aligned}$$

for $a \in [\underline{f}(X), \overline{\nu}(X)]$. For benchmark 2/28 purchase loans with full documentation, Figure 8 displays $\pi_0^{ex,-}(a)$, $\pi_1^{ex,-}(a)$ and $\pi_2^{ex,-}(a)$ as a function of a. As the fee threshold a increases, the ex-post probability that a borrower whose fees do not exceed a is fully informed decreases and the ex-post probability that the borrower is uninformed increases. The ex-post probability of the borrower being partially informed is a hump shaped function of a.

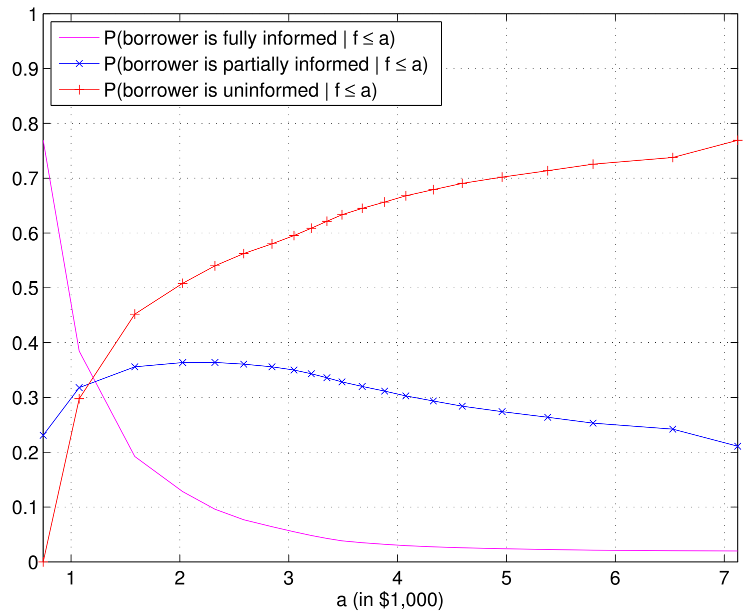

Figure 8: **Ex-post distribution of borrower informedness given fee threshold** The figure shows the ex-post distribution of borrower informedness given that broker fees do not exceed a certain threshold. The data consist of 2/28 purchase loans with full documentation originated in California. The state vector x is equal to the benchmark specification in Table 4. We set $N = 1$ and $\pi_2 = 0.75$.

Let $\pi_0^{ex,+}(a) = Prob(\text{borrower is fully informed}|f > a)$ for $a \in [\underline{f}(X), \overline{\nu}(X))$, and let $\pi_1^{ex,+}(a)$ and $\pi_2^{ex,+}(a)$ denote similar ex-post probabilities for the borrower being partially informed or uninformed. Bayes' rule implies

$$\pi_i^{ex,+}(a) = \frac{\pi_i - \pi_i^{ex,-}(a)F_{f|X}(a)}{1 - F_{f|X}(a)}, \quad i = 0, 1, 2.$$

For $\underline{f}(X) \leq a < b < \overline{\nu}(X)$ and $\pi_0^{ex}(a,b) = Prob(\text{borrower is fully informed}|a < f \leq b)$, and similar definitions for $\pi_1^{ex}(a,b)$ and $\pi_2^{ex}(a,b)$, we have

$$\pi_i^{ex}(a,b) = \frac{\pi_i - \pi_i^{ex,-}(a)F_{f|X}(a) - \pi_i^{ex,+}(b)(1 - F_{f|X}(b))}{F_{f|X}(b) - F_{f|X}(a)}, \quad i = 0, 1, 2.$$

Figure 9 displays $\pi_1^{ex}(a - \$100, a)$ and $\pi_2^{ex}(a - \$100, a)$ as a function of a, again for benchmark 2/28 purchase loans with full documentation. As observed fees increase, more and more likelihood is attributed to the borrower being uninformed. Since fully informed borrowers always pay $\underline{f}(X)$, $\pi_0^{ex}(a - \$100, a) = 0$ for all $a \in [\underline{f}(X), \overline{\nu}(X))$.

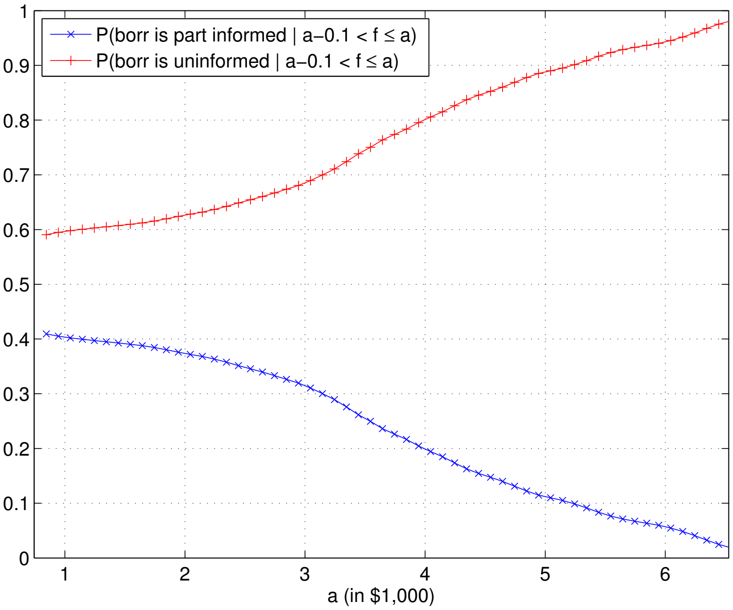

Figure 9: **Ex-post distribution of borrower informedness given fee interval** The figure shows the ex-post distribution of borrower informedness given that broker fees are within the interval $(a - \$100, a]$, for $a \in [\underline{f}(X), \overline{\nu}(X))$. The data consist of 2/28 purchase loans with full documentation originated in California. The state vector x is equal to the benchmark specification in Table 4. We set $N = 1$ and $\pi_2 = 0.75$.

6. Robustness

In this section, we extend the empirical analysis to loans originated in Florida, New York or Illinois. We also investigate the sensitivity of our results to particular model assumptions.

6.1 Loans originated in Florida, New York or Illinois

For 2/28 loans originated in Florida, New York or Illinois, Tables B.2, B.3 and B.4 in the appendix report expected broker fees.[23] A striking feature of the data is that both the size and the race effect are present in each state. For example, the difference in expected

[23] Untabulated results for 30-year FRMs support the findings presented in this section.

fees between large and small 2/28 purchase loans with full documentation is about \$600 in California, \$900 in Florida, \$700 in New York and \$100 in Illinois. And the difference in expected fees between 2/28 purchase loans with full documentation in more versus less diverse zip codes is \$600 in California, \$200 in Florida, \$500 in New York and \$200 in Illinois.

To further highlight the fact that we observe higher conditional fees for larger loans and for loans originated in more diverse zip codes, Table 11 reports the non-crossing quantile regression coefficients for the loan amount and race conditioning variables. For 2/28 loans, we find that a marginal increase in the size of the loan is associated not only with a shift of the conditional broker fee distribution to the right—that is, towards larger fees—but also with an increase in the right-skewness of the fee distribution.[24] A marginal decrease in the fraction of white population in the zip code is also associated with a shift of the conditional fee distribution to the right.

Table 11: **Non-crossing quantile regression coefficients for loan amount and race** The table reports the results from non-crossing quantile regressions of broker fees (in \$1,000) on the state vector x. It lists the regression coefficients for loan amount and race for the 25% (columns marked "25") and 75% (columns marked "75") quantiles. Statistically insignificant regression coefficients are underlined. The data include all 2/28 loans and are stratified by state, loan purpose and documentation level. "F" stands for full documentation and "L" for low documentation.

	Purchase								Refinance							
	CA		FL		NY		IL		CA		FL		NY		IL	
	25	75	25	75	25	75	25	75	25	75	25	75	25	75	25	75
	Loan amount (\$100,000)															
F	0.3	0.8	0.3	1.3	0.8	2.0	-0.2	0.3	0.6	1.5	0.7	1.8	1.3	2.7	0.0	1.1
L	0.3	1.0	0.2	1.2	0.7	1.8	0.0	0.6	0.5	1.3	0.3	1.3	0.9	1.9	-0.1	0.6
	Race (nominal fraction)															
F	-1.5	-2.5	-0.7	-0.7	-1.1	-1.9	-0.3	-0.8	-1.1	-1.2	-1.1	-1.2	-1.3	-0.9	-0.7	-0.5
L	-1.8	-2.5	-0.9	-0.9	-2.0	-2.5	-0.5	-0.7	-1.7	-1.5	-0.7	-0.7	-1.8	-1.1	-0.7	-0.8

[24] Even though the loan amount coefficient for 2/28 full-documentation purchase loans originated in Illinois is negative for the 25% quantile, it is positive, significant and larger in absolute values for quantiles above 75%. Closer inspection of this stratum of loans reveals that a marginal increase in the size of the loan is associated with a wider but more right-skewed fee distribution, implying higher expected fees for larger loans. For 2/28 low-documentation purchase loans originated in Illinois, the regression coefficients are negative and significant between the 25% and the 75% quantile, and negative but insignificant at and outside of the interquartile range. Hence a marginal decrease in the race variable is associated with a shift of the center but not the tails of the conditional fee distribution to the right, implying higher expected fees for loans originated in more diverse neighborhoods.

Another robust feature of the data is that a partially informed borrower that consults with just one friend leaves a large fraction of the overall surplus from the mortgage on the table. For benchmark 2/28 loans and $N = 1$, the expected value of a partially informed borrower's γ varies between 75-76% in California, 71-73% in Florida, 70-80% in New York and 70-74% in Illinois when $\pi_2 = 0.75$, between 71-73% in California, 68-70% in Florida, 67-77% in New York and 66-70% in Illinois when $\pi_2 = 0.5$, and between 66-67% in California, 62-64% in Florida, 62-73% in New York and 59-65% in Illinois when $\pi_2 = 0.25$. For California and Florida, conditional on π we find little variation in the expected value of γ across different strata of the data. But for loans originated in New York or Illinois, the fraction of the overall surplus from a mortgage to a partially informed borrower that goes to the broker is generally lower for 2/28 purchase loans with full documentation and higher for 2/28 refinance loans with low documentation.

6.2 Model extension

We now lift the assumption that, conditional on X and N, the borrower's valuation for the loan is independent of the borrower's level of informedness. One could imagine that borrowers with a high valuation for their loan are more inclined to consult with friends than borrowers with a low valuation, mainly because their benefits from becoming informed would be higher. Alternatively, high borrower valuation could be indicative of overly optimistic borrowers, which in turn could be indicative of less informed borrowers.

Let $F_{1,\nu}(a) = Prob(\nu \leq a|\text{borrower is partially informed})$ and $F_{2,\nu}(a) = Prob(\nu \leq a|\text{borrower is uninformed})$. Then the equilibrium distribution of broker fees is given by

$$\begin{aligned} F_f(a) &= \pi_0 + \pi_1 Prob(\min\{f_1, \ldots, f_N, \nu\} \leq a|\text{borr part informed}) + \pi_2 F_{2,\nu}(a) \\ &= \pi_0 + \pi_1 \left(1 - Prob(\nu > a|\text{borr part informed}) \prod_{i=1}^{N} Prob(f_i > a)\right) + \pi_2 F_{2,\nu}(a) \\ &= \pi_0 + \pi_1 \left\{1 - (1 - F_{1,\nu}(a))(1 - F_f(a))^N\right\} + \pi_2 F_{2,\nu}(a), \end{aligned} \quad (13)$$

where $a \in [\underline{f}, \overline{\nu}]$. Equation (13) can be rewritten as

$$F_f(a) + \widetilde{\kappa}_1(a)(1 - F_f(a))^N - \widetilde{\kappa}_0(a) \quad = \quad 0, \tag{14}$$

where $a \in [\underline{f}, \overline{\nu}]$, $\widetilde{\kappa}_1(a) = \pi_1(1 - F_{1,\nu}(a))$ and $\widetilde{\kappa}_0(a) = \pi_0 + \pi_1 + \pi_2 F_{2,\nu}(a)$.

In Appendix A we prove that the equilibrium distribution of broker fees is well-defined by Equation (14), and that $F_f \geq \min\{F_{1,\nu}, F_{2,\nu}\}$.

Proposition 4 *Let $F_{1,\nu}$ and $F_{2,\nu}$ be some distribution functions with support on $[\underline{f}, \overline{\nu}]$. For $N \geq 1$, $\pi_0, \pi_1, \pi_2 > 0$ and $\pi_1 N \leq 1$, the equilibrium distribution of broker fees is well-defined by Equation (14), has support on $[\underline{f}, \overline{\nu}]$ and is unique.*

Proposition 5 *For $N \geq 1$, $\pi_0, \pi_1, \pi_2 > 0$ and $\pi_1 N \leq 1$, the equilibrium distribution of broker fees satisfies $F_f(a) \geq \min\{F_{1,\nu}(a), F_{2,\nu}(a)\}$, for all $a \in [\underline{f}, \overline{\nu}]$.*

Equation (13) identifies $F_f(a)$ as a function of $F_{1,\nu}(a)$ and $F_{2,\nu}(a)$. Since we observe the fees a borrower pays to the broker but not the value the borrower assigns to obtaining the mortgage we need to invert the relationship in (13) and express $F_{1,\nu}(a)$ and $F_{2,\nu}(a)$ as a function of $F_f(a)$. To do so we have to impose an over-identifying restriction that links $F_{1,\nu}(a)$ and $F_{2,\nu}(a)$. For $a \in [\underline{f}, \overline{\nu}]$ and some positive scalar α, consider the specification

$$F_{1,\nu}(a) \quad = \quad w_\alpha(F_{2,\nu}(a)), \qquad F_{2,\nu}(\underline{f}) = 0, \tag{15}$$

where $w_\alpha(\cdot)$ has support on $[0, 1]$ and is defined as

$$w_\alpha(x) \quad = \quad \begin{cases} (1 - \alpha^x)/(1 - \alpha), & \text{if } \alpha \neq 1 \\ x, & \text{if } \alpha = 1. \end{cases} \tag{16}$$

Figure 10 shows and Lemma 2 in Appendix A proves that as long as $F_{2,\nu}$ is a well-defined distribution function, so is $F_{1,\nu}$ defined in (15).

Figure 10 and Lemma 2 also show that for $\alpha < 1$, $F_{1,\nu} \geq F_{2,\nu}$, and that for $\alpha > 1$, $F_{1,\nu} \leq F_{2,\nu}$. It implies that if α is less than one, the borrower valuation distribution conditional on

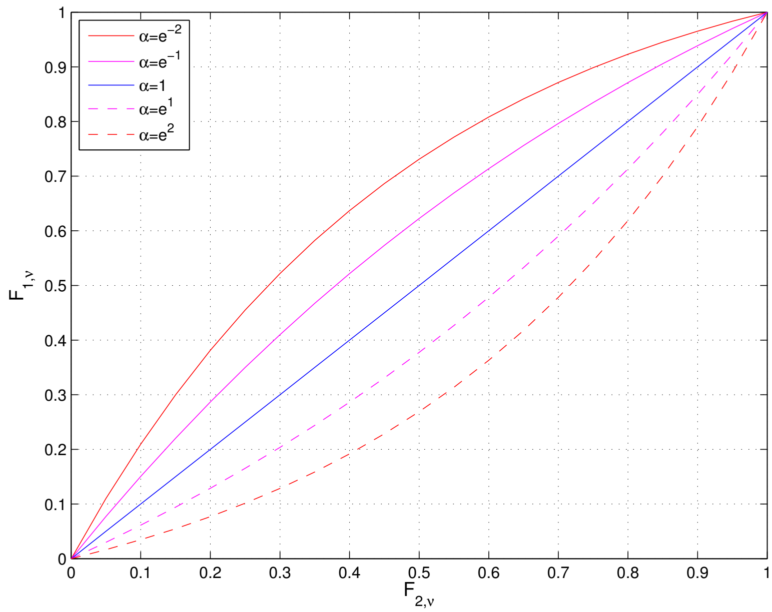

Figure 10: **Relation between conditional borrower valuation distributions for different** α The figure shows $F_{1,\nu}$ as a function of $F_{2,\nu}$ as defined in Equation (15), for different values of α.

the borrower being partially informed lies to the left of the borrower valuation distribution conditional on the borrower being uninformed. That relationship is reversed for α larger than one. Only for α equal to one is the borrower's valuation for the loan independent of the borrower's level of informedness. This highlights the fact that different choices for α capture different relations between $F_{1,\nu}$ and $F_{2,\nu}$.

Combining Equation (15) with Equation (13) yields

$$F_f(a) \quad = \quad \pi_0 + \pi_1 \left\{1 - (1 - w_\alpha(F_{2,\nu}(a)))(1 - F_f(a))^N\right\} + \pi_2 F_{2,\nu}(a), \tag{17}$$

which allows us to compute $F_{2,\nu}(a)$, and hence $F_{1,\nu}(a)$, as a function of F_f. The next proposition states that the same conditions on the model parameters that ensure the existence of F_ν in Section 2 are sufficient to ensure that $F_{2,\nu}$, and hence $F_{1,\nu}$, is well-defined:

Proposition 6 *Let F_f be some distribution function with support on $[\underline{f}, \overline{\nu}]$. For $N \geq 1$, $\pi_0, \pi_1, \pi_2 > 0$ and $\pi_1 N \leq 1$, $F_{2,\nu}$ given by Equation (17) is a well-defined distribution function as long as $F_f(\underline{f}) \geq x$, where $x \in (0,1)$ solves $x + \pi_1(1-x)^N - (1-\pi_2) = 0$. $F_{2,\nu}$ has support* *on* $[\underline{f}, \overline{\nu}]$ *and is unique.*

In Table 12 we replicate the results from Tables 7 and 8 for different levels of α. A finding that is robust to different specifications of the relation between $F_{1,\nu}$ and $F_{2,\nu}$ is that a partially informed borrower that consults with only one friend shares a large fraction of the overall surplus from the mortgage with the broker. Even though that fraction declines somewhat as $F_{1,\nu}$ shifts to the right—that is, as α becomes larger—Table 12 reveals that for benchmark 2/28 loans originated in California the average fraction remains above 60%.

Table 12: **Expected broker fees and γ for 2/28 loans for different** α The table reports, in rows identified by "partial" and "none", expected broker revenues (in $1,000) for partially informed and uninformed borrowers, for different levels of α in Equation (15). The table also reports, in rows identified by "γ", the expected value of γ in Equation (2) when the borrower is partially informed. Columns marked "e^{-2}" ("e^{-1}" ,"1" ,"e^{1}" ,"e^{2}") correspond to $\alpha = e^{-2}$ ($\alpha = e^{-1}$, $\alpha = 1$, $\alpha = e^{1}$, $\alpha = e^{2}$). The data include all 2/28 loans originated in California and are stratified by loan purpose and documentation level. The state vector x is the benchmark state vector in Table 4 and $N = 1$. The MC estimates are based on 100,000 scenarios.

		Purchase					Refinance				
Docs	Info	e^{-2}	e^{-1}	1	e^{1}	e^{2}	e^{-2}	e^{-1}	1	e^{1}	e^{2}
						$\pi_2 = 0.75$					
	partial	2.7	2.8	3.0	3.1	3.2	3.2	3.4	3.6	3.8	3.9
Full	none	4.1	4.0	4.0	3.9	3.9	4.9	4.8	4.8	4.7	4.7
	γ	0.81	0.78	0.75	0.71	0.68	0.81	0.79	0.75	0.72	0.68
	partial	2.5	2.7	2.9	3.0	3.2	3.5	3.7	3.8	4.0	4.2
Low	none	4.0	4.0	3.9	3.9	3.9	5.1	5.1	5.0	5.0	4.9
	γ	0.81	0.78	0.75	0.71	0.68	0.82	0.79	0.76	0.72	0.69
						$\pi_2 = 0.5$					
	partial	2.9	3.0	3.1	3.2	3.3	3.5	3.6	3.8	3.9	4.0
Full	none	4.5	4.4	4.3	4.2	4.1	5.4	5.3	5.2	5.1	4.9
	γ	0.77	0.75	0.72	0.68	0.65	0.77	0.75	0.72	0.69	0.66
	partial	2.8	2.9	3.0	3.1	3.2	3.8	3.9	4.0	4.1	4.2
Low	none	4.5	4.4	4.3	4.2	4.1	5.7	5.5	5.4	5.3	5.2
	γ	0.77	0.74	0.71	0.68	0.65	0.78	0.76	0.73	0.70	0.67
						$\pi_2 = 0.25$					
	partial	3.2	3.2	3.3	3.4	3.4	3.9	3.9	4.0	4.0	4.1
Full	none	5.2	5.1	4.9	4.7	4.6	6.2	6.0	5.9	5.7	5.4
	γ	0.70	0.68	0.66	0.64	0.61	0.70	0.68	0.66	0.64	0.61
	partial	3.1	3.2	3.2	3.3	3.3	4.1	4.2	4.2	4.3	4.4
Low	none	5.2	5.1	4.9	4.7	4.6	6.4	6.3	6.1	5.9	5.7
	γ	0.70	0.68	0.66	0.64	0.61	0.71	0.69	0.67	0.65	0.63

Recall that as α becomes larger, partially informed borrowers' loan valuations generally become larger relative to uninformed borrowers' valuations. Hence as α becomes larger, expected broker fees paid by partially informed borrowers increase whereas expected fees paid by uninformed borrowers decrease. We find, however, that the deviations in expected fees from the values reported in Section 5 for $\alpha = 1$ are fairly small. Note that our estimates for broker reservation values do not change as α moves away from 1. Hence the fees paid by fully informed borrowers are the same as those reported in Table 7, for all choices of α. Since variation of α does not alter the observed fee distribution, expected fees across all borrowers do not depend on α. In particular, independent of the relation between $F_{1,\nu}$ and $F_{2,\nu}$, expected fees are higher for larger than for smaller loans, and for loans originated in more versus less diverse neighborhoods.

7. Conclusion

We develop an equilibrium model for broker fees and show how the equilibrium distribution of fees depends on borrowers' loan valuations and their level of informedness. Since we do not observe borrowers' loan valuations nor how informed they are, we invert the equilibrium relationship and express the distribution of borrower valuations as a function of the distribution of fees, subject to the level of informedness in the borrower population. Estimates for the fee distribution, as a function of loan and borrower characteristics, are obtained using quantile regressions with an explicit non-crossing constraint that prevents any overlaps among quantile functions.

Using data from, formerly, one of the largest subprime lenders, we document that conditional fees are generally higher for larger loans than for smaller loans, and for loans originated in more diverse neighborhoods than loans originated in less diverse neighborhoods. If, all else the same, the distribution of informedness is the same among borrowers that take out large loans and borrowers that take out small loans, our findings are consistent with large borrowers assigning a higher value to their loan than similarly informed small borrowers. And if borrowers draw from the same distribution of loan valuations independent of the racial diversity within their zip code, then our results are consistent with borrowers in less diverse neighborhoods being more informed than otherwise similar borrowers in more diverse neighborhoods. In any case, our findings suggest that brokers expect to earn high conditional fees from contact with large borrowers and with borrowers in more diverse zip codes, and hence have an incentive to exert additional effort to locate and attract such borrowers.

While fully informed borrowers pocket the entire spread between their value for the loan and the minimum fee for which the broker is willing to originate it, uninformed borrowers are left with nothing. Partially informed borrowers, on the other hand, split the overall surplus from the mortgage with the broker. If partially informed borrowers consult with only one friend, the share of the overall surplus that goes to the broker is rather large. Indeed, for a wide range of loan and borrower characteristics, such borrowers generally retain less than 40% of the overall surplus. If, however, a partially informed borrower were to consult with one extra friend, that fraction would drop dramatically, often by more than 10 percentage points.

References

Ambrose, B., J. Conklin, 2012. Mortgage Brokers, Origination Fees, Price Transparency and Competition. Forthcoming, Real Estate Economics.

Berndt, A., B. Hollifield, P. Sandas, 2012. What Broker Charges Reveal about Mortgage Credit Risk. Working paper, Carnegie Mellon University.

Bondell, H., B. Reich, H. Wang, 2010. Non-crossing Quantile Regression Curve Estimation. Biometrika 97, 825–838.

Glasserman, P., 2004. Monte Carlo Methods in Financial Engineering. Springer, New York, NY.

ICF Macro, 2009. Design and Testing of Truth-in-Lending Disclosures for Closed-end Mortgages. Available at www.federalreserve.gov.

Koenker, R., P. Ng, 2003. SparseM: A Sparse Matrix Package for R. Journal of Statistical Software 8, available at www.jstatsoft.org/v08/i06.

Lacko, J., J. Pappalardo, 2007. Improving Consumer Mortgage Disclosure. Federal Trade Commission Staff Report.

Woodward, S., R. Hall, 2009. The Equilibrium Distribution of Prices Paid by Imperfectly Informed Customers: Theory and Evidence from the Mortgage Market. Working paper, Stanford University.

A. Proofs

Lemma 1 *For $N \geq 1$ and $0 < \kappa_1 < \kappa_0 < 1$, $p(x) = x + \kappa_1(1-x)^N - \kappa_0$ has exactly one root in $[0,1]$, and it is in $(0,1)$.*

Proof. A proof of Lemma 1 can be found in Woodward and Hall (2009). To make our paper self-contained, we reproduce it here. The function $p(x)$ is continuous, and $p(0) = \kappa_1 - \kappa_0 < 0$ and $p(1) = 1 - \kappa_0 > 0$. So there exists at least one root between 0 and 1. To rule out the possibility of more than one root, consider

$$p'(x) = 1 - \kappa_1 N(1-x)^{N-1}$$

and

$$p''(x) = \begin{cases} \kappa_1 N(N-1)(1-x)^{N-2} \geq 0, & \text{if } N > 1 \\ 0, & \text{if } N = 1. \end{cases}$$

If $N = 1$ or if $N > 1$ and $\kappa_1 N \leq 1$, then $p'(x) > 0$ in $(0,1)$, meaning there is only one root of $p(x)$ in $[0,1]$, and it is in $(0,1)$. If $N > 1$ and $\kappa_1 N > 1$, then $p'(x^*) = 0$ if and only if

$$x^* = 1 - (\kappa_1 N)^{-\frac{1}{N-1}}.$$

Because $p''(x)$ is non-negative, $p'(x)$ is positive for $x^* < x \leq 1$ and negative for $0 \leq x < x^*$. So again $p(x)$ has only one root in $[0,1]$, and it is in $(0,1)$.

Proposition 1 *Let F_ν be some distribution function with support on $[\underline{f}, \overline{\nu}]$. For $N \geq 1$, $\pi_0, \pi_1, \pi_2 > 0$ and $\pi_1 N \leq 1$, the equilibrium distribution of broker fees is well-defined by Equation (4), has support on $[\underline{f}, \overline{\nu}]$ and is unique.*

Proof. Equation (4) implies that $F_\nu(\overline{\nu}) = 1$ if and only if $F_f(\overline{\nu}) = 1$. The broker's participation constraint implies that $F_f(a) = 0$ for $a < \underline{f}$. Hence F_f has support on $[\underline{f}, \overline{\nu}]$. Let

$y = F_f(a)$ and $x = F_\nu(a)$. According to Equation (4),

$$y + \pi_1(1-x)(1-y)^N - \pi_0 - \pi_1 - \pi_2 x \;\; = \;\; 0.$$

Differentiation w.r.t. x yields $y' - \pi_1(1-y)^N - \pi_1 N(1-x)(1-y)^{N-1}y' - \pi_2 = 0$. For $x \in [0,1]$ and $\pi_1 N \leq 1$, we have $\pi_1 N(1-x)(1-y)^{N-1} < 1$. (Note that for $N = 1$, $\pi_1 N < 1$. For $N > 1$ and $x = 0$, y solves Equation (4) with $\kappa_1 = \pi_1$ and $\kappa_0 = \pi_0 + \pi_1$. According to Lemma 1, $y \in (0,1)$.) Hence

$$y' \;\; = \;\; \frac{\pi_1(1-y)^N + \pi_2}{1 - \pi_1 N(1-x)(1-y)^{N-1}} > 0,$$

which implies that $F_f(a)$ is monotonically increasing in $a \in [\underline{f}, \overline{\nu}]$. Uniqueness follows immediately from Lemma 1.

Proposition 2 *For $N \geq 1$, $\pi_0, \pi_1, \pi_2 > 0$ and $\pi_1 N \leq 1$, the equilibrium distribution of broker fees satisfies $F_f(a) \geq F_\nu(a)$ for all $a \in [\underline{f}, \overline{\nu}]$.*

Proof. For any $a \in [\underline{f}, \overline{\nu}]$, Equation (5) implies

$$Prob(f > a) \;\; = \;\; \delta \, Prob(\nu > a),$$

where $0 < \delta = \pi_1(1 - F_f(a))^N + \pi_2 < 1$. Therefore, $Prob(f > a) \leq P(\nu > a)$ and hence $F_f(a) \geq F_\nu(a)$.

Proposition 3 *Let F_f be some distribution function with support on $[\underline{f}, \overline{\nu}]$. For $N \geq 1$, $\pi_0, \pi_1, \pi_2 > 0$ and $\pi_1 N \leq 1$, F_ν given by Equation (5) is a well-defined distribution function as long as $F_f(\underline{f}) \geq x$, where $x \in (0,1)$ solves $x + \pi_1(1-x)^N - (1-\pi_2) = 0$. F_ν has support on $[\underline{f}, \overline{\nu}]$ and is unique.*

Proof. What we have to show is that

(a) F_ν is unique and $F_\nu(\overline{\nu}) = 1$ (immediate)

(b) $F_\nu(\underline{f}) \geq 0$

(c) $F_\nu(a)$ is monotonically increasing in a for $a \in [\underline{f}, \overline{\nu}]$.

With regard to (b), $F_\nu(\underline{f}) \geq 0$ if $p(x^*) \leq 1$, where $x^* = F_f(\underline{f})$ and

$$p(x) = \frac{1-x}{\pi_1(1-x)^N + \pi_2}.$$

We have $p(0) = \frac{1}{\pi_1+\pi_2} > 1$ and $p(1) = 0$, so there is at least some $x \in (0,1)$ for which $p(x) = 1$. Taking derivatives of $p(x)$ w.r.t. x, for $x \in (0,1)$, yields

$$\begin{aligned} p'(x) &= \frac{\pi_1(N-1)(1-x)^N - \pi_2}{(\pi_1(1-x)^N + \pi_2)^2} \\ &= \frac{-\delta + \pi_1 N(1-x)^N}{\delta^2} \\ &= \frac{-\delta + \pi_1 N\delta(1-y)(1-x)^{N-1}}{\delta^2} \\ &= -\frac{1 - \pi_1 N(1-y)(1-x)^{N-1}}{\delta}, \end{aligned}$$

where $\delta = \pi_1(1-x)^N + \pi_2$. As long as $\pi_1 N \leq 1$, $p'(x) < 0$. (Note that for $N = 1$, $\pi_1 N < 1$. For $N > 1$ and $x \in (0,1)$, $(1-y)(1-x)^{N-1} < 1$.) Hence $p(x^*) \leq 1$ if and only if $x^* \geq x$, where $x + \pi_1(1-x)^N - (1-\pi_2) = 0$. According to Lemma 1, $x \in (0,1)$.

With regard to (c), let $y = F_\nu(a)$ and $x = F_f(a)$ for some $a \in [\underline{f}, \overline{\nu}]$. Equation (5) implies $y = 1 - p(x)$. Differentiation w.r.t. x yields $y' = 1 - p'(x)$, which implies that $y' \geq 0$ as long as $\pi_1 N \leq 1$.

Proposition 4 *Let $F_{1,\nu}$ and $F_{2,\nu}$ be some distribution functions with support on $[\underline{f}, \overline{\nu}]$. For $N \geq 1$, $\pi_0, \pi_1, \pi_2 > 0$ and $\pi_1 N \leq 1$, the equilibrium distribution of broker fees is well-defined by Equation (14), has support on $[\underline{f}, \overline{\nu}]$ and is unique.*

Proof. Equation (14) implies that $F_\nu(\overline{\nu}) = 1$ if and only if $F_f(\overline{\nu}) = 1$. The broker's participation constraint implies that $F_f(a) = 0$ for $a < \underline{f}$. Hence F_f has support on $[\underline{f}, \overline{\nu}]$.

Let $y = F_f(a)$, $x_1 = F_{1,\nu}(a)$ and $x_2 = F_{2,\nu}(a)$. According to Equation (14),

$$y + \pi_1(1-x_1)(1-y)^N - \pi_0 - \pi_1 - \pi_2\, x_2 \;=\; 0.$$

Differentiation w.r.t. x_1 yields $y' - \pi_1(1-y)^N - \pi_1 N(1-x_1)(1-y)^{N-1}y' - \pi_2 x_2' = 0$. For $x_1 \in [0,1]$ and $\pi_1 N \le 1$, we have $\pi_1 N(1-x_1)(1-y)^{N-1} < 1$. (Note that for $N = 1$, $\pi_1 N < 1$. For $N > 1$ and $x_1 = 0$, y solves Equation (4) with $\kappa_1 = \pi_1$ and $\kappa_0 = \pi_0 + \pi_1 + \pi_2 x_2$. According to Lemma 1, $y \in (0,1)$.) Hence

$$y' \;=\; \frac{\pi_1(1-y)^N + \pi_2 x_2'}{1 - \pi_1 N(1-x_1)(1-y)^{N-1}},$$

which implies that $F_f(a)$ is monotonically increasing in $a \in [\underline{f}, \overline{\nu}]$. Uniqueness follows immediately from Lemma 1.

Proposition 5 *For $N \ge 1$, $\pi_0, \pi_1, \pi_2 > 0$ and $\pi_1 N \le 1$, the equilibrium distribution of broker fees satisfies $F_f(a) \ge \min\{F_{1,\nu}(a), F_{2,\nu}(a)\}$, for all $a \in [\underline{f}, \overline{\nu}]$.*

Proof. First consider the case where $F_{1,\nu}(a) \le F_{2,\nu}(a)$ for some $a \in [\underline{f}, \overline{\nu}]$. Equation (14) implies $F_f(a) + \pi_1(1 - F_{1,\nu}(a))(1 - F_f(a))^N - 1 + \pi_2 - \pi_2 F_{1,\nu}(a) \ge 0$, or

$$\begin{aligned} F_{1,\nu}(a) \;&\le\; \frac{F_f(a) + \pi_1(1-F_f(a))^N - 1 + \pi_2}{\pi_1(1-F_f(a))^N + \pi_2} \\ &=\; 1 + \frac{F_f(a) - 1}{\pi_1(1-F_f(a))^N + \pi_2}. \end{aligned}$$

We obtain $1 - F_f(a) \le \delta(1 - F_{1,\nu}(a))$, where $0 < \delta = \pi_1(1-F_f(a))^N + \pi_2 < 1$. Therefore, $1 - F_f(a) \le 1 - F_{1,\nu}(a)$ and hence $F_f(a) \ge F_{1,\nu}(a)$. A similar argument proves the result for $F_{1,\nu}(a) \ge F_{2,\nu}(a)$.

Lemma 2 *The function $w_\alpha(x)$ defined in Equation (16) increases monotonically from zero at $x = 0$ to one at $x = 1$. For $\alpha < 1$, $F_{1,\nu} \ge F_{2,\nu}$, and for $\alpha > 1$, $F_{1,\nu} \le F_{2,\nu}$.*

Proof. For a given α, we differentiate $w_\alpha(x)$ w.r.t. x:

$$w'_\alpha(x) = \begin{cases} -1/(1-\alpha)\,\alpha^x \ln(\alpha), & \text{if } \alpha \neq 1 \\ 1, & \text{if } \alpha = 1. \end{cases}$$

Since $-\ln(\alpha)/(1-\alpha)$ is negative for $\alpha > 1$ and for $\alpha < 1$, $w'_\alpha(x)$ is always positive.

For a given $x \in [0,1]$, we also differentiate $w_\alpha(x)$ w.r.t. $\alpha \neq 1$:

$$\frac{dw_\alpha(x)}{d\alpha} = \frac{-x\alpha^{x-1}}{1-\alpha} + \frac{1-\alpha^x}{(1-\alpha)^2}.$$

Hence $dw_\alpha(x)/d\alpha \leq 0$ if and only if $d_x(\alpha) \equiv -x\alpha^{x-1}(1-\alpha) + 1 - \alpha^x \leq 0$. Since $d_x(1) = 0$, $dw_\alpha(x)/d\alpha \leq 0$ if and only if $d'_x(\alpha) \geq 0$ for $\alpha \leq 1$ and $d'_x(\alpha) \leq 0$ for $\alpha \geq 1$. Since

$$d'_x(\alpha) = x(1-x)\alpha^{x-2}(1-\alpha),$$

the latter is indeed the case.

Proposition 6 *Let F_f be some distribution function with support on $[\underline{f}, \overline{\nu}]$. For $N \geq 1$, $\pi_0, \pi_1, \pi_2 > 0$ and $\pi_1 N \leq 1$, $F_{2,\nu}$ given by Equation (17) is a well-defined distribution function as long as $F_f(\underline{f}) \geq x$, where $x \in (0,1)$ solves $x + \pi_1(1-x)^N - (1-\pi_2) = 0$. $F_{2,\nu}$ has support on $[\underline{f}, \overline{\nu}]$ and is unique.*

Proof. What we have to show is that

(a) $F_{2,\nu}(\overline{\nu}) = 1$ (immediate)

(b) $F_{2,\nu}(\underline{f}) \geq 0$

(c) $F_{2,\nu}(a)$ is monotonically increasing in a for $a \in [\underline{f}, \overline{\nu}]$

(d) $F_{2,\nu}$ is unique.

With regard to (c), let $y = F_{2,\nu}(a)$ and $x = F_f(a)$. According to Equation (17),

$$x = \pi_0 + \pi_1 \left\{1 - (1 - w_\alpha(y))(1-x)^N\right\} + \pi_2 y. \quad \text{(A.1)}$$

Differentiation w.r.t. x yields

$$1 = \pi_1 \left(w'_\alpha(y) y' (1-x)^N + (1 - w_\alpha(y)) N (1-x)^{N-1} \right) + \pi_2 y',$$

or

$$y' = \frac{1 - \pi_1 (1 - w_\alpha(y)) N (1-x)^{N-1}}{\pi_1 w'_\alpha(y) (1-x)^N + \pi_2}.$$

Because $w'_\alpha(y) > 0$ (see proof of Lemma 2), we have $\pi_1 w'_\alpha(y)(1-x)^N + \pi_2 > 0$. Hence $y' \geq 0$ as long as $\pi_1 N \leq 1$.

With regard to (b), note that for any given x, the derivative of the right-hand side of Equation (A.1) w.r.t. y is strictly positive. (Recall that $w'_\alpha(y) > 0$.) Hence for any given x, there is at most one value of y that satisfies Equation (A.1). According to Lemma 1 there exists a unique $x \in (0,1)$ such that $x + \pi_1(1-x)^N - (1 - \pi_2) = 0$. For this value of x, y equals zero. And since $y' \geq 0$, we obtain $F_{2,\nu}(\underline{f}) \geq 0$ as long as $F_f(\underline{f}) \geq x$.

With regard to (d), for a given $x = F_f(a)$ suppose that $y < \overline{y}$. Because $w'_\alpha(y) > 0$, we have $-\pi_1(w_\alpha(\overline{y}) - w_\alpha(y))(1-x)^N < \pi_2(\overline{y} - y)$ and hence

$$\pi_1 w_\alpha(y)(1-x)^N + \pi_2 y < \pi_1 w_\alpha(\overline{y})(1-x)^N + \pi_2 \overline{y},$$

meaning that there can only be one solution to Equation (A.1).

B. Results for FRMs and Loans Originated in Florida, New York or Illinois

Table B.1: **Expected broker fees and γ for 30-year FRMs** The table reports, in rows identified by "full", "partial" and "none", expected broker fees (in $1,000) for fully informed, partially informed and uninformed borrowers. The also table reports, in rows identified by "γ", the expected value of γ in Equation (2) when the borrower is partially informed. The data include all 30-year fixed-rate mortgages originated in California and are stratified by loan purpose and documentation level. The state vector x is specified as in Table 4 and $N = 1$. The MC estimates are based on 100,000 scenarios.

Docs	Info	bm	large	small	80% wt	50% wt	bm	large	small	80% wt	50% wt
						$\pi_2 = 0.75$					
	full	0.6	0.6	0.6	0.4	0.8	0.9	0.9	1.0	0.9	1.0
Full	partial	2.5	2.7	2.3	2.3	2.7	3.4	3.6	3.1	3.2	3.5
	none	3.4	3.8	3.1	3.2	3.6	4.5	5.0	4.0	4.4	4.7
	γ	0.75	0.74	0.76	0.74	0.75	0.75	0.74	0.76	0.74	0.75
	full	1.0	1.0	1.0	0.9	1.0	0.8	0.7	0.9	0.6	1.0
Low	partial	3.0	3.3	2.7	2.8	3.2	3.2	3.5	3.0	3.0	3.5
	none	3.9	4.4	3.5	3.6	4.2	4.4	4.9	3.9	4.1	4.6
	γ	0.76	0.76	0.76	0.76	0.76	0.75	0.74	0.76	0.75	0.75
						$\pi_2 = 0.5$					
	full	0.6	0.6	0.6	0.4	0.8	0.9	0.9	1.0	0.9	1.0
Full	partial	2.6	2.9	2.4	2.4	2.9	3.5	3.8	3.2	3.4	3.7
	none	3.7	4.2	3.3	3.5	3.9	4.9	5.5	4.3	4.8	5.0
	γ	0.71	0.70	0.72	0.71	0.71	0.71	0.71	0.72	0.71	0.72
	full	1.0	1.0	1.0	0.9	1.0	0.8	0.7	0.9	0.6	1.0
Low	partial	3.1	3.5	2.8	2.9	3.3	3.4	3.7	3.1	3.1	3.6
	none	4.2	4.7	3.7	3.9	4.6	4.8	5.3	4.2	4.5	5.0
	γ	0.72	0.72	0.73	0.73	0.72	0.72	0.71	0.73	0.71	0.72
						$\pi_2 = 0.25$					
	full	0.6	0.6	0.6	0.4	0.8	0.9	0.9	1.0	0.9	1.0
Full	partial	2.8	3.1	2.5	2.6	3.0	3.7	4.1	3.4	3.6	3.9
	none	4.3	4.8	3.8	4.1	4.5	5.6	6.3	4.9	5.4	5.7
	γ	0.65	0.64	0.67	0.65	0.66	0.65	0.64	0.67	0.65	0.66
	full	1.0	1.0	1.0	0.9	1.0	0.8	0.7	0.9	0.6	1.0
Low	partial	3.3	3.6	2.9	3.1	3.5	3.6	4.0	3.3	3.4	3.9
	none	4.8	5.4	4.2	4.4	5.2	5.4	6.2	4.7	5.2	5.6
	γ	0.66	0.66	0.67	0.67	0.66	0.66	0.65	0.67	0.65	0.66

Table B.2: **Expected broker fees and γ for 2/28 loans originated in Florida** The table reports, in rows identified by "full", "partial" and "none", expected broker fees (in $1,000) for fully informed, partially informed and uninformed borrowers. Rows identified by "overall" report the expected fees across all borrowers. The also table reports, in rows identified by "γ", the expected value of γ in Equation (2) when the borrower is partially informed. The data include all 2/28 mortgages originated in Florida and are stratified by loan purpose and documentation level. The state vector x is specified as in Table 4 and $N = 1$. The MC estimates are based on 100,000 scenarios.

		Purchase					Refinance				
Docs	Info	bm	large	small	80% wt	50% wt	bm	large	small	80% wt	50% wt
						$\pi_2 = 0.75$					
	full	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.3	0.3	0.4
Full	partial	2.5	2.8	2.2	2.4	2.6	3.5	3.9	3.0	3.3	3.6
	none	3.8	4.3	3.3	3.7	3.9	5.0	5.8	4.3	4.9	5.2
	overall	3.4	3.9	3.0	3.3	3.5	4.6	5.3	3.9	4.4	4.8
	γ	0.71	0.71	0.72	0.71	0.72	0.73	0.73	0.74	0.73	0.74
	full	0.3	0.3	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.4
Low	partial	2.5	2.7	2.3	2.4	2.6	2.5	2.8	2.2	2.4	2.6
	none	3.8	4.3	3.4	3.7	3.9	3.8	4.3	3.3	3.7	3.9
	overall	3.4	3.8	3.0	3.3	3.6	3.4	3.9	3.0	3.3	3.5
	γ	0.72	0.71	0.72	0.71	0.72	0.71	0.71	0.72	0.71	0.72
						$\pi_2 = 0.5$					
	full	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.3	0.3	0.4
Full	partial	2.7	3.0	2.4	2.6	2.8	3.7	4.2	3.1	3.5	3.8
	none	4.2	4.8	3.6	4.1	4.3	5.6	6.5	4.7	5.4	5.8
	γ	0.68	0.68	0.69	0.68	0.68	0.70	0.69	0.71	0.70	0.70
	full	0.3	0.3	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.4
Low	partial	2.7	3.0	2.4	2.6	2.8	2.7	3.0	2.4	2.6	2.8
	none	4.3	4.8	3.7	4.1	4.4	4.2	4.8	3.6	4.1	4.3
	γ	0.68	0.67	0.69	0.68	0.68	0.68	0.68	0.69	0.68	0.68
						$\pi_2 = 0.25$					
	full	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.3	0.3	0.4
Full	partial	2.9	3.3	2.6	2.8	3.0	4.0	4.5	3.4	3.8	4.1
	none	5.0	5.7	4.2	4.9	5.1	6.5	7.6	5.4	6.4	6.7
	γ	0.62	0.62	0.63	0.62	0.62	0.64	0.63	0.65	0.64	0.65
	full	0.3	0.3	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.4
Low	partial	2.9	3.2	2.6	2.8	3.1	2.9	3.3	2.6	2.8	3.0
	none	5.0	5.7	4.4	4.9	5.2	5.0	5.7	4.2	4.9	5.1
	γ	0.62	0.61	0.63	0.62	0.62	0.62	0.62	0.63	0.62	0.62

Table B.3: **Expected broker fees and γ for 2/28 loans originated in New York** The table reports, in rows identified by "full", "partial" and "none", expected broker fees (in $1,000) for fully informed, partially informed and uninformed borrowers. Rows identified by "overall" report the expected fees across all borrowers. The also table reports, in rows identified by "γ", the expected value of γ in Equation (2) when the borrower is partially informed. The data include all 2/28 mortgages originated in New York and are stratified by loan purpose and documentation level. The state vector x is specified as in Table 4 and $N = 1$. The MC estimates are based on 100,000 scenarios.

		Purchase					Refinance				
Docs	Info	bm	large	small	80% wt	50% wt	bm	large	small	80% wt	50% wt
						$\pi_2 = 0.75$					
	full	0.0	0.0	0.0	0.0	0.0	0.5	0.6	0.5	0.4	0.7
Full	partial	2.2	2.7	1.7	2.0	2.4	3.7	4.5	2.9	3.5	3.9
	none	3.4	4.2	2.7	3.2	3.7	5.1	6.1	4.0	4.9	5.2
	overall	3.0	3.7	2.4	2.8	3.3	4.7	5.6	3.7	4.5	4.8
	γ	0.70	0.71	0.70	0.70	0.71	0.76	0.76	0.76	0.76	0.76
	full	0.6	0.6	0.6	0.5	0.6	0.7	0.7	0.7	0.5	1.0
Low	partial	3.6	4.1	3.2	3.3	3.9	4.2	4.8	3.7	4.0	4.4
	none	4.9	5.6	4.3	4.6	5.3	5.3	6.1	4.5	5.1	5.5
	overall	4.6	5.2	3.9	4.2	4.9	4.9	5.7	4.2	4.7	5.2
	γ	0.76	0.75	0.77	0.75	0.76	0.80	0.79	0.81	0.80	0.80
						$\pi_2 = 0.5$					
	full	0.0	0.0	0.0	0.0	0.0	0.5	0.6	0.5	0.4	0.7
Full	partial	2.3	2.9	1.8	2.1	2.5	3.9	4.7	3.1	3.7	4.1
	none	3.9	4.7	3.0	3.6	4.1	5.5	6.7	4.3	5.4	5.7
	γ	0.67	0.67	0.67	0.66	0.68	0.73	0.73	0.73	0.72	0.73
	full	0.6	0.6	0.6	0.5	0.6	0.7	0.7	0.7	0.5	1.0
Low	partial	3.8	4.3	3.3	3.5	4.1	4.4	4.9	3.8	4.1	4.6
	none	5.4	6.1	4.6	5.0	5.7	5.6	6.5	4.7	5.4	5.8
	γ	0.73	0.72	0.74	0.72	0.74	0.77	0.77	0.79	0.77	0.78
						$\pi_2 = 0.25$					
	full	0.0	0.0	0.0	0.0	0.0	0.5	0.6	0.5	0.4	0.7
Full	partial	2.6	3.2	2.0	2.4	2.8	4.2	5.0	3.3	4.0	4.3
	none	4.6	5.6	3.6	4.3	4.8	6.2	7.6	4.9	6.1	6.4
	γ	0.62	0.62	0.61	0.61	0.62	0.68	0.68	0.67	0.67	0.68
	full	0.6	0.6	0.6	0.5	0.6	0.7	0.7	0.7	0.5	1.0
Low	partial	4.1	4.6	3.6	3.7	4.4	4.6	5.2	3.9	4.3	4.8
	none	6.1	7.0	5.2	5.7	6.4	6.2	7.2	5.2	6.0	6.4
	γ	0.68	0.67	0.69	0.66	0.69	0.73	0.72	0.75	0.73	0.73

Table B.4: **Expected broker fees and γ for 2/28 loans originated in Illinois** The table reports, in rows identified by "full", "partial" and "none", expected broker fees (in $1,000) for fully informed, partially informed and uninformed borrowers. Rows identified by "overall" report the expected fees across all borrowers. The also table reports, in rows identified by "γ", the expected value of γ in Equation (2) when the borrower is partially informed. The data include all 2/28 mortgages originated in Illinois and are stratified by loan purpose and documentation level. The state vector x is specified as in Table 4 and $N = 1$. The MC estimates are based on 100,000 scenarios.

		Purchase					Refinance				
Docs	Info	bm	large	small	80% wt	50% wt	bm	large	small	80% wt	50% wt
						$\pi_2 = 0.75$					
	full	0.2	0.2	0.3	0.2	0.3	0.4	0.4	0.5	0.3	0.5
Full	partial	1.6	1.5	1.6	1.5	1.7	2.1	2.2	1.9	2.0	2.2
	none	2.5	2.6	2.4	2.4	2.6	3.1	3.5	2.7	3.0	3.2
	overall	2.2	2.3	2.2	2.1	2.3	2.8	3.1	2.5	2.7	2.9
	γ	0.70	0.69	0.71	0.69	0.70	0.71	0.70	0.73	0.71	0.71
	full	0.4	0.4	0.4	0.4	0.4	0.2	0.2	0.3	0.1	0.3
Low	partial	1.9	2.0	1.9	1.8	2.0	2.4	2.4	2.3	2.3	2.5
	none	2.8	3.0	2.6	2.7	2.9	3.4	3.6	3.2	3.3	3.6
	overall	2.6	2.7	2.4	2.5	2.7	3.1	3.3	3.0	3.0	3.2
	γ	0.71	0.70	0.73	0.71	0.72	0.74	0.73	0.75	0.74	0.74
						$\pi_2 = 0.5$					
	full	0.2	0.2	0.3	0.2	0.3	0.4	0.4	0.5	0.3	0.5
Full	partial	1.7	1.7	1.7	1.6	1.8	2.2	2.4	2.0	2.1	2.3
	none	2.8	2.9	2.7	2.7	2.9	3.4	3.9	3.0	3.3	3.5
	γ	0.66	0.64	0.68	0.65	0.66	0.68	0.66	0.69	0.67	0.68
	full	0.4	0.4	0.4	0.4	0.4	0.2	0.2	0.3	0.1	0.3
Low	partial	2.0	2.1	2.0	1.9	2.1	2.5	2.6	2.5	2.4	2.6
	none	3.1	3.4	2.9	3.0	3.2	3.8	4.1	3.5	3.7	3.9
	γ	0.68	0.66	0.69	0.67	0.68	0.70	0.69	0.72	0.70	0.71
						$\pi_2 = 0.25$					
	full	0.2	0.2	0.3	0.2	0.3	0.4	0.4	0.5	0.3	0.5
Full	partial	1.9	1.9	1.9	1.8	2.0	2.4	2.6	2.2	2.3	2.5
	none	3.4	3.5	3.2	3.3	3.4	4.0	4.6	3.4	3.9	4.1
	γ	0.59	0.58	0.61	0.59	0.60	0.62	0.60	0.64	0.61	0.62
	full	0.4	0.4	0.4	0.4	0.4	0.2	0.2	0.3	0.1	0.3
Low	partial	2.2	2.3	2.1	2.1	2.3	2.7	2.8	2.6	2.6	2.8
	none	3.7	4.0	3.3	3.6	3.8	4.4	4.8	4.0	4.3	4.5
	γ	0.62	0.60	0.63	0.61	0.62	0.65	0.63	0.67	0.65	0.65