Asymmetric information and the supply-chain of mortgages: The case of Ginnie Mae loans*

Jonathan Becker[1], Kenneth Hendricks[1,2], Jean-François Houde[1,2] and Diwakar Raisingh[3]

[1]University of Wisconsin-Madison
[2]NBER
[3]Department of Justice

February 6, 2023

1 Introduction

The voluminous literature studying the mortgage market has focused mostly on the retail (origination) segment in which financial intermediaries acquire loans from borrowers. According to the traditional banking model, vertically integrated financial institutions (traditional banks) fund mortgages using deposits and then keep the resulting loans on their balance sheets. Retail margins capture the difference between the interest rate on loans and the marginal cost of attracting deposits. In practice, in the U.S., only a relatively small share of conforming loans are funded this way.[1] Instead, the majority of conforming loans made by traditional banks are eventually pooled into mortgage-backed securities (MBS) and sold on secondary markets to investors such as pension funds, hedge funds or foreign banks. The banks retain the servicing rights and earn a monthly fee for collecting and distributing the monthly payments of the borrower. This fee is equal to the note rate of the loan minus the agency's fee for insuring the loan against default and the coupon paid to the investors. This structure corresponds to the *originate-to-distribute* model. Securitization greatly increases the supply of funds to the U.S. mortgage market. It also opens the door to non-depository mortgage specialists (brokers and correspondent lenders) that compete with traditional banks to attract borrowers. For mortgage specialists, the primary funding source is the wholesale market, in which other financial institutions (both traditional and shadow banks) compete to acquire loans via auctions or posted-prices.

*Pleminary and incomplete. Please do not cite without the permission of the authors.

[1]Loans are said to be conforming if they satisfy the underwriting criteria of three federal agencies Freddie Mac, Fannie Mae, and Ginnie Mae.

This paper provides the first comprehensive analysis of the originate-to-distribute supply chain for conforming mortgages. We advance a simple modeling framework for loan valuation in the wholesale and secondary markets for mortgages. The prices that banks are willing to pay for loans in the wholesale market depend upon their resale prices in the secondary market, and on the stream of fees that they can earn from servicing the loans. These prices in turn depend on the duration of the loan: the flow of payments to the bank for servicing the loan and to the investors for funding the loan end when the loan is prepaid, either by the borrower or by the agency in the event of default. Thus, early prepayment is the main source of risk that banks and investors face. The main friction is private information about this risk: originators who sell in the wholesale market have more information about loan duration than the banks, and the banks have more information about loan duration than investors when they sell the loans in the secondary market. These asymmetries in information can give rise to adverse selection, increase the costs of intermediation, and prevent the efficient flow of funds.

The main goals of this paper are to validate our model of loan valuation, test for adverse selection in the two markets, and quantify their effects. The empirical analysis requires data covering the entire lifespan of a loan, from origination, to possible sale in the wholesale market, to sale in the secondary market, and to payment history and duration. Until now, such data have been unavailable to researchers, but we have been able to obtain access to three data sources that, once combined, allow us to construct the full history of a loan. A unique aspect of our study is the use of proprietary data on loan acquisition by auction from the FinTech company Optimal Blue (OB), which operates one of the largest loan exchange platforms. Sale by auction has facilitated entry into the primary market of lenders with limited capital who specialize in mortgage origination, and is a rapidly growing segment of the wholesale market.

We use the data on loan duration and auctions to establish several important facts..First, loan survival is positively correlated with loan and borrower characteristics. We estimate the probability of a loan surviving for different periods and find that the characteristics can explain a significant percentage of the variation in loan survival. Second, bidders value loan duration. We regress bids on the loan characteristics observed by bidders and find they bid more for loans with longer expected duration. Third, the auction is a common value auction in which the bidders are asymmetrically and privately informed about loan survival. Using the residuals of the bid regressions as measures of the bidders' signals, we find that a bidder's signal, and the maximum among it's rivals, are positively correlated with loan survival. The correlation with own signal is lower than the correlation with the maximum rival signal, and varies across the bidders, which suggests that the bidders are differentially informed. These facts provide strong support for our model of valuations.

Buyers in the wholesale market observe a subset of the characteristics observed by originators selling the loan. For example, the buyers do not observe the borrower's age, race, gender, or choice of fees, which are significant predictors of loan duration. Since these characteristics are not individually priced by the buyers, sellers have an incentive to use their private information to keep and securitize loans that they believe are more likely to survive, and to sell the others. Bidders in turn should respond to this adverse selection by lowering their bids. We use exogenous variation in the capacity of sellers to act on this incentive to determine whether, and by whom, the loans sold in the wholesale market are adversely selected.

A similar situation arises in the secondary market. Banks with a large volume of loans can choose to sell their loans in a To-Be-Announced (TBA), multi-issuer security or in a custom pool security. The TBA market is a highly liquid, forward market in which a seller and a buyer agrees to trade a volume of loans at a specified price and future date. Importantly, the investors do not observe the characteristics of the individual loans, because the loans are not selected at the time of the trade. The custom pool market is less liquid, but investors observe the characteristics of the individual loans in the custom pool security when the trade is made. As a result, the large banks have an incentive to sell loans that are less likely to survive in the TBA market, and loans that more likely to survive in the custom pool market, where they can be sold at higher price.

Banks must also decide whether to sell the loan in high coupon security or a low coupon security. They can get more money up front by selling a loan in a high coupon security (i.e., high price, low fee) or more money later by selling the loan in a low coupon security (i.e., low price, high fee). Clearly, this choice depends on the bank's assessment of the loan's duration. It should place loans that are more likely to survive in a low coupon security, and loans that are more likely to be prepaid early in a high coupon security. We test this hypothesis for loans sold in the TBA market only since we observe the security prices.in this market but not in the custom pool market.

To assess the importance of adverse selection in the secondary market, we construct a test that is commonly used in the insurance literature (Chiappori & Salanie 2000). Specifically, we test whether the 12-month survival rate is higher for loans placed in custom pool securities than in TBA securities, and for loans placed in low coupon TBA securities than for high coupon TBA securities. The results strongly support the hypothesis of adverse selection, both on observables and unobservables. The magnitude of the adverse selection effects are especially large for pool choice, even after controlling for loan characteristics.

The correlation between coupon choice and survival can be due to moral hazard. Lenders may be causing borrowers to refinance their loans early in order to obtain higher service income on the new loan. We explore this hypothesis using the sample of Ginnie Mae loans where the coupon is uniquely determined by the note rate, so banks have no choice. We find that the unconditional correlation between survival and service income is indeed positive but small, and it is essentially zero.once we condition on loan attributes. Therefore, we cannot reject moral hazard on observables but we can reject it on unobservables, which suggests that the correlation between coupon choice and survival is likely due to adverse selection.

This paper contributes to three important literatures at the intersection of Industrial Organization and Finance. First, it contributes to a large literature measuring the importance of adverse selection in financial and insurance markets.[2] Second, it builds on a growing number of papers analyzing the Industrial Organization of the mortgage industry (e.g. Stanton et al. (2014), Allen et al. (2019), Robles-Garcia (2019), Buchak et al. (2020)).

Finally, it contributes to the empirical auction literature; in particular we use insights from important empirical papers studying the design of auctions in financial markets, and auctions with common values (e.g. Hendricks et al. (2003), Haile (2002), Hortaçsu & McAdams (2010), Hortaçsu & Kastl (2012), Cassola et al. (2013)).

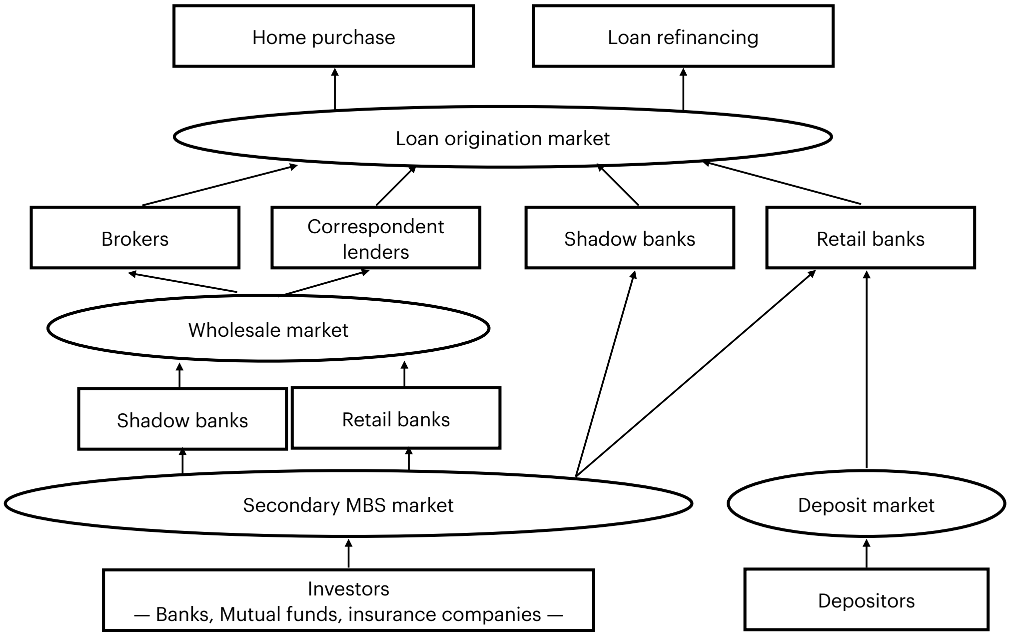
2 Supply Chain of Mortgages in the US

Figure 1 illustrates the flow of funds from investors and depositors to borrowers. In the retail market, banks and mortgage specialists provide loans to borrowers seeking to buy a home or refinance an existing mortgage. Borrowers are presented with a rate sheet of note rates and upfront payments, known as "discount points", that specify the amount that the borrower would need to pay at closing to lower the note rate (andreas et al. 2013). Based on these rate sheets, the borrower selects a lender and a note rate. The lender must then decide whether to keep the loan on its books, sell the loan to another lender in the wholesale market, or securitize the loan and sell it in the secondary market. Loans that are originated, securitized, and sold in the secondary market by vertically integrated banks are called *retail* loans. Loans that are sold in the wholesale market and then securitized and resold in the secondary market are called *non-retail* loans. The wholesale channel is split between brokers and correspondent lenders. A *broker* matches a borrower to a bank who underwrites and funds the loan at closing. A *correspondent lender* is a mortgage specialist or bank who underwrites and funds the mortgage at closing and then sells the loan to an ""aggregator" or sponsor bank for securitization, typically a few days after the closing date. In our sample of Ginnie Mae loans, which covers the period from 2013 to 2022, roughly 40% of all securitized loans were originated by correspondent lenders, and 10% by brokers. The remaining 40% of loans were retail.

The traditional way that correspondent lenders sold their loans was in a posted-price market. Buyers post wholesale rate sheets daily that specify the prices they were willing to pay for loans. These prices are known as*lock* prices because they vary with the lock-in period, which can be0, 30, 60, or 90 days. They are also seller-specific, and contingent on a very coarse binning of loan characteristics. Given these characteristics and a lock-in period, the correspondent lender selects a buyer, and that buyer then agrees to buy the loan shortly *after* closing. With this sales mechanism, both the seller and the buyer incur holding costs. They also bear the risk of "fall-out" which are loan applications that do not result in a closed loan, either because the borrower does not qualify

[2] Closely related to this paper include the work of Agarwal et al. (2012) and Downing et al. (2009). See Einav et al. (2010) for a review of the insurance literature.

Figure 1: Flow of funds in the mortgage industry



or because they turn down the offer.

More recently, online platforms such as *Optimal Blue* have provided correspondent lenders with the option to sell their loans individually in online auctions. Buyers in the auction have the same loan information they would have in the posted price market. However, the auction allocation is likely to be more efficient (i.e., the buyer with the highest valuation is more likely to get the loan), since the bids that buyers submit are a continuous function of loan characteristics. The auction may also be less costly, because only the seller incurs holding costs. The data do not allow us to distinguish between posted price sales and auction sales, but auction sales represent a growing segment of the wholesale market. Specifically,auction sales on the OB platform accounted for over 75% of loan sales in2019.

Most mortgages are securitized. The securitization process involves pooling many different loans and issuing a mortgage-backed security (MBS) backed by these loans. The security is then sold (in tranches) to investors such as foreign banks, hedge funds, and pension funds. In the case of conforming loans, the GSEs, Fannie Mae and Freddie Mac, purchase the loans from the banks, insure them against default, and issue the MBSs. In the case of government-insured loans, Ginnie Mae insures the loans against default, but does not issue securities directly. Banks are responsible for delivering and managing loan packages, subject to Ginnie's underwriting rules. In both cases, the banks typically retain the servicing rights to the loans. They collect the interest payment from the borrower each month, pay the MBS coupon to investors and the guarantee fee to the agency, and keep the difference, known as the *service income*. Both investors and banks bear the risk of loan prepayment.

Most agency MBSs are sold in the To-Be-Announced (TBA) market. TBA trades are forward contracts: a seller and buyer agree to trade a volume of loans(par value) of agency MBS at a specified price and future date (settlement date). The time between the date of the trade and date of the settlement is typically several weeks. When the trade is made, the mortgages in the MBS are not known, in part because they often do not yet exist. Instead, the two parties agree upon the issuer (agency guarantor), coupon, and maturity of the loans. In selecting the pool of loans, the seller has an incentive to deliver loans that are more likely to default or be refinanced. However,according to Vickery & Wright (2013), the buyer understands this incentive,anticipates that loans will be adversely selected, and prices accordingly.[3] In principle, the market could unravel, but the lack of loan information makes the market more homogeneous and liquid, and the liquidity premium more than offsets the adverse selection discount (see Vickery & Wright (2013)). An important benefit of the TBA market for lenders in the primary and wholesale markets is that it locks in a resale price for the loans that they originate or buy. Banks with large volumes of loans can choose to sell their loans in another,less liquid, securities market. It is called the "specified-pool" or "custom-pool" market, because the characteristics of the loans in the agency MBS are known when the trade is made. During our sample period, roughly 20% of the securitized loans are pooled in custom TBA-eligible securities.

Some banks such as Quicken Loans or Bank of America rely almost exclusively on the retail channel to acquire loans. In contrast, several large shadow-banks such as Pennymac originate a very small quantity of loans directly, and rely on the wholesale market to acquire loans. However, most banks and shadow banks manage a diversified portfolio of retail and non-retail loans.

As originators in the retail market, these lenders have to decide whether to keep a loan or sell it in the wholesale market and, as buyers in the wholesale market, they have to set lock prices in the posted price market and submit bids in the auction market. As sellers in the secondary market,the banks need to decide whether to sell in the TBA or the specified pool market and, in these markets, choose the coupon for their loan. In what follows, we study these decisions in the context of the auction market.

3 Data

In this section, we describe the three main sources of data used in the empirical analysis and how we track loans across these data sets. We focus on 30-year fixed rate mortgages insured by Ginnie Mae, a public corporation responsible of insuring default risks for loans qualifying for FHA, VA, and

[3] This price is known in the finance literature as the "cheapest-to-deliver" price.

rural housing subsidies.[4] Because Ginnie Mae guarantees loans with higher LTV ratios, borrowers tend to be riskier (both in terms of default and prepayment). FHA borrowers in particular are often first-time home-buyers who eventually transit to conventional products after building enough equity. As a result, these loans cannot be sold to the GSEs, Fannie Mae and Freddie Mac, which implies that this segment operates more or less independently of the other segments. The other reason for focusing on Ginnie Mae loans is that Ginnie Mae does not discriminate between lenders when setting its guarantee fee. It is fixed exogenously and the same for all lenders.[5]

3.1 Loan Securitization and Performance

The first data set, *eMBS*, provides detailed information all mortgage backed securities insured by one of the three agencies that were issued from January 2013 to the present and their component loans. In our analysis, we use data Ginnie Mae securities issued between October 2013 and December 2019. The characteristics of the MBS include the CUSIP (security identifier), coupon rate, issuance date, issuer and servicer identity, maturity, and par amount. The characteristics of the loans include: the CUSIP with which it is associated, the subsidizing agency, the loan type (purchase, refinance, etc.), original principal balance, note rate, loan-to-value (LTV), debt to income ratio, FICO score, number of units on the property, state, origination type (retail or non-retail), and the identity of the issuer which, in case of Ginnie Mae, is the servicer. For each component loan, we observe the unpaid principal balance on a monthly basis until it has either been paid off or defaulted. We use the unpaid balance for each loan in March 2022 to measure the loan duration until full payment.

We use the above information to infer the service income earned each month by the servicer. In regards to the security prices, we do not observe the prices of the custom pool securities but are able to obtain the daily agency TBA MBS prices from Bloomberg.[6]

Table 1 1 provides summary statistics on the characteristics of the loans. The sample consists of 30-year, fixed rate Ginnie Mae loans issued between October 2013 to December 2019 that had at least 6 months of performance data.[7] Note rates are typically quoted on a 1/8 percentage points (p.p) grid. The note rate varies between 3% and 5 % and averages 4%. The average loan size is $210K, but it varies a lot. The loans typically have a very high LTV, with 58% of them having

[4]This term is the most common, accounting for 93% of the loans in our data. The remainder is divided between fixed rate mortgages with different maturities (6%) and adjustable rate mortgages (1%).

[5]The GSE's also charge a fixed, monthly g-fee, but they allow the lender to "buy down" the fee by converting the flow into an upfront payment or, alternatively, to "buy up" the fee and receive an upfront payment from the GSE (see andreas et al. (2013)).

[6]Bloomberg sources its pricing data from Trade Reporting and Compliance Engine (TRACE), which is a database of trades maintained by Financial Industry Regulatory Authority (FINRA). This database contains the universe of TBA bond trades for which one of the parties was registered with FINRA. TBA MBS trades are typically made with a FINRA registered dealer so TRACE should contain nearly all trades (see Gao et al. (2017)). The daily price of a TBA security corresponds to the last observed trading price as of that date.

[7]We truncate the sample to avoid the Covid shock to the markets.

Table 1: Summary of Mortgages in Ginnie Mae MBS's

(a) All Loans: 2013-2019

VARIABLES	Mean	SD	P-10	P-90
Note rate	4.2	.56	3.5	4.9
Loan (x100K)	2.2	1.1	1	3.5
LTV	95	8.4	85	101
Credit Score	688	54	625	769
DTI	41	9.6	28	53
1(DTI > 40)	.58	.49	0	1
1(VA)	.34	.47	0	1
1(Second lien)	.06	.24	0	0
1(Purchase)	.76	.43	0	1
1(Retail)	.39	.49	0	1
1(Corr.)	.49	.5	0	1
Survival: 12m	89	31	0	100
Survival: 36m	57	50	0	100
Observations	751794			

(b) Matched sample: 2018-2019

VARIABLES	Mean	SD	P-10	P-90
Note rate	4.4	.61	3.5	5.1
Loan (x100K)	2.3	1	1.2	3.6
LTV	96	7	86	102
Credit Score	687	52	626	767
DTI	43	10	30	50
1(DTI > 40)	.63	.48	0	1
1(VA)	.29	.45	0	1
1(Second lien)	.019	.14	0	0
1(Purchase)	.83	.37	0	1
1(Retail)	.0023	.048	0	0
1(Corr.)	.97	.16	1	1
Survival: 12m	82	38	0	100
Survival: 36m.	30	46	0	100
Observations	53843			

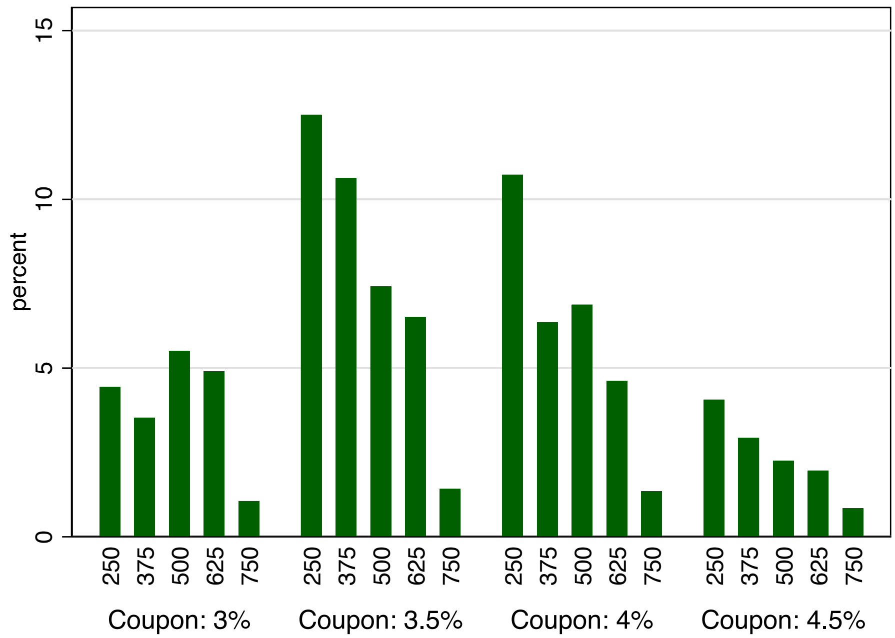
Reports the summary statistics of loans that entered Ginnie Mae 30 Year Fixed Rate MBS's issued between October 2013 and December 2019 that had at least 12 months of performance data. LTV, debt to income ratio, and FICO scores are missing for some of the mortgages in our data set.

values between 95 and 100. Moreover, the FICO scores of the borrowers are a relatively low 687, which is just above 670, the cutoff between "fair" and "good" credit. The LTV's and FICO scores are consistent with the goal of the subsidy programs. The largest subsidy category is FHA loans. These loans increased in popularity after the financial crisis, replacing privately securitized subprime loans, because the FHA program allow borrowers with low credit-scores and/or high LTV ratios to access the mortgage market (although they do incur higher insurance payments over the life of the contract).

Coupon rates are quoted on a 1/2 p.p grid. Figure 2 gives the frequency distribution of the coupon rate of the Ginnie Mae securities in our sample. Most pools pay out a coupon that is between 3.5% to 5.0% and this accounts for over 90% of the pools in the data.

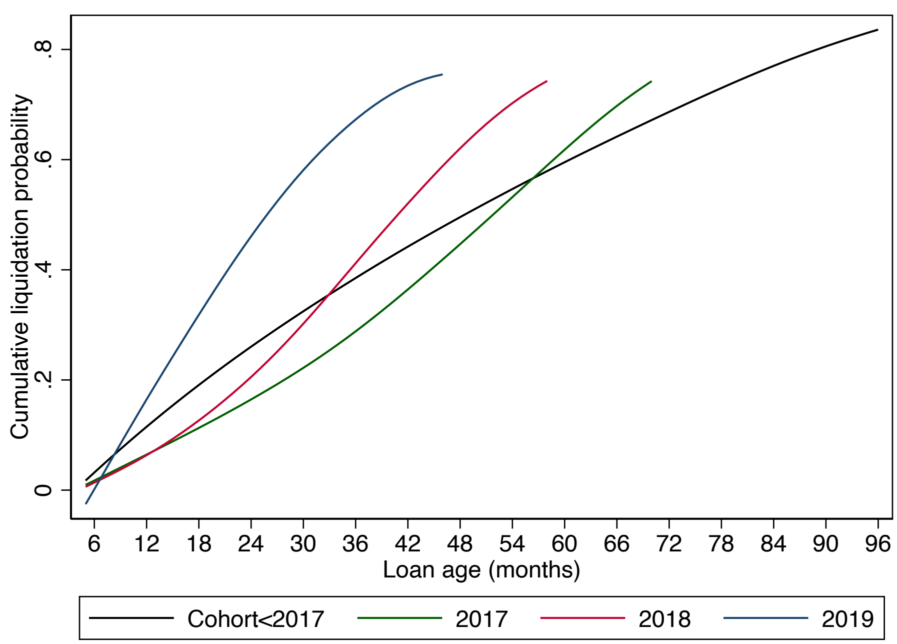
Since consumers can exert the option to pre-pay their loan early or default, very few loans last until maturity. Figure 3 quantifies this risk for the cohorts of Ginnie Mae loans originated between 2013 and 2019. The pre-2017 cohorts faced relatively stable interest rates. On average, 70% of those loans were pre-paid within the first six years, and about 10% were pre-paid within the first year. The risk of early pre-payment increased substantially for loans originated in 2018 and 2019, due to the dramatic decline in mortgage interest rates observed between 2019 and 2021. This affected especially the 2019 cohort. Nearly 50% of those loans were prepaid within the first

Figure 2: Density of Ginnie Mae's MBS coupons and service income between 2013 and 2019



The x-axis groups loans based on the MBS coupon (bottom number, c) and service income (top number, $r - c$). The sample excludes infrequently used coupons (≤ 2.5 and ≥ 5).

Figure 3: Evolution of pre-payment risk across cohorts



18 months, and 30% of loans were pre-paid during the first year. Since our data on auctioned loans covers the 2018-2020 period, we use the risk of early pre-payment as our primary measure of loan performance.

3.2 Auction Market

The mortgage auction data comes from Optimal Blue, a FinTech firm that operates the largest loan exchange platform in the market. Mortgage originators use this platform to sell loans to banks to free up capital that they can use to originate more mortgages. The OB auction is a first-price, sealed bid auction that lasts one to two hours. The seller usually invites all the buyers in its network of buyers (typically 8 to 15 buyers) to submit bids, called *bulk* bids, for the loan. For more specialized loans, the seller may invite fewer buyers. The set of potential bidders varies across lenders, but vary very little across time. Forming a relationship is costly because it involves both parties conducting due diligence as to the reliability and underwriting standards of their counterparty. In most auctions, all invited bidders submit bulk bids and nearly all do, because bidding is essentially cost-less and is a way of maintaining the relationship with the seller. A bidder can always submit a low bid if it does not want to purchase the mortgage.

An unusual feature of the auction is that the seller lender can always choose to sell the loan at the buyer's lock price if this price is higher than its bulk bid. Thus, from the seller's perspective, each buyer's actual bid is the maximum of its lock price and bulk bid, and the winning bid is the maximum of the bulk bids and lock prices. The other unusual feature of the auction is that, after observing the bids, the seller can decide not to sell the loan, in which case it either resells the loan in a later auction or sells it in the secondary market. This event is common for conforming loans, but not for Ginnie Mae loans, since the seller of these loans bears the costs of securitization.

We focus on 30-year, fixed rate mortgages eligible for Ginnie Mae insurance sold by auction during the period January 2018 to January 2019..For each mortgage, we (and the bidders) observe the following characteristics: original principal balance, loan-to-value ratio (LTV), note rate, loan type (purchase, refinance, etc.), property type, number of units, and zip code.

Since mortgages are sold individually, each mortgage is associated with an auction. For each auction, we observe the following variables: auction date, a seller id, the number of invited bidders, the value of each submitted bulk bid if it is above the bidder's lock price and the lock price if it is not, and the associated bidder id. The sellers and bidders have unique identifiers, but we can infer their identities from eMBS and our data set on the origination market.

The characteristics of the Ginnie Mae loans sold at auction are reported in Table 3. The summary statistics are quite similar to those reported in Table 1 with three key differences: (i) more loans are used to purchase a property as opposed to being refinanced, (ii) the note rate is substantially higher, and (iii) more loans have an LTV near 100. The higher prevalence of loans for purchasing a home is likely driving the other two differences. The larger note rate corresponds

Table 2: Survival probability regression

VARIABLES	(1) $1(T>12)$	(2) $1(T>36)$	(3) $1(T>12)$	(4) $1(T>12)$
Note rate	-9.53	-15.4	-10.1	-11.4
	(0.21)	(0.31)	(0.31)	(0.60)
Loan amount (x100K)	-6.24	-18.0	-9.11	-11.4
	(0.26)	(0.45)	(0.42)	(0.76)
Loan amount squared (x100K)	0.79	3.54	1.21	1.83
	(0.066)	(0.11)	(0.10)	(0.24)
1(VA)	-6.58	-5.55	-11.7	-14.2
	(0.29)	(0.29)	(0.40)	(0.64)
1(Second lien)	5.85	10.9	6.56	1.06
	(0.37)	(0.56)	(0.43)	(1.05)
LTV	0.044	0.12	0.070	0.25
	(0.0062)	(0.0092)	(0.011)	(0.032)
Credit score groups = 2, 630-690: Fair	-0.30	-1.11	-0.72	-2.20
	(0.14)	(0.21)	(0.25)	(0.56)
Credit score groups = 3, 690-720: Good	-0.87	-1.79	-1.95	-3.75
	(0.18)	(0.27)	(0.32)	(0.71)
Credit score groups = 4, 720-850: Excellent	-0.73	-0.91	-2.32	-4.37
	(0.19)	(0.29)	(0.34)	(0.71)
Loan type = 2, Refi: Not Streamlined, Not Cash Out	-4.05	-5.05	-5.08	-3.21
	(0.23)	(0.33)	(0.47)	(0.97)
Loan type = 3, Refi: Cash out	-6.26	-8.00	-7.10	-5.95
	(0.20)	(0.26)	(0.33)	(0.75)
Loan type = 4, Refi: Streamlined	-3.82	-3.20	-8.86	1.67
	(0.36)	(0.64)	(0.81)	(2.29)
1(DTI > 40)	-1.00	-1.89	-1.37	-2.00
	(0.074)	(0.12)	(0.15)	(0.33)
1(Retail)	0.35	1.10	-1.41	6.73
	(0.18)	(0.25)	(0.33)	(5.87)
1(Correspondent)	0.53	0.27	-0.57	-1.33
	(0.15)	(0.22)	(0.28)	(1.04)
Constant	140	140	147	139
	(1.48)	(1.95)	(2.29)	(4.40)
Observations	748,612	631,127	273,303	48,551
R-squared	0.128	0.194	0.152	0.160
Pool date FE	yes	yes	yes	yes
State FE	yes	yes	yes	yes
Issuer FE	yes	yes	yes	yes
Sample	2013-18	2013-18	2018-19	Matched
Mean dep. variable	89.1	57	83.1	82.3

Robust standard errors in parentheses

The sample in columns (1-3) correspond to a 10% random sample of Ginnie Mae loans issued between 2013 and 2019. Standard errors are clustered at the note-rate/issuing date level.

Table 3: Summary of Mortgages Sold at Auction

Variable	N	Mean	SD	Pctl10	Pctl90
Note Rate	42,852	4.660	0.472	4	5.250
Original Loan Amount	42,852	217,173	103,622	108,080	343,000
LTV	42,852	94.486	8.339	85	100
1(LTV $\in$ (75, 80])	42,852	0.024	0.152	0	0
1(LTV $\in$ (95, 100])	42,852	0.715	0.451	0	1
Debt to Income Ratio	42,852	42.232	10.018	29.280	54.480
Monthly Income	40,501	6,042	3,189	2,900	10,130
1(Monthly Income $\geq$ \$20K)	40,501	0.006	0.076	0	0
FICO	42,852	681.438	52.224	622	760
1(Purchase)	42,852	0.831	0.375	0	1
1(Retail)	42,852	0	0	0	0
1(Agency = FHA)	42,852	0.620	0.485	0	1
1(Agency = VA)	42,852	0.295	0.456	0	1
1(Agency = Rural Housing)	42,852	0.085	0.279	0	0
1(Paid off within 12 months)	39,889	0.111	0.314	0	1
i_PaidOffOrDelinquentInOneYear	39,889	0.122	0.328	0	1

to a more valuable loan because, all else equal, a higher interest rate leads more income.

The value of each bid corresponds to the wholesale price for a \$100 loan. A bid of \$100 corresponds to paying the par-value of a loan. The mortgages are typically securitized and sold on the TBA MBS market and the MBS's fetch prices above the par value of the loan (loans generate interest). The TBA MBS price effectively acts as a floor on the bids that the seller will accept since it knows that buyers will get at least that price when they securitize the loan. From a bidder's perspective, the TBA MBS price corresponds to an observable reserve price. There are some cases in which the seller accepts a price that is below the TBA MBS price and near the loan's par value. This can occur for loans that have an expected short duration (high prepayment risk) that will yield limited buyer interest. This is not to say that sellers lose money on those loans, since consumers typically pay upfront fees to the mortgage originators.

Table 4 summarizes the main outcome variables in the auction data. The bids are typically above the par value of the loan and the highest bid is typically well above the par value. The money left on the table (highest bid less the second highest bid) is on average 0.20. While this is a relatively small amount relative to the par value, it is quite substantial portion of the bid after we net out the MBS price.

The bids are nearly always above the MBS price which is consistent with the idea that it acts as an observable floor on the bids. In comparison, the lock price is effectively secret and it is common to see bids below the reserve.

Table 4: Summary of Bids

Variable	N	Mean	SD	Pctl10	Pctl90
# Bids in Auction	42,852	6.731	4.301	1	12
# Serious Bids in Auction	42,852	6.192	4.264	1	12
Bid	288,454	104.214	1.487	102.325	105.780
Serious Bid	265,333	104.345	1.331	102.614	105.826
Highest Bid	42,779	104.765	1.450	102.869	106.436
Money on Table[a]	32,518	0.237	0.467	0.020	0.534
Money on Table (Auctions $\geq$ 2 Serious Bids)	31,259	0.212	0.300	0.020	0.491
TBA MBS Price Associated with Loan[b]	42,779	103.180	1.349	101.703	104.891
Bid - TBA MBS Price	288,454	1.070	0.994	0.109	2.040
Serious Bid - TBA MBS Price	265,333	1.248	0.632	0.440	2.069
Highest Bid - TBA MBS Price	42,779	1.586	0.759	0.710	2.487
Auction Reserve Price[c]	33,571	104.356	1.373	102.505	105.890
Bid - Reserve Price	231,768	−0.222	0.890	−0.875	0.460
Serious Bid - Reserve Price	212,977	−0.115	0.547	−0.703	0.470
Highest Bid - Reserve Price	33,571	0.397	0.572	−0.059	1.049
Reserve Price - TBA MBS Price	33,571	1.179	0.850	0.177	2.149

[a] Money on Table is the highest bid less the second highest bid and is only computed for auctions with at least two bidders.
[b] If a mortgage can go into MBS's with two different coupons, we choose the price of the MBS with the higher coupon.
[c] Data on the reserve price was not available for a subset of the auctions.Serious bids are those that are above the TBA MBS Price - 1/32. The latter is a bit of a buffer.

3.3 Origination Market and Matching

The third data set is on loans originated between 2013 and 2022. It provides detailed information on borrower characteristics (including three-digit zip code and county), the identity of the originators, and fees. To assemble this data set, we combine data from three sources. First, CoreLogic provides data on the name of the retail originators (including brokers and correspondents) for all new mortgage transactions. Second, publicly available data from HMDA contain information on the universe of new mortgage transactions (including rejections), as well as key demographic characteristics of borrowers such has race, gender and income. Third, Optimal Blue's RateLock data set provides rich information on upfront lending fees (or points) that consumers are paying at closing.

To match HDMA loans with OB auction loans, we conduct a strict match on the following loan characteristics: sponsoring agency, loan term, property location (zip code and county), note rate, occupancy status, number of units, dummy variable for whether the loan was a home purchase, and a dummy variable for whether the loan was an adjustable rate mortgage. We then perform a fuzzy match on the continuous characteristics: loan amount, yearly income, DTI ratio and CLTV ratio. We were able to match approximately 85% of auction loans to a unique loan in HMDA. We employ similar procedures to match OB auction loans to loans in eMBS issuances and HDMA originations to loans in eMBS issuances for the period 2013 to 2022. We are able to match x% of HDMA originations with a unique loan in eMBS and match over 91% of auctioned loans with a unique loan in eMBS.

Using data from HMDA, we find that the wholesale market is highly concentrated, with the top-4 banks purchasing more than 45% of all agency loans. The retail market is less concentrated, with the top-4 originators acquiring less than 20% of the mortgages. This difference highlights the importance of scale when selling loans in the secondary market,.and the lack of scale when selling loans in the wholesale market. Large issuers can acquire a diversified pool of loans in the retail and wholesale markets, enabling them to produce higher value securities and to achieve lower servicing costs. By contrast, the barriers to entry in the retail market are much lower. Originators need only to invest in local networks of loan officers and real-estate agents, and they require relatively little liquidity to operate when they can sell loans at competitive prices in the wholesale market. Thus, the emergence of the online auction market has generated potentially important gains from trade: economies of scale upstream and geographic segmentation downstream.

We use the matched HMDA-eMBS data set to identify which originators are using which channels to sell loans, and the volumes that they are selling. We classify originators into three groups: *retail lenders* who securitize and service more than 95% of the loans that they originate, *correspondent-only lenders* who rely exclusively on the wholesale market to finance their retail operation, and *hybrid lenders* who sell loans using both the retail and the wholesale channel.

Table 5 provides summary statistics on the three lender types for Ginnie Mae loans.[8] The first

[8]The results for Fannie Mae and Freddie Mac are similar.

Table 5: Firm size distribution across origination channels

Lender types	Number	C4 %	Market share	Acquisition channel Wholesale	Acquisition channel Direct	Median days to securitize Wholesale	Median days to securitize Direct
Correspondent	429	0.15	0.05	0.97	0.03	24	17
Hybrid	782	0.22	0.53	0.45	0.55	26	18
Retail	120	0.56	0.41	0.02	0.98	25	19

Source: HMDA and eMBS. The sample includes all matched loans between HMDA and eMBS originated between 2018 and 2021. Correspondent lenders are defined as originators who securitize directly less than 5% of their volume. Retail lenders are defined as originators who securitize directly more than 95% of their volume. Hybrid lenders active in both secondary and wholesale markets. The first columns counts the number of lenders in HMDA that are matched with eBMS. The second column measures the market share of the top-4 lenders within each group, and the third column measures the overall market share of each lender types. The acquisition channel measures the share of loans originated by each group that are either sold to the wholesale market, or sold directly to the secondary market. The number of days to securitize is the median number of days between origination date and the issuing date of the security (once per month).

group, correspondent-only lenders, is the largest in terms of numbers, but they only represent 5% of originations. These lenders tend to be small, as indicated by the low within-segment concentration level (the top-4 concentration measure is 15%).[9] The opposite is true for retail-only lenders. The top-4 lenders originate 56% of the loans in that group, and their overall category market share is 41%. This group includes two of the largest originators: Bank of America and Quicken Loans. The middle category of hybrid lenders has the largest market share with 55%, and they tend to be much larger than the correspondent-only lenders.

On average, these lenders securitize directly 55% of the loans they originate, and sell the remainder on the wholesale market. For these lenders, as well as for correspondent-only lenders, the cost of originating loans is directly affected by the expected resale value of loans in the wholesale market. We estimate that, for roughly 65% of agency loans originated in the US, the wholesale price of loans has a direct impact of the price that consumers pay.

The last two columns of the panel highlight another important difference between the retail and wholesale acquisition channels. For loans acquired directly on the retail market, the median time between closing and MBS issuance is 14 days. In contrast, loans acquired on the wholesale market are securitized 26 days after origination. This highlights the main financial cost of selling loans on the wholesale market: lenders must keep loans on their balance sheet longer before receiving a payment.

[9] Since we focus on loans that are matched between HMDA and eMBS, this only represents a subset of all lenders in the US. Our measure of the importance of the correspondent channel is therefore under-stated.

4 A Model of Loan Valuations

MBSs are known as *pass-through securities*. They generate two sources of income for a bank that retains servicing rights: (i) an upfront security price, $P(c)$, from the sale of the loan and (ii) monthly service income (stochastic). The monthly service income is determined by the difference between the note rate associated with the mortgage (r), the coupon c that must be paid to investors, and the guarantee fee (or *g-fee*) paid to the agency (g). These three variables are measured in percentage points (p.p.). The excess corresponds to the gross profit margin on monthly servicing activities. Coupons are chosen from a discrete grid with 0.50 p.p. increments, while the note rates are typically selected from a finer 0.125 p.p. grid increment. When loan i is placed in an MBS with coupon choice of c, the revenue to the bank for a \$100 tranche is given by the upfront payment $P(c)$ plus the discounted value of service income

$$R_i = \underbrace{\sum_{\tau=1}^{T} \delta^{\tau} L_{\tau,i}}_{\text{service multiple } (M_i)} \times \underbrace{\frac{r_i - g - c}{1200}}_{\text{service income}} \tag{1}$$

where δ is the discount rate that the banks uses to weight future cash flows (relative to the upfront cash payment), $L_{\tau,i}$ is the loan balance at the end of month τ, and T is the term of the loan. In practice, the service multiplier, M_i, is a random variable, since it depends on the borrower's decision to default or refinance. Loans that terminate later are higher value loans to banks and investors, because they generate cash-flows for a longer period of time. The service income captures how the servicer collects a fraction of the interest produced by the loan.

Since payments are made monthly and the units of r_i, g, and c are in percentage points, we divide by 1200 to obtain the fraction of the monthly interest payment received.

The discount factor δ may vary across lenders and time because of reserve requirements and the need for liquidity. However, the Ginnie market is composed mostly of shadow banks with similar liquidity needs so, in what follows, we assume that δ is the same across banks. More generally, banks put more weight on liquidity than MBS investors, especially shadow-banks. This generates gains from trade even in the presence of asymmetric information (Downing et al. 2009).

The service income of a loan is common to all lenders. However, the stream of payments and duration of the loan is uncertain, and sellers and buyers in both the wholesale and secondary markets need to form beliefs about M_i when they make their decisions. These beliefs will depend upon the information that they have about the borrower and the loan.

4.1 Security Valuations

After acquiring loans, banks must decide on which loans to place in which security. For any given Ginnie loan and type of security (i.e., custom pool or TBA), banks can select two characteristics:

the delivery month and the coupon rate. Banks typically select the earliest delivery date available (next calendar month accounting for a 1-2 weeks of delivery time), and so we focus on the coupon choice. Subject to restrictions on the servicing income imposed by the agency, banks can choose the coupon that maximizes their expected revenue from the loan. In the finance literature, this problem is described as the "best execution" of an MBS. This choice depends on the security price $P(c)$, which is increasing in c, and the bank's beliefs about the loan duration (M_i). Banks face a tradeoff between increasing their expected revenue from payments by choosing a lower coupon, or increasing upfront revenue from selling the security by increasing the coupon value. The rules for coupon choice are the same for each type of security.

Ginnie Mae charges a fixed g-fee of 0.06 p.p. for all banks and restricts the coupon choice such that the spread $(r-c)$ is between 0.25 and 0.75. This effectively imposes a minimum and maximum markup on banks. Since note rates are typically quoted on a 1/8 p.p. grid, it implies that for most loans, banks do not face a coupon choice. However, for note rates ending with 0.25 or 0.75, lenders can choose a high or low coupon security. For instance, a 4.25 note rate mortgage can be pooled in a 4% coupon security (high) or in a 3.5% coupon security (low). In the latter case, the bank earns a servicing income of $r-c-g = 0.69$ p.p., compared to 0.19 p.p. with the 4% coupon, but receives a lower upfront payment from selling the loan.

Define $\overline{M}_i = E[M_i]$ as the expected duration of loan i. Since Ginnie does not collect other upfront payments from banks, the optimal coupon for eligible loans is a binary discrete choice problem:

$$c_i^* = \begin{cases} c_L & \text{If } \overline{M}_i > \frac{P_H - P_L}{(c_H - c_L)/1200}, \\ c_H & \text{If } \overline{M}_i < \frac{P_H - P_L}{(c_H - c_L)/1200}. \end{cases} \tag{2}$$

where c_L and c_H denote the low and high coupons available for loan i, and $P_L < P_H$ are the associated security prices. Equation 2 implies that only high-duration loans are placed in the low-coupon option, with the threshold determined by the difference in the security prices relative to the difference in coupons.

Thus, our model predicts that banks will select a low coupon if they believe a loan will last for a long time, and a high coupon if they believe the loan will be pre-paid early. In practise, sellers' preferences for liquidity could also affect this decision. Sellers who value upfront cash payment (low δ) are more likely to choose the high-coupon option, which could mitigate the adverse selection problem.

A similar adverse selection problem arises when banks choose to place a loan in a custom security. As we discussed previously, the key difference between a custom pool security and a TBA security is that investors observe the characteristics of loans for the former, but not for the latter. As a result, the price of the TBA security is fixed and common across banks, but the price of the custom pool security will depend upon the expected duration of the loans in the pool. For example, suppose a bank has n loans to securitize and allocates the loans between a custom pool security and a TBA security in order to maximize its revenues. Index the loans by their expected duration from highest to lowest with $i = 1$ being the highest, and assume that M_i is distributed independently across loans. Then the optimal strategy for the bank is to assemble the custom pool by including only the highest duration loans and finding the value of i to maximize

$$\pi(i) = P_S(i; c)i + (n - i)P_T$$

where $P_S(i; c)$ is the price of the custom security when i is the marginal loan and c is the coupon, and $P_T(c)$ is the price of the TBA security with coupon c. Since the value of the security is decreasing in i, the revenue-maximizing rule (ignoring integer problems and assuming an interior solution) is to choose i such that the marginal revenue of the loan is equal to TBA security price. This model predicts that (i) the optimal custom security price $P_S(i^*; c)$ is strictly greater than $P_T(c)$, and (ii) the duration of the loans in the custom security are on average higher than duration of the loans in the TBA security.

Following Chiappori & Salanie (2000), the predictions of our model of security valuation can be tested empirically by measuring the correlation between the choice of a high coupon or multi-issuer pool ("low deductible"), and the probability that a loan gets prepaid early ("accident"). This test can be implemented via regressions by measuring differences in the early prepayment probability across the two coupons and contract types. We conduct the coupon test for TBA securities only, since we do not observe the prices of custom pool securities. The lack of price data for custom pools means that we also cannot test whether the prices of custom securities are higher than the prices of TBA securities.

4.2 Auction Valuations

An important implication of our model of loan valuations is that a bidder's bid on a loan should be increasing in the probability of survival. We can test that prediction by regressing the probability of the event :a loan surviving and bids on loan characteristics and see if the factors that predict higher survival also predict higher bids. Alternatively, we can construct an estimate of the probability of survival based on the survival regression and then regress the bids on that probability.

We also use the auction data to learn about the information structure of the market. Bidders are assumed to be risk neutral. Each bidder j observes a private signal S_{ij} about the value of loan i. Let Z_i denote the vector of loan characteristics that is observed by each bidder, and let ι_i denote the identity of the seller. In a **private value** (PV) model, bidder j's willingness-to-pay for loan i is given by

$$W_{ij} = P_j(c) + \bar{R}_i + S_{ij} \tag{3}$$

where

$$\bar{R}_i = \left(\frac{r_i - c - g}{1200}\right) E[M_i | Z_i, \iota_i].$$

and S_{ij} is an idiosyncratic value shock (e.g., costs or liquidity).

Here $\bar{R}_i$ is common to each bidder since they face the same service fee, and have the same expectation of the service multiplier, M_i. Bidders compete away this value in the auction, so the dispersion in bids reflects dispersion in the value shock and dispersion in security prices.

The latter comes from bidders who can earn higher prices by selling the loan in the custom pool rather than in the multi-issuer pool. Unlike the price of the multi-issuer pool, custom pool prices are bidder-specific, since they depend on the characteristics of the pool. Each bidder knows its own custom pool price, but not those of other bidders.

The bidders' beliefs about M_i are conditional on the identity of the seller because of the possibility of adverse selection. In originating loans, sellers observe loan characteristics (e.g., the borrower's age, race, and choice of points) that are not observed by bidders and are important predictors of loan survival. Since these characteristics are not individually priced in the bids, sellers have an incentive to use their private information to keep the loans that they believe are more likely to survive and sell the others. We use exogenous variation in the capacity of sellers to act on this incentive to determine whether, and by whom, the loans sold in the wholesale market are adversely selected.

In the **common value** (CV) model, the value of the asset is given by

$$P_j(c) + \left(\frac{r_i - c - g}{1200}\right) M_i.$$

Each bidder j gets an informative signal S_j about M_i, and these signals cause bidders to have different beliefs about the survival of loans with characteristics z_i. Bidders may use different methods to process the data or have different incentives to invest in information acquisition, resulting in heterogeneous interpretation of the data. We allow the distribution of the signals to vary across bidders to reflect possible differences in the informativeness of the signal.

Our tests assume that the auction has a pure strategy, monotone increasing equilibrium. Let $B_{ij} = \beta_j(S_j; Z_i)$ denote bidder j's equilibrium bid function for loans with characteristics z_i and let l_{ij} denote bidder j's lock bid for loan i. The threshold signal that solves $\beta_j(s_j; Z_i) = l_{ij}$ is denoted by $s_j^*(Z_i, l_{ij})$. Then, under PV, for any $S_{ij} \geq s^*(Z_i, l_{ij})$

$$E[M_i|S_{ij} = s; Z_i, \iota_i] = E[M_i|B_{ij} = b; Z_i, \iota_i] = E[M_i|Z_i, \iota_i].$$

The first equality follows from monotonicity and the second from the independence of S_{ij} and M_i. By contrast, under CV,

$$E[M_i|S_{ij} = s; Z_i.\iota_i] = E[M_i|B_{ij} = b; Z_i, \iota_i]$$

in strictly increasing in b. Thus, ex-post measures of loan performance should not vary with b if the auction is PV and be strictly increasing in b if the auction is CV. We refer to this test as the **monotonicity** test. Another bid-level test is to condition expected survival on the bidder's bid b and the maximum rival bid. Ex post measures of early prepayment should not vary with the maximum rival bid if the auction is PV, but be strictly increasing in this bid if the auction is CV. In fact, the slope should be higher than the slope for b since the maximum rival bid is a summary statistic of multiple signals and therefore more informative.

A third, related test focuses on winning bids only and is known as the **Winner's Curse** test. Bidder j wins the auction if it submits the highest bid. Let $B_{i,-j}$ denote the vector of bids submitted by j's rivals for loan i. Then, under PV,

$$E[M_i|B_{ij} = b, \max\{B_{i,-j}\} < b; Z_i, \iota_i] = E[M_i|Z_i, \iota_i]$$

and under CV,

$$E[M_i|B_{ij} = b, \max\{B_{i,-j}\} < b; Z_i, \iota_i] < E[M_i|B_{ij} = b; Z_i, \iota_i]$$

If the auction is PV, there is no selection effect: Since ex post measures of loan survival do not vary with b, they also do not vary when b is the winning bid. By contrast, if the auction is CV, there is a selection effect. Winning is "bad news" because it means that rivals have lower signals, which implies lower survival rates.

We run these tests on the sample of Ginnie Mae loans where bidders *cannot* choose the coupon rate - i.e., the note rate does not end in .25 or .75. For these loans, the coupon is the same for each bidder regardless of whether it sells the loan in a multi-issuer pool or a custom pool.

5 Empirical patterns on survival and bids

6 Empirical Analysis of the Wholesale Market

The second prediction of our model of financial intermediation in that banks' willingness-to-pay for loans include a common-value component.

Following Henricks et al. (2003), we test this hypothesis by measuring the correlation between bids in the OB auctions and ex-post loan performance, measured using the 12-month survival probability of FHA and VA loans. We performance this analysis using auctions conducted between January 2018 and December 2019. We further drop auctions for loans for which banks have a coupon choice, which reduces the importance of unobserved heterogeneity in values.

We test the common-value hypothesis by estimating the following linear survival model:

$$Y_i = \lambda \text{Bid}_{ij} + \text{Fixed-effects} + Z_i\beta + \epsilon_i \tag{4}$$

where $Y_i = 100 \times 1(T_i > 12)$, B_{ij} is the bid of bank j, and Z_i includes loan and originator characteristics. We use the same set of control variables included in the adverse-selection test, augmented with information regarding the auction date (instead of MBS issuance month), and originator and borrower county fixed-effects. When controlling for loan characteristics, we also include auction date $\times$ note rate fixed-effects, which absorbs the common believes that bans have about the resale price in the secondary market. However, since about 30% of loans end up in custom-pools, banks' information set include idiosyncratic believes about the value of single-issuer pools.

Table 6 summarizes the results. We measure banks' bids in two ways. First, specifications (1)-(3) control for the *Net Bid* of bank j (Bid - TBA price). Column (1) controls only for the auction date. The estimate of λ shows a strong positive correlation between bids and ex-post performance. A one standard-deviation increase in bids lead to a 2 percentage point increase in survival, or roughly a 11% decrease in pre-payment risk. This correlation confirms that banks incorporation information about loan survival when pricing loans. However, since banks are symmetrically informed about loan characteristics, this does not necessarily lead to a Winner's Curse problem. Column (2) controls for all characteristics observed by banks in the auction, and λ measures the information content in bids beyond observed differences across auctions. The residual correlation between bids and survival drops by roughly 2/3. A standard-deviation increase in bids, is associated with a 2 percent reduction in pre-payment risk (or 0.36% increase in survival).

Specifications (4-5) measures banks' bids using the residual of the following Hedonic regression model:

$$\text{Net bid}_{ij} = Z_i \beta_j + \text{Fixed-effects} + \eta_{ij}. \tag{5}$$

This regression is estimated separately for each bank j. We do so to account for the fact that exist important asymmetries between banks in how borrower characteristics are priced. The residual bids η_{ij} therefore better proxies for the signal that individual bidders' receive about the value of each loan. Column (4) measures the correlation between this signal and the 12-month survival. The estimate of λ reveals again a positive correlation between individual bids and performance. A one standard-deviation increase in the residual value is associated with a 0.43 percentage point increase in survival.

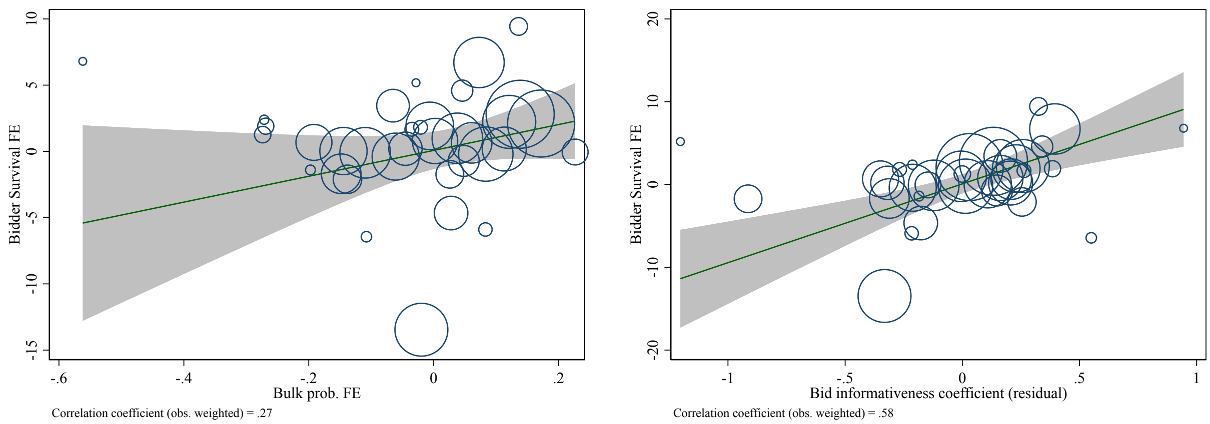
These two sets of results confirm that the ex-post duration of loans is increasing in banks' private signal; consistent with the common-value hypothesis. The second testable implication of the common-value model is banks submitting the highest bid receive an overly optimistic signal of the common component of loan value (proxied by the 12-month survival). This potentially leads to a Winner's Curse problem. We test this hypothesis by including the maximum bid of rival banks within the same auction as an additional control; both using bid levels and residuals. If the bid of rivals is predictive of loan performance (beyond the information content of a bank's own bid), banks would benefit from pooling information. Using net bids (column 4), we find that the second-highest bid is *more* informative about performance than individual bids. A one-standard deviation increase in the maximum rival bid (0.72) is associated with 1.64% increase in performance, or 10% decrease in the risk of early pre-payment.

The last column provides a second measure of the potential importance of the Winner's Curse. We measure the correlation between the winning margin (difference between the first and second highest bid), and the ex-post performance. Under the common-value hypothesis, loans acquired are of low quality when the highest bidder receives an outlier signal of the loan value. This leads to a phenomenon analogous to adverse-selection: banks acquire "lemons" when the gap between the highest and second highest bid is largest. The point estimate show that a 1$ winning margin is associated with 3.4% decrease in survival. In our sample, difference between the 95% and 5% percentile of the distribution of winning margin is 0.75$, an the 99th percentile of the distribution of winning margin is $1.25. To put this number in perspective, the marginal effect of loan size on survival evaluated at the average loan value is about -7%. Winning an auction by a $1 margin is therefore equivalent, in terms of early-prepayment risk, to a 50k increase in loan size (about half standard-deviation).

Although the common-value component is economically meaningful, the correlation between bids and performance is smaller in magnitude than the adverse-selection problem induced by the choice of placing loans in custom MBS's. There are some reasons to believe that our test under-state the importance of Winner's Curse. Our analysis of variance suggests that banks value differently the same observed characteristics when preparing their bids, which can be due to heterogeneity in the quality of signals that banks receive about loan performance. For instance, we observe systematic differences across banks in their propensity to submit bulk bids (as opposed to lock prices). On average 60% of bids are classified as bulk, but 25% of bidders submit bulk bids more than 80% of the time, and 25% of bidders use lock prices more than 45% of the time. Since lock prices are based on a coarse grouping of consumers, it is likely that banks different their ability to predict performance.

To illustrate the importance of information quality on the Winner's Curse, Figure 4a plots the distribution of the difference between the highest bid, and the highest lock price. Recall that banks on the OB platform can sell loans using the best wholesale rate-sheet of banks in their network, or by running an auction. The highest lock price measures the best available "posted" price of the loan after origination. The difference between the highest bid and this price therefore is a measure of the gain from using auctions. For the average loan, this difference is $0.75, which is slightly below the standard deviation of the net-bids across all auctions. The gain from using auctions is therefore substantial.

This difference is also stable over time within originators. Since buyer networks are the same in the lock price and auction markets, this difference is attributable mostly to differences in the information structure. In the auctions, banks are able to submit individual bids for each loan, while



(a) Distribution of gains from using auctions over posted prices

(b) Correlation between information and performance

Figure 4: Effect of information on wholesale prices and bank performance

in the posted price market banks design a relatively small number of wholesale rate sheets. Banks are therefore better information to price loans in the auctions, which alleviate the Winner's Curse problem and lead to higher wholesale prices for originators.

Figure 4b provides further evidence of the impact of signal quality on the wholesale market. The x-axis measures the correlation between bids and loan survival at the bank level, by estimating a separate λ for each bank in equation 5. This is a measure of the "informativess" of banks' bids. The y-axis plots the estimated fixed-effects of the highest-bidder obtained by regressing the 12-month survival on loan characteristics. This variable is centered at zero, and banks with positive values on average acquire loans that perform better than what their observed characteristics suggest. This is a measure of bank productivity in the wholesale market. The figure shows a clear positive correlation between bid informativeness and loan quality. Recall that the point estimate of λ is 0.41.

Banks with signal quality higher than 0.5 on average acquire loans that are 5% more likely to survive. In contrast, banks with signal quality between -.5 and 0.25 acquire loans that are 5% less likely to survive. To the extent that these low-performing banks are maximizing profit, the presence of banks with lower quality signals put downward pressure on wholesale prices, and therefore lead to higher borrowing cost for consumers.

7 Evidence of Adverse Selection in MBS Market

We use the issuer's choice of coupon and type of security to test for adverse selection in the MBS market. We focus on the sample Ginnie Mae securities for FHA and VA 30 year fixed-rate loans issued between 2013 and 2019.

7.1 Coupon Choice

The adverse selection test for the coupon choice is based on the following empirical discrete choice model. The net-present value of a loan with note-rate r_i and MBS coupon c is given by:

$$R_i(c) = (r_i - c - g)M_i + P_{i,c} + \varepsilon_{i,c}, \tag{6}$$

where $\varepsilon_{i,c}$ is a choice-specific preference shock for a security with coupon $c \in \{L, H\}$. It rationalizes the fact that the lender's coupon choice may be driven by factors other than the tradeoff between security price and loan duration. We parameterize the choice model as follows:

$$\Delta\varepsilon_i = \varepsilon_{i,H} - \varepsilon_{i,L} = Z_i\alpha + \eta_i \tag{7}$$

$$\overline{M_i} = Z_i\beta + \nu_{i,1} \tag{8}$$

where $\overline{M}_i$ denotes the lender's expected discounted flow of payments conditional on loan characteristics Z_i and an information shock $\nu_{i,1}$ The preference shocks can depend upon loan characteristics, and we interpret the information shock as representing borrower characteristics that are observed by the lender but not by the econometrician. Loan duration is given by

$$m_i = \overline{M_i} + \nu_{i,2} = Z_i\beta + e_i \tag{9}$$

where $\nu_{i,2}$ is the repayment shock and $e_i = \nu_{i,1} + \nu_{i,2}$. We assume that lenders have correct beliefs about loan duration.

By definition, $c_H - c_L = 0.5$ and define $\Delta P_i = P_{i,H} - P_{i,L}$. Then the choice of a high-coupon security is characterized by the following inequality:

$$\begin{aligned} C_i &= H \text{ iff } (r_i - c_H - g)\overline{M_i} + P_{i,H} + \varepsilon_{i,H} > (r_i - c_L - g)\overline{M_i} + P_{i,L} + \varepsilon_{i,L} \\ &\Rightarrow \underbrace{\eta_i - 0.5\nu_{i,1}}_{w_i} > Z_i(\beta/2 - \alpha) - \Delta P_i \equiv Z_i\gamma - \Delta P_i \end{aligned} \tag{10}$$

Similarly, the event that a loan survives at least 12 months is given by

$$Y_i = 100 \text{iff} M_i > \kappa \Rightarrow e_i > \kappa - Z_i\beta.$$

where κ is the value of M_i associated with a loan that is prepaid at the end of 12 months.

The ability of lenders to select the coupon leads to adverse selection if e_i and w_i are positively correlated. This occurs for instance if selection is entirely driven by the tradeoff between security prices and loan duration (i.e., $\eta_i = 0$). Selection can be *advantageous* if the factors entering the prices for high and low coupon securities are negatively correlated with duration. This can arise for instance if the security prices are set after the MBS pool is assembled, or if a bank's private benefit form assembling a diversified pool outweighs the benefit of placing a longer duration loan in a low coupon.

Our model of coupon choice implies that the probability of loan survival is lower for loans placed in high-coupon securities. We measure this correlation using a linear probability model[10]

$$Y_i = Z_i\beta + \lambda 1\{C_i = H\} + u_i. \tag{11}$$

in which we regress the 12-month survival indicator on loan characteristics and the coupon indicator. Since Y_i is equal to the survival indicator multiplied by 100, λ measures the difference in the survival probability in percentage points between loans placed in low and high coupons. If $\lambda > 0$, then the survival probability of high-coupon loans is lower than that of low-coupons:

$$\Pr(M_i > \kappa | w_i > \gamma Z_\gamma - \Delta P_i) < \Pr(M_i > \kappa | w_i < \gamma Z_\gamma - \Delta P_i).$$

The key assumption underlying our estimation approach is that the repayment shock $\nu_{i,2}$ is independent of the information shock $\nu_{i,1}$ and the preference shock η. This assumption is plausible in our context because the lender observes the borrower's type. We estimate the model on loans that are eligible for a coupon choice (i.e., note rates ending in .75 or .25 digits) and that are sold in multi-issuer pools. In order to identify the source of selection, we estimate λ varying the set of controls Z_i. The price of the TBA MBS securities is a function of only aggregate information available to investors at the trading date (typically a few weeks before the pool issuance date). Our baseline specification therefore compares the ex-post performance of eligible loans that are securitized on the same date. We do so by controlling for note-rate/issuance date fixed-effects (i.e., $r \times t$). Next we also condition on the identity of the issuer (i.e., aggregator or retail originator) selling the loan. We do so by augmenting the fixed-effects to compare loans securitized on the same date, with the same note rate, and the same issuer. Finally, we also control for characteristics of the loans that are observed by the lender/issuer, but not by the investor buying the TBA security. This includes the financial attributes of the borrower, contract characteristics, as well as the origination channel. Note that the investor also does not observe the identity of the issuer.

Panel A in Table 7 summarizes the results. The first row presents the estimate of λ across different sets of controls. The last two columns compare the estimate for loans acquired through the retail and wholesale channels. The results confirm that the coupon choice reflects issuer's beliefs about loan duration. Loans placed in high-coupons are 3.96% less likely to survive the first year, compared to loans sold on the same date and with the same note rate. This difference is reduced to 2.64% when we condition on loan attributes, and is further reduced to 1.61% when we condition on the identity of the issuer. We therefore conclude that more than half of the adverse-selection in the MBS market is due to observed characteristics of the loans. The identity of the issuer in

[10]We use the linear model rather than a probit or logit model because Z_i include a rich set of fixed effects.

particular is an important predictor of the coupon choice and loan survival. We find that a large fraction of issuers never select the low coupon option when it is available, and those lenders on average sell loans that are more likely to be pre-paid early. The last two columns further decompose the effect between retail and wholesale loans. Estimating the survival regression separately in the two samples attenuates the correlation between the coupon choice and survival, but the difference between the two channels is not significant. This suggests that issuers acquiring loans through the retail channel are not better informed about the loan performance than issuers acquiring loans through the wholesale channel.

7.2 Custom Pool vs TBA Security Choice

In Panel B, we estimate a similar survival model to test for adverse selection in the TBA market by comparing the performance of loans placed in multi-issuer and custom pools. In particular, if issuers securitize high-performing loans in custom pools (because the security price is increasing in loan duration), loans placed in multi-issuer pools should have lower survival probability:

$$Y_i = Z_i\beta + \lambda 1\{\text{Multi-issuer}_i\} + u_i. \tag{12}$$

The results strongly confirm this hypothesis. As before our baseline specification includes note-rate x issuance date fixed-effects. Without conditioning on other loan or issuer characteristics, we find that loans placed in multi-issuer pools are 10% more likely to be pre-paid within the first year. Given that the average 12-month survival probability during our sample period was 89%, this difference suggests a very severe adverse-selection problem. Once again, conditioning on loan and issuer characteristics attenuates substantially this difference. In column (3), we find that the difference in survival between multi-issuer and custom securities is 2.87%. A major source of adverse selection in this market is therefore that issuers selling under-performing loans in the secondary market are more likely to use multi-issuer pool securities. Since these lenders tend to be smaller on average, this indicates that larger banks also tend to acquire higher quality loans in the retail or wholesale markets.

7.3 Moral Hazard Test

Finally, the previous two tests of adverse selection are based solely on the correlation between ex-post performance and the security choices. However, the insurance literature has long recognized that a positive correlation between prices and performance can be due either to adverse-selection or moral hazard. In our context, moral hazard can be caused by the ability of lenders to convince borrowers to repay their loans early. For instance, lenders may encourage a borrower to refinance its loan so that they can earn higher service income on the new loan.

We can test this hypothesis using the sample of loans that are not eligible for a coupon choice.

These are loans with service income of 0.375, 0.5 or 0.625 before applying the guarantee fee (i.e., $r - c$). Since we focus only on single-unit mortgages, everything else being equal, loans with higher service income are strictly more profitable for banks. We test the moral hazard hypothesis by estimating the following linear probability model:

$$Y_i = \lambda_1 1(r_i - c_i = .5) + \lambda_2 1(r_i - c_i = 0.625) + \theta r_i + Z_i \beta + u_i. \tag{13}$$

This equation describes the relationship between interest rates and loan attributes and the propensity of consumers to re-pay their loan. Since high-rate loans are more likely to be pre-paid default, we expect $\theta < 0$. The coefficients λ_1 and λ_2 allow for this relationship to be discontinuous around the digits determining the coupon choice. We infer that lenders engage in strategic pre-payment if $\lambda_2 > \lambda_1 > 0$.

Panel C summarizes the results. The top panel measures the effect of service income on performance for retail loans. Without conditioning on the identity of the lender, we find that loans with higher service income are more likely to survive. As with the selection test, this difference is largely due to unobserved heterogeneity across lenders. Once we condition on the identity of sellers, the highest margin loans are more likely to survive (relative to loans with the lowest service income group), but the difference is zero for loans placed in 50 bps coupons. Once we also condition on loan attributes, this difference is eliminated. We therefore cannot reject the hypothesis that banks engage is strategic prepayment based on observed attributes of the loans. However, this effect is economically much smaller than the selection effect, and could be driven by selection of which loans are securitized vs kept on the banks' balanced sheet. Therefore, we conclude that the correlation between security characteristics and loan performance is most likely due to adverse selection based on observed and unobserved loan attributes.

Table 6: Correlation between bids and short-term loan survival

(a) Bid levels

VARIABLES	(1) Bids	(2) Bids	(3) Winning Bid
Net Bid (/SD)	0.30	0.11	1.02
	(0.064)	(0.049)	(0.25)
Max rival bid (/SD)		1.45	
		(0.23)	
Observations	625,205	625,205	60,108
R-squared	0.231	0.232	0.233
Sample	All	All	All
Loan characteristics	yes	yes	yes
County FE	yes	yes	yes
Originator FE	yes	yes	yes
Period FE	t x r	t x r	t x r

(b) Bid residuals

VARIABLES	(1) Bids	(2) Bids	(3) Winning Bid
Residual bid (/SD)	0.38	0.22	0.44
	(0.071)	(0.061)	(0.15)
Max rival residual bid (/SD)		0.54	
		(0.14)	
Observations	625,205	610,722	60,108
R-squared	0.232	0.233	0.233
Sample	All	All	All
Loan characteristics	yes	yes	yes
County FE	yes	yes	yes
Originator FE	yes	yes	yes
Period FE	t x r	t x r	t x r

The sample corresponds to matched loans between OB auctions and eMBS. The sample excludes outlier bids: net bids less than -2, auctions with winning net-bids less than the 0.5% percentile, and auctions for which the runner-up bid difference is greater than the 99.5% percentile. Auctions with outlier LTV and FICO are also excluded: LTV less than 60% or more than 97.75 for FHA and more than 110% for VA loans, and loans with FICO score less than 560. Loan characteristics include: note-rate spread (relative to Freddie-Mac survey), loan amount (linear and square), monthly income, LTV, indicator for maximum LTV, indicator for DTI greater than 60 and between 50 and 60, FICO score, indicator for monthly income above 20,000 or reported zero, indicator for first-time buyer, indicator for purchase loan, indicator for high-balance loan, indicator for FHA loan, ratio of loan size over average price per sq. foot in zip code (from Zillow), log average price per sq. foot in zip code (from Zillow) log average growth rate in house price in zip code (from Zillow), and total bidders invited.

Table 7: Adverse-selection and Moral Hazard in the MBS market for Ginnie-Mae loans

VARIABLES	(1)	(2)	(3)	(4) Retail	(5) Wholesale
Panel A: Coupon choice					
1(High coupon)	-3.96	-2.63	-1.61	-0.93	-0.90
	(0.35)	(0.33)	(0.26)	(0.31)	(0.28)
Loan characteristics	no	yes	yes	yes	yes
Fixed effects	$r \times t$	$r \times t$	$r \times t \times f$	$r \times t \times f$	$r \times t \times f$
Panel B: Pool type					
1(Multi-issuer pool)	-10.0	-4.27	-2.87	-3.07	-2.62
	(0.29)	(0.22)	(0.22)	(0.23)	(0.22)
Loan characteristics	no	yes	yes	yes	yes
Fixed effects	$r \times t$	$r \times t$	$r \times t \times f$	$r \times t \times f$	$r \times t \times f$
Panel C: Moral hazard test					
Spread $(r - c)$: 500 bbs	0.40	0.046	-0.39	-0.31	-0.45
	(0.15)	(0.15)	(0.11)	(0.12)	(0.13)
Spread $(r - c)$: 625 bbs	1.05	0.60	-0.065	-0.11	-0.046
	(0.16)	(0.16)	(0.11)	(0.13)	(0.13)
Loan characteristics	rate+loan	all	all	all	all
Fixed effects	t	t	$t \times f$	$t \times f$	$t \times f$

The sample in Panel B includes all loans sold in Ginnie Mae II securities between 2013 and December 2019. Panel A and C excludes loans sold in single-issuer securities. Panel A further excludes loans not eligible for coupon choice, and Panel C includes only loans that are not eligible for a coupon choice. Loan characteristics include: insurance premium, first-time buyer indicator, loan type (purchase, cash-out refi, etc), days since origination, state fixed-effect, retail channel, correspondent channel, seller fixed-effects. Robust standard errors are clustered at the origination month/note-rate level.

References

Agarwal, S., Chang, Y. & Abdullah, Y. (2012), 'Adverse selection in mortgage securitization', *Journal of Financial Economics* **105**, 640–660.

Allen, J., Clark, R. & Houde, J.-F. (2019), 'Market power and search frictions in negotiated-price markets', *Journal of Political Economy* **127**(4), 1550–1598.

andreas, F., Goodman, L., Lucca, D., Madar, L., Molley, L. & Willen, P. (2013), 'The rising gap between primary and secondary mortgage rates', *FRBNY Economic Policy Review* pp. 17–39.

Buchak, G., Matvos, G., Piskorski, T. & Seru, A. (2020), Beyond the balance sheet model of banking: Implications for bank regulation and monetary policy. NBER working paper 25149.

Cassola, N., Hortaçsu, A. & Kastl, J. (2013), 'The 2007 subprime market crisis in the EURO area through the lens of ECB repo auctions', *Econometrica* **81**(4), pp. 1309–1345.

Chiappori, P.-A. & Salanie, B. (2000), 'Testing for asymmetric information in insurance markets', *Journal of Political Economy* **108**(1).

Downing, C., Jaffee, D. & Wallace, N. (2009), 'Is the market for mortgage-backed securities a market for lemons?', *Review of Financial Studies* **22**(7).

Einav, L., Finkelstein, A. & Levin, J. (2010), 'Beyond testing: Empirical models of insurance markets', *Annual Review of Economics* **2**, pp.311–36.

Gao, P., Schultz, P. & Song, Z. (2017), 'Liquidity in a market for unique assets: Specified pool and to-be-announced trading in the mortgage-backed securities market', *Journal of Finance* **LXXII**(3).

Haile, P. (2002), 'Auctions with resale markets: An application to u.s. forest service timber sales', *American Economic Review* **91**(3).

Hendricks, K., Pinkse, J. & Porter, R. H. (2003), 'Empirical implications of equilibrium bidding in first-price, symmetric, common-value auctions', *Review of Economic Studies* **70**, 115–145.

Henricks, K., Pinkse, J. & Porter, R. H. (2003), 'Empirical implications of equilibrium bidding in first-price, symmetric, common value auctions', *Review of Economic Studies* .

Hortaçsu, A. & Kastl, J. (2012), 'Valuing dealers' informational advantage: A study of Canadian treasury auctions', *Econometrica* **80**(6), pp.2511–2542.

Hortaçsu, A. & McAdams, D. (2010), 'Mechanism choice and strategic bidding in divisible good auctions: An empirical analysis of the turkish treasury auction market', *Journal of Political Economy* **118**(5), pp. 833–865.

Robles-Garcia, C. (2019), Competition and incentives in mortgage markets: The role of brokers. Working paper, Staford GSB.

Stanton, R., Walden, J. & Wallace, N. (2014), 'The industrial organization of the us residential mortgage market', *Annual Review of Financial Economics* **6**, 259–288.

Vickery, J. & Wright, J. (2013), 'Tba trading and liquidity in the agency mbs market', *Federal Reserve Bank of New York Policy Review* .