# Adverse Selection in the Wholesale Mortgage Market*

Jonathan E. Becker

This Version: July 16, 2024

Click here for the most recent version.

[Preliminary: Please do not cite.]

**Abstract**

This paper studies the role of private information in determining the allocation of loans through the wholesale mortgage market. Mortgage originators who sell loans in the wholesale market are often approved to securitize their mortgages with the Government Sponsored Enterprises (GSEs) by selling through the *cash window*. Originators decide for each loan whether to (a) sell to third-party investors in the wholesale market, or (b) sell loans to the GSEs for a lower price while retaining future servicing rights and payments. I develop a simple model to explain the determinants of this choice. Originators make their decision based on a private signal about loan quality and any information conveyed by wholesale investors' price offers. I estimate the model on a novel linked dataset that includes GSE and investor price offers from a large loan trading platform. My results suggest that originators possess private information relevant to a loan's future cash flows and that they leverage this private information when deciding between selling and securitizing. Furthermore, the price offers of wholesale investors are informative for originators, allowing them to refine their beliefs about loan quality and strategically allocate loans. Compared to a world where only coarse-grained loan attributes are priced, the ability to price individual loans increases the number of loans obtained by wholesale investors but these additional loans are adversely selected into the market.

*I'm grateful for detailed conversations with my advisors, Ken Hendricks and Jean-François Houde, as well as Elise Marifian, Dayin Zhang, Panle Barwick, Anthony Defusco, Ashley Swanson, Jason Allen, and Jason Choi. I'm also grateful for feedback from participants in the UW-Madison IO workshop: Chris Sullivan, Lorenzo Magnolfi, Dan Sacks, Karam Kang, and Martin O'Connell. Finally, I've benefited greatly from bi-weekly mortgage data conversations with Alina Arefeva, Dimitris Friesen, and Dena Lemonosov, without whose extensive familiarity with the mortgage market I would doubtless have remained in the dark.

# 1 Introduction

The U.S. residential mortgage market has three primary components: the origination market, the wholesale market, and the securitization market. The wholesale market stands between loan origination and securitization, and is an important driver of decisions in those markets. Despite its size and importance, the wholesale market has received relatively little attention from economists and policy makers. In the origination market, borrowers acquire mortgage loans from banks, credit unions, and mortgage companies. Originators charge fixed up-front fees to underwrite and issue the mortgage, and borrowers agree to pay back the loan with interest over a set period of time. For many small- and mid-sized originators, keeping mortgages on their balance sheets is either unwise or infeasible. These originators have two options–they may either securitize the loan and retain servicing rights, or they can release servicing rights by selling the loan on the wholesale market.[1] In the former case, the originator sells the loan to a Government-Sponsored Enterprise (GSE)—either Fannie Mae or Freddie Mac–but remains in charge of collecting payments, with the right to keep a portion of those monthly payments as their "servicing fee." In the latter case, the originator hands over the loan (including servicing rights and responsibilities) to an "aggregator"–typically a large bank or mortgage company–in exchange for a one-time up-front payment. Both options involve selling the mortgage loan to third-parties, so institutions which rely on these two sources of revenues are said to follow the "originate-to-distribute" model.

In these transactions, loan originators may have access to private information not available to buyers. This private information could include characteristics of the borrower or the loan terms which affect the expected value of the loan (e.g., via the borrower's risk of prepayment or default). For instance, originators know whether borrowers have purchased points or accepted credits, both of which involve tradeoffs between up-front fees and long-term rates–borrowers who accept higher rates over the long term come with a greater risk of prepayment. Despite being informative about loan valuations, not all of this information is made available to loan buyers. This information asymmetry raises the possibility that originators who both use the wholesale market and securitize loans with the GSEs will use

[1]For the duration of this paper, I will use the term *servicing retention* to refer to a broad class of servicing relationships, including not just cases where servicing is performed by the loan originator, but also cases where servicing is contracted out to a subservicer, or transferred simultnaeous with securitization to an approved third-party servicer working with Fannie Mae or Freddie Mac. While the incentives for these various optionss may vary slightly, existing mortgage data sources do not provide details on these servicing contracts, so distinguishing between them in a model is infeasible. For our purposes, the key characteristic of all these servicing options is that the originator has an additional stake in loan via the servicing rights over and above the initial price received for selling the loan.

the wholesale market to offload their less valuable loans.

This paper has three broad goals. First, we aim to establish whether loans selected for sale in the wholesale market are an adversely selected sample. Second, we want to investigate the driving forces behind the adverse selection. While originators may have private information about a loan's likely performance, so too may investors. Because originators are not forced to sell in the wholesale market, some of this information may be conveyed to originators through the wholesale price. Understanding the source and magnitude of private information in the wholesale market is important for addressing any market design proposals. Third, we aim to understand the consequences of private information and adverse selection for the allocation of loans through the wholesale market.

To address these questions, we use wholesale market data from a major auction platform to examine whether private information drives adverse selection of loans into the wholesale market. We develop a simple model of loan valuations, private information, and wholesale price formation, and test the adverse selection theory by looking at differences in ex-post loan performance across sales channels (adjusting for observable loan characteristics). We find some evidence that loans sold at auction are adversely selected, and suggest that both originators' informational advantage in wholesale transactions and the non-committal structure of wholesale transactions are responsible. The combined effect is that given the structure of the wholesale market and the informational asymmetries between originators and investors, wholesale transactions allow for originators to retain the servicing rights for idiosyncratically high-value loans. At the same time, however, the ability to price loans at the loan-level allows investors to secure servicing rights for a greater volume of loans than they would otherwise, which most likely represents an overall efficiency gain.[2]

This paper contributes to the literature on the wholesale mortgage market. The wholesale market has received relatively scant attention in the literature, despite being both large and important. Most work has focused primarily on wholesale funding relationships, not specifically loan trades. Stanton et al. (2014) examine wholesale market relationships, and argue that connections through funding arrangements mean that the mortgage market is effectively more highly concentrated than a superficial examination of mortgage origination would suggest. Zhang (2020) examines the importance of wholesale funding, by looking at small banks who rely on wholesale funding advances from the Federal Home Loan Bank system for their origination business. Finally, Benson et al. (2023) examine the withdrawal of

[2] Aiello (2022) examines financially constrained mortgage servicers, arguing that they pursued aggressive foreclosure options more often than less constrained servicers, leading to a decrease in market efficiency. Also, see Kim et al. (2022a), Diop and Zheng (2023), and Sandler (2023).

banks from the loan aggregation market and the corresponding move into wholesale funding alongside the arrival of non-bank servicers. Similar to these papers, this paper examines an important source of wholesale funding; however, we focus narrowly on the buying and selling of loans in the wholesale market. These loan trades are closely tied to the credit lines and wholesale advances discussed in the literature, since they provide the cash originators need to meet their payment obligations.

This paper also contributes to a growing literature on the securitization decisions of mortgage originators. Kim et al. (2018) and Kim et al. (2022b) discusses differences in the securitization incentives faced by banks and non-banks, and find that as a result non-banks tend to issue riskier loans. Buchak et al. (2023) examine the role of integrated intermediaries and argue that originators with the capacity to service their securitized loans have additional advantages in retaining customers. Similarly, Huh and Kim (2022) analyze the effects of securitization opportunities on borrowers, and argue that the structure of securitization options has large effects on the rates faced by borrowers.

Finally, this paper contributes to the literature on private information and adverse selection in the mortgage market. Recent work has tended to focus on role of private information along two fronts: (1) large institutions choosing whether and how to package loans into mortgage backed securities (MBSs) on the basis of private information, and (2) MBS investors deciding which securities to deliver in (non- pre-specified) To-Be-Announced (TBA) trades. The majority MBS trades occur in the To-Be-Announced (TBA) forward market, where trades do not specify the identity of securities to be delivered at settlement. Relative to the market for specified pools, the TBA market is extremely liquid; however, adverse selection is likely to occur since sellers have flexibility in choosing securities to deliver. Downing et al. (2009) document adverse selection MBS market for special purpose vehicles market amounting to 4-6 basis points in terms of yield-to-maturity. However, Vickery and Wright (2013) argue that adverse selection is unlikely to undermine the TBA market, as the magnitude of adverse selection is swamped by the liquidity value of TBA trades. Private information is valuable in the securitization process because not all securities are tradeable on the TBA market; thus, lenders also face a choice of which loans are placed into TBA-eligible securities. Becker et al. (2023) examine this decision in the context of Ginnie Mae loans, and provide evidence of that loans are adversely selected into TBA-eligible securities. Huh and Kim (2023) also examine this decision and argue that high-quality loans are pooled separately and traded outside the TBA market, suggesting a strong role for adverse selection. Mayock and Shi (2022) examine the related question of adverse selection in mortgage servicing rights, and argue that servicing rights are transferred on loans that are more risky *both* ex-ante and ex-post.

Most closely related to this paper is a small literature on selection of balance sheet loans. Agarwal et al. (2012) look at loans originated before the financial crisis and argue that during this period, banks retained low-default-risk loans on their portfolio while securitizing higher-default-risk loans. More recently, Shi et al. (2023) examine banks' securitization decisions and argue that banks keep high-quality loans on their balance sheet, though unlike this paper, the driving mechanism is the agencies' guarantee fees and not necessarily private information about loan quality.

In the remainder of this section, I outline the structure of the paper. Section 2 introduces the mortgage market, covering institutions and incentives at each point in the life of a mortgage. Section 3 highlights existing sources of mortgage data and discusses my method for combining these sources into a single dataset that tracks loans from application and origination, through securitization, to repayment/delinquency. Section 4 presents descriptive evidence about differential performance between sold and securitized loans to motivate detailed investigation of adverse selection. Section 5 outlines a simple joint model of loan performance, originator beliefs, and bid formation. It also discusses measurement and modeling choices required to estimate this model. Section 6 provides model estimates and discusses the incentives created by the information structure of the wholesale auction environment. Section 7 presents two simple counterfactuals—one where the auction market is shut down and replaced by the existing posted-price market, and one where sellers must choose to sell all loans to investors or securitize all loans with the GSEs. We highlight the allocative consequences of allowing fine-grained pricing in the presence of privately informed originators who can choose between channels. Section 8 discusses broader implications of our findings for origination and securitization markets.

## 2 Market Background

This project focuses on the wholesale mortgage market. However, the origination and securitization markets are important for contextualizing the wholesale market. The origination market supplies loans to the wholesale market, while the securitization market drives demand for loans in the wholesale market. This section provides additional detail on all three markets.

## 2.1 The Origination Market

In the origination market, borrowers search for a mortgage by visiting lenders, either by searching for a loan directly or employing a broker to search on their behalf. Historically, these lenders have predominantly been branches of depository institutions located in the borrower's community–small community banks or local branches of large national associations. In recent years, an increasing volume of loan origination has been performed by non-depository mortgage companies who have little to no local branch presence and must reach borrowers online. Depositories fund loan origination through deposits, whereas mortgage companies typically fund origination by drawing down a wholesale line of credit which is repaid when the loan is sold or securitized.

## 2.2 The Wholesale Market

### 2.2.1 The Posted Price Market

The bulk of wholesale loan trades occur in a *posted price* market, where prices are set according to investors' *rate sheets*. Rate sheets delineate how much an institution is willing to pay for a loan with certain characteristics. The main source of variation in rate sheet pricing comes in the form of a "reference rate," which is adjusted daily to reflect current market conditions and is set separately by interest rate. Added to this are fixed price adjustments for loans with certain characteristics.[3] Most rate sheets price using adjustments for coarse bins on credit score, loan-to-value ratio, and loan amount; while loan amount is priced idiosyncratically, almost all institutions set uniform prices for loans within 20-point FICO bins and 5% LTV bins,[4] and a small adjustment may be added for loans originated under special loan programs such as Fannie Mae's Green Financing program or Freddie Mac's Home Possible program.[5] These rate-sheets form the backbone of the posted price market, allowing originators to shop among the standing rate-sheet offers when selling loans.

Origination is not costless for lenders, and many (especially smaller) institutions lack the ability to self-fund their origination business.[6] By providing a straightforward means of

[3]For these adjustments, loan characteristics are usually binned, and the placement and pricing of these bins is relatively stable through time

[4]The placement of these bins mirrors the loan-level price adjustments set by Fannie Mae and Freddie Mac as discussed in Section B.2.

[5]A list of specialty loan programs for Freddie Mac can be found at:
- https://sf.freddiemac.com/working-with-us/origination-underwriting/mortgage-products
and for Fannie Mae at:
- https://multifamily.fanniemae.com/financing-options/specialty-financing

[6]Many can't fund their business entirely through deposits, either due to structural factors (e.g., regula-

selling loans, the posted-price market ensures that small lenders can continue to originate mortgages. Originators can fund their origination business by either (a) drawing down a line of credit which they pay back using proceeds from loan sales, or (b) using pre-arranged transactions where the posted-price offers set in advance a price to be received upon closing.[7] The posted price market can be used for both funding models.

### 2.2.2 The Auction Market

In recent years, an increasing number of wholesale trades have occurred via auctions conducted through online trading platforms. This project looks at auctions on the largest such platform—Optimal Blue.[8] Using this platform, originators can set up an auction shortly after a loan has closed. Once an auction has been set up, originators can invite bidders from a (typically short) list of approved counterparties–the *seller's network*. Invited bidders are notified and given an opportunity to submit bids. If they choose not to submit bulk bids, their posted price offer (or *lock price*) remains in place, and originators may sell at that price. Most auctions occur within a short time window–typically 1-2 hours.[9] These auctions are effectively single-loan auctions.

After the auction concludes, loans are not automatically transferred to the highest bidder. Instead, all bids are summarized for the originator, and originators are presented with multiple options for distributing loans. Sellers can incorporate their own valuations for their loans and account for any idiosyncratic costs of dealing with specific investors. The platform will then optimize on behalf of originators, advising which loans to sell to investors and which to keep. As such, Optimal Blue auctions are highly *non-committal*–originators are not required to sell to any particular investor, and they may choose not to sell at all.

tion barring certain institutions from collecting deposits) or volume considerations (e.g., small banks have insufficient deposits to sustain large mortgage volumes).

[7]In these pre-arranged transactions, small originators may enter into financing agreements with larger institutions, whereby the small institution handles the origination process and agrees to sell the loan to the large institution at a pre-determined price immediately upon closing. Because the origination process is lengthy and involved, originators absorb interest rate risk by "locking-in" borrowers' interest rate offers for a set period of time. As long as the prospective borrower closes on their home before this period elapses, they can exercise the option to borrow at the agreed rate regardless of changes to market conditions in the interim (though rate-lock contracts may include an option to extend the duration of their locked rate in the event of unforeseen complications). Typically a price discount is applied for loans with longer lock periods (typically for rates locked 0, 15, 30, 45, and 60 days out) to mitigate against interest rate risk.

[8]While platforms differ in their features, Optimal Blue's auctions are likely common to others.

[9]Originators *can* choose a longer window if desired. However, this is only common for certain types of loans–for instance, sellers dealing in "scratch-and-dent" loans often allow for up to ten days of bidding.

## 2.3 The Securitization Market

In the securities market, mortgages are bundled into pools backed by borrowers' payment obligations. Most mortgages in the U.S. are ultimately securitized,[10] with the bulk of securitization done by Fannie Mae, Freddie Mac, and Ginnie Mae in the *Agency MBS market*.[11] Securitization provides benefits to both investors and originators. For investors, mortgage backed securities offer stable returns by pooling across loans with differing characteristics and risk profiles (e.g., loans originated across geographic markets). For originators securitizing a mortgage with the GSEs, they retain a stake in the loan by acting as the loan's servicer, collecting payments from the borrower and delivering those payments to the GSEs for disbursement to investors. For loans that trade hands prior to securitization, servicing rights and responsibilities are either *retained* by the originator/seller or *released* to the buyer.[12]

This project focuses on loans eligible for securitization with Fannie Mae or Freddie Mac, who specialize in securitizing low-risk loans. Fannie and Freddie only acquire loans that satisfy the conforming loan limit–a maximum loan amount adjusted by year and county to reflect local housing market conditions. Furthermore, borrowers must meet minimum standards for a loan to be eligible for securitization. Borrowers need a minimum credit score of 640 and a maximum back-end debt-to-income ratio of 50%. Home purchase loans must meet an 80% loan-to-value ratio (i.e., a 20% down-payment) requirement or pay private mortgage insurance until the effective loan-to-value ratio based on outstanding principal falls to 80%.[13] In addition to imposing loan-specific standards, the regulator of the GSEs, the Federal Housing Finance Agency (FHFA), maintains strict requirements for loan originators to sell and service loans with the Enterprises.[14]

When loans are delivered to the GSEs, originators are compensated through one of

[10] Only 27% of mortgage debt is unsecuritized and remains on the balance sheets of financial institutions, with 86% of this unsecuritized debt being held by depositories.

[11] In theory, mortgage backed securities can be created by private institutions, however this *private label* market largely collapsed after the 2008 financial crisis; post-crisis, the private label MBS market accounts for 3.2% of outstanding residential mortgage debt, compared to 66.4% in the Agency RMBS channel (Goodman et al. (2023)).

[12] While servicing rights can trade separately from loans, the wholesale market deals almost-exclusively in released-servicing loans where loans and servicing rights trade together.

[13] Borrowers not meeting these requirements may be eligible for a mortgage backed by an explicit government guarantee through the Federal Housing Administration (FHA), the Veteran's Administration (VA), of the US Department of Agriculture (USDA). Loans with such a guarantee are almost exclusively securitized by Ginnie Mae and will not be considered in this project. It should be noted that for sellers to be eligible to securitize mortgages with Ginnie Mae, they must also remain in good standing with the Enterprises. See Ginnie Mae 5500.3, Rev. 1.

[14] Available at: https://www.fhfa.gov/Media/PublicAffairs/Documents/Fact-Sheet-Enterprise-Seller-Servicer-Min-Financial-Eligibility-Requirements.pdf

three mechanisms: Lender Swaps, Portfolio Securitization, or Structured Securitization.[15] In lender swaps and structured securitization, originators are provided compensation in the form of a mortgage-backed security (typically one backed by the loans being exchanged) which they may sell to investors in the securities market. With portfolio securitization (or *cash-window* transactions), originators (typically small and medium-sized originators) give mortgage loans to the GSEs in exchange for cash.[16] In cash window transactions, the GSEs take on the burden of securitization and assemble pools comprised of loans acquired from multiple originators. Additionally, the GSEs assume risks involved in selling those securities on the MBS market–risks that are typically assumed by the originator in traditional loan swap transactions. Cash window prices thus come with a slight discount, meaning the value to the originator of securitizing loans through the cash window tends to be lower than the value of the securities backed by those loans. As such, the cash window is typically used out of necessity–predominately by small originators who lack the origination volume sufficient to create securities backed exclusively by their own loans.

In our setting, the Optimal Blue platform interfaces with Fannie Mae and Freddie Mac. For the GSE prices, the auction system queries the GSEs and returns a cash window price for each loan. While originators can interact with the GSEs directly, Optimal Blue streamlines this process, so much so that some sellers use the platform exclusively to sell to the GSEs.

# 3 Data

## 3.1 Sources of Mortgage Data

The data for this project come from three primary sources. First, mortgage origination data is provided through the Consumer Finance Protection Bureau under the Home Mortgage Disclosure Act (HMDA). HMDA includes the near-universe of U.S. residential mortgage originations, including *all* residential mortgages from firms which originate greater than 100 loans per year. In addition to originations, HMDA also provides data on wholesale market purchases from qualifying financial institutions. For each loan in the HMDA data, we observe loan characteristics, borrower characteristics, and the identity of the originator.

The second source of data is auction data from a large loan trading platform–Optimal

[15] See, for example: https://capitalmarkets.fanniemae.com/media/4271/display

[16] Historically, slight differences between Fannie Mae and Freddie Mac led to Fannie Mae having a cash-window price advantage over Freddie Mac. However, with the creation of Uniform Mortgage Backed Securities (UMBS) in mid-2019, which allowed for Fannie and Freddie securities to trade together, this price advantage largely dried up and Fannie and Freddie are viewed as substitutable.

Blue. The auction data covers almost two million auctions held between January 2018 and July 2022, and it includes loans with diverse borrower and institution characteristics from roughly 200 originators. Furthermore, the auction data has bids from nearly 80 mortgage investors, as well as cash window price offers from Fannie Mae and Freddie Mac where applicable.

The final major source of data is loan-level securization and performance data from Fannie Mae and Freddie Mac. These agencies represent the vast majority of mortgage securitization for conventional loans–loans which are not backed by any government agency–after the collapse of the private-label mortgage securities market in the 2008 financial crisis. Thus any loan that is ultimately securitized is most likely observable in the securities data. For each loan securitized with one of these agencies, we not only have borrower characteristics at the time of origination, but also monthly payments and delinquencies.

## 3.2 Matched Dataset

These three sources of data can be matched with high fidelity, owing to several intermediate matches. Thus for the bulk of loans which travel through the auction platform we know: (1) extensive borrower characteristics at the time of origination, (2) bids received by the loan at auction, (3) whether the loan was securitized with one of the major agencies, (4) ex-post performance of the loan *if* securitized, and (5) whether the loan traded hands, either between origination and auction or between auction and securitization. Furthermore, we know for each originator on the auction platform: (1) all mortgage loans they originate, (2) which of these loans they choose to sell via auctions or other wholesale channels versus those they securitize themselves, and (3) the ex-post performance of the securitized loans from each of these channels. In addition to filling out our picture of loan and borrower characteristics, linking loans across multiple datasets also allows us to identify originators with systematic data errors in a given data source. A detailed description of the match between datasets is provided in Appendix F.

## 3.3 Estimation Sample

### 3.3.1 Sample Restrictions

For all exercises in this paper, we restrict attention to 30 year fixed-rate mortgages–the most common mortgage product in the U.S. by a wide margin. Furthermore, we focus on conventional loans (not guaranteed by any government agency) on one-unit single-family homes. We exclude loans with atypical features such as balloon payments and interest-only periods, as well as homes purchased as investment properties.

We also restrict attention to mortgage originated between January 2018 and February 2020. This choice of sample ensures that at least three years of payment history can be observed for every loan in our sample. These loans all experience an unforeseen interest rate shock resulting from the COVID-19 pandemic. Restricting attention to loans originated before March 2020 mitigates against the possibility that loan origination decisions were made under vastly different financial constraints.

Furthermore, my primary analysis dataset focuses on a select subset of mid-sized loan originators who are observed to sell loans directly to the GSEs. Approximately 50% of all originators on the auction platform are approved to sell to *both* Fannie Mae and Freddie Mac. Further, 30% of originators are approved with *either* Fannie Mae or Freddie Mac.[17] Only 20% of originators are approved with neither, but these originators tend to be extremely small. Most originators approved with the GSEs are approved from their first day on the platform, suggesting that their use of the auction platform is a continuation of their business model prior to platform entry.[18]

For transparency, we flag two potential sample restrictions that are not made here, but which might be important when interpreting the model results. Those are restrictions to (a) mortgage companies instead of banks, and (b) institutions who service loans in-house. First, restricting attention to mortgage companies rather than banks would allow us to focus more narrowly on the choice between wholesale and retail channels. Independent mortgage companies are non-depository institutions and have to sell their loans in order to free up cash for future originations. By contrast, banks can fund their origination business through deposits, and thus have an additional option to retain loans on their balance sheets. A more comprehensive model of banks loan-level decisions would consider all three options, not just the two that are considered in this paper. However, the banks who originate loans in my dataset are small and thus less likely to utilize the balance sheet option. Thus, I treat them as effectively facing the same binary choice that mortgage companies do.

Finally, we could distinguish between institutions who service loans in-house from those who sell-off or contract-out servicing to a third-party. Due to structural features in the market, the majority of originators who sell to the GSEs specialize in one servicing method–

[17] Owing to Fannie Mae's historical pricing advantage, 25% are approved with Fannie only, while 5% are approved with Freddie only.

[18] A small handful of originators begin to interact with after entering the auction platform. However, most of these originators only begin interacting with the GSEs *after* the Covid-19 pandemic begins, and thus fall outside our sample period.

either retaining servicing rights for all loans or selling/contracting servicing rights for all loans.[19] Some originators are *integrated* lenders who have in-house servicing capabilities, while other originators are *specialized* lenders who specialize in loan origination, but sell servicing rights to one of a handful of specialized servicers. The relationship between loan performance and loan value (to the originator) is clear in the case of institutions who service loans in-house–they continue to derive value from all loans that are not prepaid or in default. For institutions who sell or contract-out servicing, the relationship depends on details of how the servicing rights are transferred, which are unobserved and thus hard to model. Restricting attention to lenders with in-house servicing would drop about half of all loans and lenders on the platform; however, many of these dropped lenders are extremely small.[20] To avoid cutting the sample size drastically, I keep loans where the lender contracts out servicing but still sells loans to the GSEs. So long as lenders and servicers write contracts where lenders have skin in the game, their incentives will be similar to lenders who service in-house.

### 3.3.2 Loan and Originator Characteristics

Table 1 reports summary statistics for the estimation sample. The average loan size is \$256,500 and the median is \$240,000. This reflects the fact that the distribution of loan size has a long right tail, owing to loans in high-price markets like New York, Washington D.C., and San Francisco. This average loan size means that the representative loan yields about \$625 in servicing revenues per year in first few years. Notice also that the borrowers are relatively "safe." The average credit score among borrowers is nearly 750, with the vast majority of the sample above 700. The median loan-to-value ratio is 80%, and although 25% of the sample has a loan-to-value of $\geq 90$, these borrowers tend to be competitive along other dimensions and simply pay private mortgage insurance during the first years of the loan. Finally, borrowers' debt-to-income ratios remain well-below the soft "limit" of 50 set by Fannie Mae and Freddie Mac.

The average auction has 13 participating investors, submitting either coarse grained posted prices, fine-grained bulk bids, or both. Note that in the event a single investor submits both a posted price offer and a bulk bid, we observe only the highest of the two–their effective price. Most auctions receive at least one "bid" of each type, however, with the maximum (effective) bulk bid being about \$0.60 higher than the maximum posted price.

[19] This is shown in Figure 1b

[20] This owes to fact that much of servicing costs comes in the form of personnel. Surveys from the Mortgage Bankers Association suggest that the typical employee in servicing manages upwards of 500 loans. Originators with fewer loans than this will be at a considerable cost disadvantage.

Relatedly, the highest bulk bid is considerably higher than the price offers of both Fannie Mae and Freddie Mac. This is to be expected, as most investors at the auction also have the ability to sell via the cash window. Note that while the average price offer from Fannie Mae is higher than the average price offer from Freddie Mac, this is largely due to Fannie Mae prices being queried more frequently for more valuable loans. Looking at loans where both Fannie Mae and Freddie Mac price offers are observed, Freddie Mac prices tend to be higher, especially before June 2019.

In terms of loan performance, the loans in our sample have relatively poor performance compared to previous years. One year after origination, only 81% of loans 'survive'–the remaining 19% have prepaid. At two years, this figure is 49% and at three years it is only 33%. These numbers are uncharacteristically low, but not surprising in our setting. The loans in our sample experienced a historically low (and arguably unforeseen) interest rate environment due to the Covid-19 pandemic. The loans in our sample mostly had note rates ranging from 4% to 5%. By late 2020, these same borrowers would be facing rates as low as 2%, meaning that refinancing to a lower interest rate was extremely attractive, especially for larger loans. In terms of servicing revenues, this translates to about $0.23 per $100 loan volume in the first year, but only $0.46 per $100 in the first three years (compared to about $0.70 for loans that were not prepaid during this time).

A total of 35 loan originators meet the inclusion criteria for our estimation sample. Though this amounts to under half of the originators using the platform during this time, it includes a majority of the largest originators, for whom the choice between sale and securitization is most salient. Over the sample period, the average originator holds auctions for over 3,700 loans. Originators differ in the number of investors they invite to submit bids, with the average originator inviting 16 unique investors.

# 4 Descriptive Statistics

This section provides descriptive evidence of adverse selection in the wholesale market to motivate the modeling choices in Section 5. We first look at originators' propensity to sell given the spread between investor bids and GSE prices, showing that the probability of selling wholesale increases in this spread. Next, we conduct a set of bivariate probit regressions using the selection decision (i.e., the decision to sell to investors or GSEs) and various binary measures of loan performance and find that loans sold to investors tend to be more likely to default or to prepay early. Finally, we use non-parametric survival models to document differential survival between the loans sold wholesale and those sold to the GSEs, showing that loans sold to the GSEs tend to survive longer than those sold wholesale. Together, these three exercises suggest that loans sold to investors are adversely selected.

## 4.1 Investor Bids and GSE Price Offers

Figure 3 shows the distribution of the spread between the maximum investor bid and the maximum GSE price. The majority of loans have a positive spread, with an average magnitude of \$0.62 per \$100 of loan volume. This is more stark when we break down the distribution by whether a loan was sold or securitized. Loans that are sold to investors have an average spread of \$0.79 per \$100 of loan volume, and only 2% of loans sold to investors have a negative spread. Of the loans securitized with a GSE, 43% have a negative spread, and the average spread is only \$0.10 per \$100 of loan volume. Nonetheless, the fact that the majority of loans securitized with a GSE have a positive spread suggests that originators perceive some value to servicing loans even after servicing costs are accounted for.[21]

Table 2 shows that the probability of sellers selling to investors is increasing in the bid spread. At the margin, increasing the spread between investor and GSE bids by one dollar increases the probability of sale to investors by about 20% (a large increase when we consider that around 60% of loans sell to investors). Furthermore, we notice that the sale decision depends strongly on the bid spread. Loan characteristics such as loan size and credit score do partially explain the sale probability, but they are relatively unimportant in comparison to the bid spread.

## 4.2 Positive Correlation Test for Private Information

Initial descriptive evidence for adverse selection comes from the positive association test of Chiappori and Salanie (2000). I model selection using a bivariate probit model:

$$Y_i = \mathbf{1}_{\{Y_i^* \geq 0\}}, \quad Y_i^* = Z_i\beta_i + \varepsilon_i, \quad \begin{bmatrix}\varepsilon_1 \\ \varepsilon_2\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0 \\ 0\end{bmatrix}, \begin{bmatrix}1 & \rho \\ \rho & 1\end{bmatrix}\right) \tag{1}$$

[21] As discussed further in Appendix B.2, there is puzzling bimodality of the bid spread distribution when raw bid numbers are used. This difference is driven by a change in certain sellers' GSE price offers that occurs abruptly on January 1, 2019. This is most likely an artifact of the way prices are reported in the data, and can be controlled for straightforwardly in the model by separating sellers into two types–those who see a downward shift in reported GSE prices in early-2019 and those who do not.

where $Y_1$ is a binary performance outcome, $Y_2$ is a binary selection outcome, and $Z_i$ is a vector of covariates important for *both* survival and selection.

For the exercises in this section, we use two types of binary loan performance measures: (a) loan survival indicators in 6-month increments, and (b) an indicator for loan default. Loan valuations depend on the entire stream of servicing cash flows, for which binary survival is an imperfect proxy. Nonetheless, binary survival captures an important component of loan valuations–loans that survive longer pay out servicing revenues for longer. Additionally, because mortgage interest payments (and hence servicing payments) are largest at the beginning of the loan term, prepayment has the largest impact on cash flows early on. Although partial prepayment is possible (and would not be captured with binary survival measures), it is uncommon in practice; the largest source of prepayment risk in both frequency and magnitude is *full* prepayment which *is* captured by loan survival. We consider loan default separately because, though it is relatively rare, the cost of default is high. Servicers must front payments to investors, and they typically undertake much greater effort to follow-up with borrowers, enter foreclosure proceedings, etc.

For the margin of selection, we look at selection into securitization. Because auctions are non-committal, originators are not obligated to sell loans to the highest bidder and may instead opt to securitize a loan directly. If the originator has private information relevant to survival, then they will require higher bids in order to sell "good" loans to wholesalers. Thus, under adverse selection, we should see a positive correlation between selection into the retail channel and binary survival and a negative correlation between selection into the retail channel and loan default. We focus on originators who take the bulk of their loans to auction and who use a mix of third-party sales and securitizations.

Results for the estimated correlation parameters are shown in Table 3. Estimated correlations are consistent with adverse selection into auctions by sellers along two related dimensions. A reasonably strong negative correlation exists between loan default and selection into securitization. Loans retained by originators carry less default risk, accounting for observed characteristics. A positive correlation exists between loan survival and securitization, so loans retained by originators pay out servicing revenues for longer. A smaller and non-significant correlation exists between 12- and 18-month survival and securitization. However, this is unsurprising even if originators have payoff-relevant private information; many loan trades come with penalties for the seller if the borrower prepays or defaults within the first year. Both results indicate that originators choose to retain servicing rights to loans with greater cash flow potential.

### 4.3 Differential Survival

As shown in Figure 4, the relationship between loan performance across channels that we noted with binary survival also holds with the more continuous measure. The figure presents predicted loan survival trajectories using a random survival forest model discussed in more detail in Appendix C. Figure 4a shows the model predictions when only using covariates priced at auction as predictors. No difference between channels would be predicted on the basis of covariates alone, suggesting that loans sold to investors are comparable to those sold to the GSEs in terms of characteristics. Figure 4b shows the model predictions, now with the realized sale decision as an additional predictor. Now we see a significant gap between the predicted survival curves forloans sold to investors and those sold to the GSEs. Loans sold to investors have less favorable survival trajectories than those sold to GSEs. This is comparable to the approach taken by Dávila and Parlatore (2018), who measure price informativeness by comparing model fit for a model trained with and without an informative variable, though in our context the informative variable is the sale decision rather than a price. This is consistent with the story told in Section 4.2.

Using a full survival model rather than binary survival indicators better-captures the cash value of the difference between securitized and sold loans. In early years, before a loan's principal has been paid down, servicing revenues are just proportional to duration of survival. By 30 months, the average retail loan has paid out the equivalent of an additional month's servicing revenues, relative to the average wholesale loan. Since the originators in our sample originate hundreds or thousands of loans per year, these per-loan servicing differences can add up to substantial revenues.

Note that while these differences are documented only for the estimation sample, there is reason to believe that the estimation sample is representative of the wholesale market more broadly. This is discussed briefly in Appendix B.3.

## 5 Selection Model

In this section I outline a simple model of loan valuations to illustrate how incentives for adverse selection arise in the wholesale market. The model is close in spirit to Einav et al. (2012) and Einav et al. (2013).

## 5.1 Model Timing and Incentives

The model begins with originator $j$ having originated loan $i$. Initially, the originator receives a private signal $S_i$ about the quality of the loan. The originator also observes a cash window price $P_i^0$ for the loan. The originator also receives a wholesale price offer $P_{ij}$. The originator then decides between selling the loan to the GSEs and selling to a wholesale investor. If the originator chooses to sell to the GSEs, they receive the cash window price $P_i^0$ up-front and retain servicing rights and responsibilities, entitling them to a stream of servicing payments for the life of the loan. If the originator chooses to sell wholesale, they receive the investor price $P_{ij}$ up-front and release servicing rights and responsibilities. Finally, loan performance $M_i$ is realized, the loan servicer receives payments and incurs any costs of servicing and securitization. If the loan is sold to a third-party investor, the costs of servicing are paid by that investor. If the loan is securitized directly by the seller, then the seller incurs a fixed cost $c$, which represents the full cost of securitizing and servicing the loan.

### 5.1.1 The Keep/Sell Decision

The originator's decision concerning whether to sell or securitize the loan depends on their relative valuations of the two options. The value of selling the loan to the GSEs depends on the up-front GSE price, the cost of servicing, and the expected future cash flows from servicing. Originators who securitize and service their loans have realized servicing cash flows at loan age $T$ given by:

$$\text{Servicing Revenues}_T = \frac{.25}{12} \times \sum_{t=1}^{T} UPB_t \tag{2}$$

reflecting the fact that servicers receive an annualized servicing fee of 25 basis points on the loan's unpaid principle balance (UPB).[22] Thus, we construct our measure of cash flows $M_i$ as:

$$M_i := \frac{0.25}{1200 \times UPB_0} \sum_{t=1}^{T} \beta^T UPB_t. \tag{3}$$

That is, a loan's valuation is just a discounted sum of servicing fees. The maximum loan age $T$ is the amortization period of the loan in months ($T = 360$ in our estimation sample).

[22]Some servicers can negotiate a slightly different servicing rate, but total servicing revenues will remain proportional to the sum of unpaid balances.

Dividing through by 100 ensures that the cash flows are expressed in dollars of servicing revenues per \$100 of loan volume (i.e., the same scale as bids). For present purposes, I compute loan performance summary statistics using a constant discount factor $\beta = 0.98$. With this measure of cash flows, the value of selling to the GSEs and keeping servicing rights can be written as:

$$v_{ij}^{K} = P_i^0 + \mathbb{E}\left[M_i | S_{ij}, P_{ij}\right] - c.$$

Here we allow the expectation of future cash flows to depend not only on the originator's private signal, but also the price offer they receive from wholesalers. Because our auction environment is non-committal on the seller side, it is important to allow originators to learn from any private information conveyed in wholesale price offers.

Meanwhile, the value of selling to investors is just the investor price offer:

$$v_{ij}^{S} = P_{ij}.$$

The originator will choose to sell whenever $v_{ij}^S > v_{ij}^K$. In this setting, the non-committal nature of the auctions provides originators with two incentives to take loans to request bids. First, taking loans to auctions provides *option value*, since originators can choose whether to accept bids. Furthermore, bids also have an *information value*–originators can update their expectations about the value of future cash flows by conditioning on the bid(s) received at auction.

Before moving on to discuss the econometric model, we briefly flag a few important assumptions embedded in the model specification. First, note how servicing costs enter the model: they are assumed to be constant and are not allowed to vary between originators. This assumption is strong, especially in light of previous discussion of returns to scale in servicing. However, because the model aims to capture the incentives of small- and medium-sized originators, we can think of this term as embodying the average cost of servicing for this specific group. While there are differences in size between originator (as documented in Table 1), these differences pale in comparison to the difference between originators in our model and the largest industry players who may originate and sell hundreds of thousands of mortgages per year.

A second set of assumptions has to do with the way GSE prices enter the model. By not indexing GSE price $P_i^0$ with $j$, we're assuming that GSE cash window prices do not vary substantially between originators. Furthermore, the expected value $\mathbb{E}\left[M_i|\cdot\right]$ does not condition on $P^0$. That is, we assume that GSE prices are irrelevant to originators' beliefs about a particular loan's survival prospects. This is not unreasonable, as the GSEs price loans on coarse-grained characteristics, not on a loan-by-loan basis. More details on the GSEs cash window are included in Appendix B.

## 5.2 Econometric Model

We now specify some structure for the model of cash flows $M$, signals $S$, and investor prices $P$:

$$M_{ij} = \mu^M(Z_i) + \varepsilon_{ij} \tag{4}$$

$$S_{ij} = \omega_{ij} \tag{5}$$

$$P_{ij} = \mu_j^P(Z_i) + \eta_{ij} \tag{6}$$

where characteristics $Z_i$ include payoff-relevant observables such as information about the borrower(s), the loan, and the state of the economy.

We assume that the cash flow residual $\varepsilon$, the originator signal $\omega$, and the bid residual $\eta$ are jointly normally distributed:

$$\begin{pmatrix} \varepsilon \\ \omega \\ \eta \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \sigma_{MM} & \sigma_{MS} & \sigma_{MP} \\ \sigma_{MS} & \sigma_{SS} & \sigma_{SP} \\ \sigma_{MP} & \sigma_{SP} & \sigma_{PP} \end{pmatrix}\right)$$

We will assume that originators have rational expectations, in the sense that they know the deterministic component of cash flows $\mu^M$ and bids $\mu^P$, but must take expectations of the residuals $\varepsilon$, $\omega$, and $\eta$ on the basis of their true joint distribution.

The covariance parameters capture the information structure of the model. $\sigma_{MS}$ captures the informativeness of the originator's signal–if $\sigma_{MS}$ is large, the originator possesses more payoff-relevant information about the loan not contained in $Z_i$. Similarly, $\sigma_{MP}$ captures the informativeness of the bid function–large values of $\sigma_{MP}$ imply that bidders bid higher for loans that yield greater cash flows over and above what we can predict on the basis of observables. Finally, $\sigma_{SP}$ represents how accurately originators anticipate investor prices at auction over and above the deterministic component of the bid which depends on loan characteristics.

Written this way, we can see that investor prices convey information through two mechanisms, which we will call an *informativeness mechanism* and a *signal purification mechanism.* The informativeness mechanism operates through $\sigma_{MP}$; it is the direct mechanism that conveys investors' private information (if any) to originators. The signal purification mechanism operates indirectly through $\sigma_{SP}$. Because originators signals and investors bid residuals follow a joint distribution, learning the bid residual allows the originator to form more precise expectations about the cash flow residual. This is true even if $\sigma_{MP}$ is zero. That is, bid residuals may be uncorrelated with the cash flow residual, but so long as $\sigma_{MS}$ is non-zero, originators still update on the basis of the bid residual.

The value of keeping a loan can be written as:

$$v^K = P^0 + \mu^M + \omega\left(\frac{\sigma_{MS}\sigma_{PP} - \sigma_{MP}\sigma_{SP}}{\sigma_{SS}\sigma_{PP} - \sigma_{SP}^2}\right) + \eta\left(\frac{\sigma_{MP}\sigma_{SS} - \sigma_{MS}\sigma_{SP}}{\sigma_{SS}\sigma_{PP} - \sigma_{SP}^2}\right) - c$$

Equating this quantity with $v^S = \mu^P + \eta$ gives a threshold function $\bar{\eta}(\omega)$:

$$\bar{\eta}(\omega) = \frac{P^0 + \mu^M - \mu^P - c + \omega\left(\frac{\sigma_{MS}\sigma_{PP} - \sigma_{MP}\sigma_{SP}}{\sigma_{SS}\sigma_{PP} - \sigma_{SP}^2}\right)}{\left(1 - \left(\frac{\sigma_{MP}\sigma_{SS} - \sigma_{MS}\sigma_{SP}}{\sigma_{SS}\sigma_{PP} - \sigma_{SP}^2}\right)\right)}$$

which partitions the signal space into a region $\Omega_{Keep}(\eta, Z, P^0)$ where selling to the GSEs is optimal and a region $\Omega_{Sell}(\eta, Z, P^0)$ where selling to investors is optimal.

Figure 5 shows the valuation of the keep and sell options in the bid residual. The value of the sell option is straightforwardly increasing in the bid residual. By contrast, the value of the keep option depends on the covariance parameters. Intuitively, when originators are uninformed, higher bids lead originators to expect expect higher cash flows. By contrast, when the originator signals are more informative than investors' bids, higher bids can actually lead to lower expected cash flows. Other things being equal, as $\sigma_{MS}$ increases, the slope of the keep valuation becomes flatter. As $\sigma_{MP}$ increases, the slope of the keep valuation becomes steeper. The effect of increasing $\sigma_{SP}$ is ambiguous, and depends on the relative magnitudes of $\sigma_{MS}$ and $\sigma_{MP}$.

The descriptive exercise of Section 4.1 strongly suggests that we are in the world of Figure 5a; however, the slope of the keep valuation in $\eta$ is unclear.

## 5.3 Estimation

### 5.3.1 Measuring Cash Window Prices

In our setting, originators who take loans to auction yet sell to the GSEs do so almost exclusively through the cash window. The Optimal Blue platform integrates the GSEs' pricing and execution engines, allowing originators to quickly query updated cash window offers. A unique advantage of our dataset is that cash window prices are observed in the majority of relevant cases. Cash window prices are observed for approximately 35% of all loans auctioned on the platform between January 2018 and July 2022. Of the remaining loans, most were originated by originators who were not observed using the cash window at the time of origination and thus fall outside the purview of our question.

For some of the remaining cases, originators are known (based on securitization data) to have securitized the loan through the cash window. Cash window prices may not appear because the originator did not query prices through the platform; however, it can be assumed that originators would be aware of cash window prices for all eligible loans. We use a random forest prediction model to impute values of cash window prices where missing. Since the goal of this exercise is *merely* to impute missing cash window prices, *not* to recover the GSEs' valuation of the loan characteristics, we can use covariates such as origination date that are informative about prices without worrying about connecting these covariates to economic conditions (yield curve, TBA prices, etc.) known to all auction participants. As shown in Figure 6, the origination date is an important predictor of GSE prices, providing considerably more predictive power than loan and borrower characteristics alone.

### 5.3.2 Modeling Cash Flows

Because we use loans originated through February 2020, we only have three years of performance data for every loan in our sample. Despite this, cash flows in a loan's earliest years still provide a decent sense of a loan performance. Nonetheless, some loans survive for a considerable time, yielding cash flows long-after the first three years. Thus, we model cash flows using a tobit model with truncated measure $M_i$:

$$M_i^* = Z_i\beta + \alpha_{t_i} + \varepsilon_i, \qquad M_i = \begin{cases} M_i^* & \text{if} \\ \bar{M} & \text{otherwise} \end{cases}$$

where $\bar{M}$ is the three year cash flows for a loan following the typical payment schedule. Here, $Z_i$ is a vector of loan and borrower characteristics (credit score, loan amount, etc.), as well as controls for the yield curve and securities prices in the To-Be-Announced market, designed to account for observable market conditions. The term $\alpha_{t_i}$ is a (monthly) time effect which we assume to be (a) zero in expectation, conditional on loan characteristics and (b) independent of $\varepsilon$, $\omega$, and $\eta$. That is, $\alpha_{t_i}$ captures cohort-wide shocks to loan performance that are unanticipated to both sellers and investors (e.g., market-wide shocks from COVID-19).

To capture originators' beliefs about the relationship between loan characteristics and cash flows, we estimate our cash flow model on loans originated between 2013 and 2017. To guarantee that model estimates obtained on the $2013-2017$ sample reflect originators' beliefs about loans originated in $2018-2019$, we restrict our sample by dropping brokered loans and retail loans originated and securitized by large integrated institutions. We retain all correspondent loans as well as retail loans from small originators.

Figure 7 shows the distribution of three-year cash flows for $2013-2019$, both in their raw and normalized forms. We notice immediately that loans originated in 2018 and 2019 yield lower cash flows than loans originated in previous years, despite borrowers and loans having comparable characteristics, as shown in Table 4. This is due to historically low interest rates in $2020-2021$ incentivizing higher rates of refinancing activity than in previous years. This systematic divergence of realized from expected cash flows highlights the importance of $\alpha_{t_i}$ in the cash flow model. By assuming that $\alpha_{t_i}$ is zero in expectation, we are saying that past performance provides the best indication of future performance, while still allowing for aggregate shocks that drive a wedge between realized and expected cash flows.

#### 5.3.3 Modeling Auction Bids

To model the deterministic component of investors' bid functions, we estimate a regression model which is a reduced form representation of the max bid formation process:

$$Y_i = Z_i\beta + \gamma_i + \eta_i. \tag{7}$$

Here $Y_i$ is the maximum bid received for loan $i$ at auction, $Z_i$ is a set of loan and borrower characteristics, and $\gamma_i$ is a (high-dimensional) set of fixed effects.

While the covariates in $X_i$ overlap considerably with those in the performance model, the bid model incorporates additional fixed effects for important covariate bins that drive bids over and above their influence on future servicing cash flows. First, we allow for credit score by loan-to-value ratio bins that represent the loan level price adjustments assessed to securitizing institutions by Fannie Mae and Freddie Mac. As discussed in Appendix B.2, these bins occur in 20 point increments for credit score and 5% increments in loan-to-value ratio. Additionally, we bin loan amounts according to well-known loan-size thresholds used to create custom securities (see Huh and Kim (2022) and Huh and Kim (2023)).[23]

[23] Bin cutoffs are placed at 85k and 110k, then in 25k increments from 125k to 275k. These cutoffs are commonly used by issuers creating custom, single-issuer swap securities, but are also relied upon by the GSEs who pay a premium for small loans through the cash window in increments reflecting traditional loan-

Investors in the auctions face the choice to securitize loans with Fannie Mae or Freddie Mac through the cash window or by using swap transactions. As depicted in Figure 1a, the investors in the auctions fall into three rough groups: (a) investors who use exclusively swap transactions, (b) investors who use exclusively cash-window transactions, and (c) investors who use a mix of cash window and swap transactions. The largest investors by loan volume tend to rely almost exclusively on swap transactions. Smaller investors rely more heavily on the cash window, owing to a requirement that securities be assembled from comparable loans–when investors have few loans sharing similar loan size, term, maturity, and note rate characteristics, the cash window becomes their primary option for timely securitization.

To control for market conditions affecting investors' prices, we include a variable for the current TBA price, which is common knowledge to both investors and originators. Investors using loan swaps may be able to place a given loan in one of multiple securities with varying coupon rates.[24] The value of this choice depends on the trading price of securities with that coupon—differences in prices by coupon depicted in Figure 8b. Because the realized value of the coupon choice is endogenous, we control for TBA prices by using the modal coupon for a given interest rate. Furthermore, since TBA securities are traded on a forwards market, multiple securities with the same coupon trade simultaneously, with the only difference being the remaining months until settlement. Figure 8a shows trading prices for securities with a coupon of 2.0 trading at 0, 1, 2, and 3 months from settlement. We control for market conditions using a two-month settlement period to avoid the relative volatility of zero-month trades and ensure that sufficient trading volume is observed (unlike with three-month trades).

Finally, since the modeled auction prices are the result of competitive forces, it is important to control for differences in competition across sellers and loans. To do this, we add a control for the "network size" of the seller of loan $i$ which we allow to vary at the seller-by-quarter level. Intuitively, the maximum bid is likely to differ between auctions, depending on the number of (competitive) investors and the winner's curse.

## 5.4 Estimation and Identification

We estimate the full model using maximum likelihood, integrating out over the unobserved signal dimension with gauss-hermite quadrature. Full details of the estimation procedure can be found in Appendix D.3 Before moving on to results, we discuss sources of identifying

size cutoffs. For documentation of this practice, see Freddie Mac's cash specified payups characteristics at: https://sf.freddiemac.com/working-with-us/selling-delivery/delivery-options-pricing/cash-payups

[24]Per Fannie Mae and Freddie Mac guidelines, the minimum spread between a loans note rate and the coupon is 25 basis points and the maximum is 112.5 basis points.

variation in the model.

Because the model is agnostic about the interpretation of the originator's signal, we can normalize the signal variance $\sigma_{SS}$ to be 1. The signal only enters the model through originators' expectations of cash flows; thus while the magnitudes of the covariance terms $\sigma_{MS}$ and $\sigma_{SP}$ depend on the scale of the signal, their relative values $\frac{\sigma_{MS}}{\sigma_S}$ and $\frac{\sigma_{SP}}{\sigma_S}$ do not.

The variance terms can be identified outside of the maximum likelihood routine, in separate cash flow and bid regressions. Bids are observed for all loans, thus there is no need estimate $\sigma_{PP}$ together with the covariance and servicing cost terms. Similarly, cash flows are observed for all loans, thus the cash flow residual $\varepsilon$ is the residual of a the latent cash flows in a tobit model, the variance of which $\sigma_{MM}$ can be estimated offline.

For the covariance terms, $\sigma_{MP}$ can be straightforwardly identified off the observed correlation between bids and cash flows. The remaining covariance terms $\sigma_{MS}$ and $\sigma_{SP}$, as well as the cost of servicing $c$ are identified from the originator's decision to keep or sell loans. From the threshold Equation 5.2, we see that these terms all enter the model through by changing the behavorior of the threshold. Inside the threshold function are GSE prices $P^0$ and investor prices $P_{ij}$, which can move independently of expected loan performance $\mu_M$ to trace out the decision boundary, from which these terms are recovered.

# 6 Results

## 6.1 Model Estimates

Table 5 reports the important parameter estimates for the full model. Loan size is an extremely important driver of both loan performance and investor prices. Larger loans yield greater *absolute* cash flows, but lower cash flows per \$100 of loan volume because they face greater strategic incentive to refinance when rates drop. This is reflected in investor bids.

The originator's network size is important for capturing investor pricing behavior, with an additional participating investor leading to an expected increase of \$0.03 in the maximum price. Notice also that TBA price is an important control in the bid model, with a coefficient close to 1. This is to be expected, since TBA prices are a very close proxy for the value of a loan.

Note that the estimated cash flow residual $\sigma_{MM}$ is roughly one quarter the magnitude of the bid residual $\sigma_{PP}$. This is consistent with loans investors valuing loans for reasons over and above expected cash flows. As discussed in Section 5.3.3, investors reap value from loans not only through servicing revenues and cash window prices, but also through the option value of selling securities backed by their loans.

The estimated value of $\sigma_{MS}$ is positive, large, and significant, indicating that originators are privately informed about the value of loans.

Strikingly, originators have *strong* expectations about bids received at auction, as indicated by the large value of $\sigma_{SP}$. This fact should not be entirely puzzling, however. Bids are driven by investors' liquidity needs and their existing portfolios of loans eligible for securitization among other things. These are sources of idiosyncratic valuations that change over the long-run but are relatively stable in the short-run. Because originators hold numerous auctions for investors, they receive multiple bids reflecting the same idiosyncratic liquidity needs within a short period of time. This would allow originators to form more precise expectations about bids than they could on loan characteristics alone.

Note that given the estimated covariance structure, the value of keeping a loan increases in $\omega$ but actually *decreases* in the bid residual $\eta$:

$$\frac{d\hat{v}_K}{d\omega} = \left(\frac{\sigma_{MS}\sigma_{PP} - \sigma_{MP}\sigma_{SP}}{\sigma_{SS}\sigma_{PP} - \sigma_{SP}^2}\right) \approx 0.33, \qquad \frac{d\hat{v}_K}{d\eta} = \left(\frac{\sigma_{MP}\sigma_{SS} - \sigma_{MS}\sigma_{SP}}{\sigma_{SS}\sigma_{PP} - \sigma_{SP}^2}\right) \approx -0.25$$

However, this is not because investor prices are uninformative. Rather, investor prices are still positively correlated with loan performance, but much of the information conveyed by bids is already contained in the private signal. The marginal dependence of these two value derivatives on the covariance parameters is shown in Figure 10. For the estimated values of $\sigma_{MS}$ and $\sigma_{MP}$, only very low values of $\sigma_{SP}$ would allow for investor bids to increase the expected value of keeping the loan (given the signal value). Note that at these estimated values, originators accept higher price offers and reject lower price offers, consistent with the descriptive results in Section 4.1. Lastly, when the originator's private signal is higher, the price required to sell to investors is higher.

Finally, we estimate servicing costs of \$0.54 per \$100 of loan volume. This translates to servicing costs of approximately \$700–1900 per loan or about two years of servicing revenues. To put this number in context, the Mortgage Banker's Association reports servicing costs ranging from \$240 to \$320 per loan from 2013 to 2022. However, these numbers come from survey respondents consisting mostly of large integrated institutions (e.g., Rocket Mortgage or Wells Fargo). Thus, these lower survey numbers are not necessarily representative of the costs faced by originators in our setting, who tend to be much smaller. Moreover, our cost parameter incorporates the full opportunity cost of both securitization and servicing, not merely the accounting cost of servicing. This opportunity cost can include costs not well-captured in our measure of cash flows, such as the expected cost of fronting lost servicing payments in the case of loan default or the haircut taken in the event of forced GSE loan repurchase.

Figure 9 shows the model fit. The observed and modeled bid and performance distributions fit relatively well. The bid distribution is slightly flatter for the modeled bids than the actual data, suggesting that the assumption of normal errors is not overly strong. Model performance matches observed performance somewhat less well, though fit is not terrible. Relative to the data, the model predicts marginally lower censoring probability, but a longer left tail. This is because the model does not technically constrain performance to be greater than zero.

## 6.2 Evaluating Adverse Selection

We can now think about the 'ex-ante value' of going to auction as:

$$\begin{aligned} v_{ij}^A &= \underbrace{\left(P_i^0 + \mathbb{E}\left[M_i | S_{ij}, v_{ij}^K > v_{ij}^S\right] - c\right)}_{\text{Expected Retail Value of Loan w/ Rejected Bid}} \cdot \underbrace{Pr(v_{ij}^K > v_{ij}^S | S_{ij})}_{\text{Probability of Rejecting Bid}} \\ &+ \underbrace{\mathbb{E}\left[P_{ij} | S_{ij}, v_{ij}^K \leq v_{ij}^S\right]}_{\text{Value of Accepted Bid}} \cdot \underbrace{Pr(v_{ij}^K \leq v_{ij}^S | S_{ij})}_{\text{Probability of Accepting Bid}} \end{aligned}$$

which gives us a straightforward way of evaluating the value of the auction environment to originators. We can compare this quantity to the value to the originator of securitizing their loan:

$$P_i^0 + \mathbb{E}\left[M_i | S_{ij}\right] - c,$$

and the bid received when selling at auction.

Averaging over all loans in the sample, we see that the average investor bid is \$103.55, compared to an average value of securitizing of \$103.20. Thus, were originators (in the aggregate) forced to decide between securitizing all loans and selling all loans via auctions, they would prefer to sell all loans. However, the option value of the auction environment results in an average value of \$103.75. Relative to securizing all loans on their own, originators value auctions at about \$0.55 (per \$100 loan volume) or about \$1500 for an average-size loan.

## 6.3 Baseline Comparison Estimates

To put our above estimates in perspective, we can re-estimate our model for a subset of originators who do not securitize with the GSEs during this period. These are originators for whom investors do not need to account for the possibility of adverse selection. Because we observe both servicing cash flows and investor prices for all such loans, we can straightforwardly estimate a variant of the model without originator signals. The $\sigma_{MP}$ parameter in this alternative model is now more straightforwardly interpreted as the level of informativeness of the bidders. Estimates for this specification are included in Table 6.

Note that the estimated value of $\sigma_{MP}$ is substantially higher in this setting (0.108 vs 0.046). Accounting for the slightly higher investor price variance, this still amounts to sizeable increase in the estimated correlation from 0.22 to 0.48. This difference suggests that investors facing sellers with the ability to adversely select loans into the wholesale market shade their bids downward to avoid conveying more information that originators could leverage when deciding to securitize or sell.

Under this alternative model, investors pricing behavior is allowed to differ from the model with a securitization option. Investors actually submit *higher* bids for originators who do not securitize with the GSEs. Using the benchmark model to predict bids for the estimation sample, we predict that investors would offer prices that are approximately \$0.051 higher per loan, or about \$140 for an average-size loan.

# 7 Counterfactuals

In this section, I use the model estimates to conduct two counterfactual exercises designed to shed light on the effect of private information and adverse selection in the wholesale market as currently designed. The first counterfactual considers a world without wholesale market auctions, while the second counterfactual considers restricting auction access to eliminate adverse selection at the loan-level.

## 7.1 Counterfactual I: Lock-Price Market

In our first counterfactual exercise, we ask what would happen if we shut down the auction market entirely. We assume that in this world, all wholesale transactions would take place through via lock prices. Recall that in this market, investors only submit coarse-grained price offers, and do not price loans individually.

For this counterfactual, we cannot use the same data from our estimation sample. Because we observe only the maximum price offer for each investor, the lock prices we observe in the estimation sample are truncated–only observed in the case that they are higher than the investor's bulk bid. Thus, we cannot straightforwardly recover the maximum lock price for each loan in our sample. Instead, we use data from the period January 2018 to August 2018, where the true maximum lock price is known.

During this period, the lock price is higher than investor bids in 25% of auctions. In the remainder of cases where investor prices are higher than lock prices, the average difference between the maximum investor price and the maximum lock price is \$0.34. This contrasts with an average difference of \$0.60 reported above in Table 1 when the maximum *observed* lock price is used.

In the counterfactual specification, we assume that the true maximum lock price would remain unchanged from its observed value. This assumption is not completely innocuous, since lock prices reflect investor beliefs about the average loan quality for a group of loans. However, loans acquired at auction make up a small fraction of loans that investors acquire on the wholesale market. Meanwhile, lock prices are set based on the posted price market as a whole. Thus, during this time, lock prices would be unlikely to change considerably if the ability to purchase loans at auction were to disappear. Nonetheless, we would caution against interpreting these results beyond the limited application at hand.

Absent a fine-grained bid, the value to the originator of keeping a loan is now:

$$v^K = P^0 + \mu^M + \omega \left( \frac{\sigma_{MS}}{\sigma_{SS}} \right) - c$$

Equating this quantity with $v_{ij}^S = P^{Lock}$ gives a threshold in $\omega$:

$$\bar{\omega} = \frac{\sigma_{SS}}{\sigma_{MS}} \left( P^{Lock} - P^0 - \mu^M + c \right)$$

For signal values above this, the originator chooses to keep the loan. Note that because the originator cannot observe a bid residual, the decision to sell or securitize depends entirely on the originator's private signal.

To conduct this counterfactual exercise, we draw signal values for each loan conditional on the recovered bid residual:

$$\eta_{ij} = P_{ij} - \mu_{ij}^P$$

We also condition our sample draws on the performance residual. For loans prepaid within three years we directly observed the performance residual, and for loans that do not prepay within three years we obtain a lower bound for the performance residual. Finally, we condition on the keep/sell outcome using the decision rule in equation 5.2. We sample subject to these linear inequalities using Gibbs sampling as in Rodriguez-Yam et al. (2004). We simulate the counterfactual 1000 times.

Table 7 summarizes the results of our counterfactual exercise. We can think about four distinct 'types' of loans, based on whether they are kept or sold in factual and counterfactual scenarios. The majority of loans are either always sold or always securitized. However, for a still sizeable fraction of loans, the realized outcome depends on whether auctions or lock prices are used.

In a world with only lock prices, originators securitize 64% of loans and sell only 36%. When auctions are introduced, originators securitize only 45% and sell 55%. This difference owes to two types of loans.

First, 23% of loans are securitized when only lock prices are used, but sold when auctions are used. These are loans for which the originator's signal was high, but swamped by the possibility of receiving higher price offers at auction.

Second, there is a small fraction (only 4%) of loans that are sold when only lock prices are used, but securitized when auctions are introduced. These are loans for which the originator's signal was low, but still swamped in expectation by the positive information value of going to auction.

For both types of loans, auctions allow the originators to improve their allocative decisions. The first and most common loan type has a negative average cash flow residual, amounting to approximately \$100 per loan. The second and less common loan type has a strong positive average cash flow residual, amounting to approximately \$275 per loan.

These counterfactual results have implications for market efficiency. The presence of an auction market allows for a greater fraction of loans to be serviced by large integrated investors. Because originators have high estimated cost of servicing (most likely higher than large investors), auctions reduce wasteful servicing expenditures.

## 7.2 Counterfactual II: Exclusive Auction or Cash Window

In our second counterfactual exercise, we ask what would happen if loan originators were forced to choose between the auction channel and the cash window for their entire loan portfolio. As a policy, this could represent a decision on the part of the GSEs not to do business with sellers using third-party sales or (more realistically) a decision by the auction platform to exclude sellers who routinely sell loans through the cash window.

Section 6.2 argued that when averaging across all sellers, the average value of the investor bid exceeded the expected value of servicing loans. However, two sources of seller-level heterogeneity make this counterfactual non-trivial. First, sellers differ in the portfolio of loans they originate and the network of bidders they interact with. While it is beyond the scope of this paper to consider the effect of secondary markets on sellers' origination decisions and wholesale interactions, it remains true that sellers exhibit considerable heterogeneity on these dimensions. Thus, some sellers may originate loans that are of disproportionately high value to GSEs or (their network of) investors. Sellers with attractive GSE portfolios have a greater incentive to securitize loans, other things being equal.

Second, and more importantly, sellers likely have idiosyncratic costs of servicing and securitization. Lower-cost sellers will have a greater incentive to securitize their own loans, while higher cost sellers have a greater incentive to sell. The baseline model of Section 5 smoothed over differences between sellers for ease of exposition, but we can easily re-estimate the model with seller-specific costs. Now the value of keeping a loan can be written as:

$$v^K = P^0 + \mu^M + \omega \left( \frac{\sigma_{MS}\sigma_{PP} - \sigma_{MP}\sigma_{SP}}{\sigma_{SS}\sigma_{PP} - \sigma_{SP}^2} \right) + \eta \left( \frac{\sigma_{MP}\sigma_{SS} - \sigma_{MS}\sigma_{SP}}{\sigma_{SS}\sigma_{PP} - \sigma_{SP}^2} \right) - c_j$$

and sellers' decisions closely resemble their analogues from the baseline model. In order to include all sellers in the estimation, we include separate securitization costs for the top 10 sellers by origination volume and we aggregate over the remaining sellers.[25] Thus the maximum likelihood routine searches over 14 parameters–three information structure parameters and 11 securitization cost parameters.

Table 8 shows the model estimates with seller-specific costs of servicing. The information structure parameter estimates closely resemble those obtained in the baseline exercise. As before, we find that residual bids are less informative than sellers' private information. Note, however, that the parameter for the correlation between the sellers' signals and bidders' residual bids is now lower than in the baseline results.

Note also the distribution of estimated seller costs, which range from \$0.50-0.65 (per \$100 loan volume). Somewhat surprisingly, small sellers have lower estimated securitization costs than their larger, non-aggregated counterparts. As an added benefit, the counterfactual model better-rationalizes the behavior of sellers excluded from consideration when estimating the baseline model. The baseline model excluded sellers who sold exclusively to the GSEs or to third-party investors. This can now be explained by sellers having securitization costs that are either extremely low or extremely high, respectively.

[25] In Appendix C, we show that the securitization cost estimates are not overly sensitive to the choice to aggregate over small sellers.

### 7.2.1 Counterfactual Simulation

In the counterfactual world, originators must now decide whether to use auctions for their entire portfolio or use the cash window for their entire portfolio. Thus we need to compute two expected portfolio values for each seller:

$$V_j^{\mathrm{S}} = \sum_{i \mid j(i)=j} v_{ij}^{S} = \sum_{i \mid j(i)=j} \bar{b}\left(Z_{ij}\right)$$
$$V_j^{\mathrm{K}} = \sum_{i \mid j(i)=j} v_{ij}^{K} = \sum_{i \mid j(i)=j} \mu^{M}\left(Z_{ij}\right)$$

In this expectation, originators no longer have the option value of securitizing good loans and selling bad ones, so we must sum over all loans in their portfolio. For the keep value, the expected value of the portfolio is just the sum of the deterministic components of loan performance. For the sale value, recall that in Section 6.3, we showed that auction investors bid higher when facing sellers without the choice between the auction and cash window channels. Thus, rather than using realized bids for the counterfactual, we simulate bids $\bar{b}$ using the alternative model estimated on sellers with *no* adverse selection potential.

For the counterfactual, we will examine the decisions of small sellers individually. That is, we consider the loan portfolios of each individual seller, rather than treating them as a single large seller whose loan portfolio is the aggregate of all small sellers' portfolios. However, since we do not estimate separate securitization costs for each small seller, we will treat each seller as though their costs were given by the model's small seller aggregate cost. To the extent that there is true underlying heterogeneity in costs between small sellers, we should treat counterfactual predictions with caution.

Table 9 shows the change in firms' average portfolio values under the counterfactual. In the counterfactual environment, the majority of sellers would choose to sell loans exclusively to third-party investors. However, a small handful of sellers (2 large sellers, and 4 small sellers) would opt to securitize their entire portfolio with the GSEs. Overall, while over 60% of loans are sold to third-party investors in the existing auction environment, approximately 90% of loans are sold to third-party investors in the counterfactual world. On average, investors would acquire loans of comparable quality, but would pay \$0.16 less per loan.

Most sellers are worse off under the counterfactual, losing on average about half of one year's servicing revenues in overall value. However, a small number of sellers experience a slight increase in welfare under the counterfactual. All such sellers choose to sell to investors in the counterfactual. For these sellers, the higher bids investors submit when facing a seller who cannot adversely select loans for sale outweighs the benefit of being able to choose between keeping and selling loan-by-loan. This welfare increase raises the question as to why these sellers do not already commit to selling all loans to investors. A comprehensive answer lies outside the scope of this paper, but one potential explanation is that sellers who service large loan portfolios cannot credibly commit to selling all loans.

Before concluding, note one limitation of this counterfactual exercise. In the counterfactual, we hold fixed the characteristics of sellers' origination portfolios. That is, we do not currently consider how sellers' origination activities might be affected by the counterfactual policy change. Appendix E briefly explores the impact of auction use on loan origination, finding some evidence that origination volumes increase in response. However, the potential increase in origination volume is small relative to the size of originators' loan portfolios, and is thus not a first-order concern for us here.

# 8 Conclusion

This paper has provided evidence to suggest that wholesale mortgage markets experience an adverse selection problem arising when informed originators meet with (relatively) uninformed investors. This adverse selection incentive is exacerbated by the structure of certain wholesale market transactions, which are non-committal for the originators.

This matters for a few reasons. First, there are straightforward efficiency concerns. Investors in wholesale markets often have elaborate servicing infrastructure in place; if adverse selection results in small originators keeping and securitizing more loans on their own, this suggests that servicing rights may not be allocated efficiently. Additionally, adverse selection in the wholesale market can spill over into the origination and securities markets by changing the incentives of originators to make loans and by changing the mix of loans available to be securitized. While these spillovers are beyond the scope of this paper, they nonetheless raise important considerations for the study of the origination and securitization markets. Future work needs to be attentive not only to the presence of the wholesale market, but also the potential role of information asymmetries within it. This paper takes a first step toward understanding both where these asymmetries appear and why they persist.

# References


Sumit Agarwal, Yan Chang, and Abdullah Yavas. Adverse selection in mortgage securitization. *Journal of Financial Economics*, 105(3):640–660, 2012.

Darren J Aiello. Financially constrained mortgage servicers. *Journal of Financial Economics*, 144(2):590–610, 2022.

Jonathan Becker, Kenneth Hendricks, Jean-François Houde, and Diwakar Raisingh. Asymmetric information and the supply-chain of mortgages: The case of ginnie mae loans. 2023.

David Benson, You Suk Kim, and Karen Pence. Bank aggregator exit, credit supply, and borrower outcomes in the mortgage industry. 2023.

Greg Buchak, Vera Chau, and Adam Jørring. Integrated intermediation and fintech market power. *Available at SSRN 4552425*, 2023.

Sharad Chaudhary. A close look at the gse cash window, 2020. URL https://www.machinesp.com/post/a-close-look-at-the-gse-cash-window.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters, 2018.

Pierre-André Chiappori and Bernard Salanie. Testing for asymmetric information in insurance markets. *Journal of political Economy*, 108(1):56–78, 2000.

Eduardo Dávila and Cecilia Parlatore. Identifying price informativeness. Technical report, National Bureau of Economic Research, 2018.

Yongheng Deng, John M Quigley, and Robert Van Order. Mortgage terminations, heterogeneity and the exercise of mortgage options. *Econometrica*, 68(2):275–307, 2000.

Moussa Diop and Chen Zheng. Mortgage servicing fees and servicer incentives during loss mitigation. *Management Science*, 69(11):7118–7150, 2023.

Chris Downing, Dwight Jaffee, and Nancy Wallace. Is the market for mortgage-backed securities a market for lemons? *The Review of Financial Studies*, 22(7):2457–2494, 2009.

Liran Einav, Mark Jenkins, and Jonathan Levin. Contract pricing in consumer credit markets. *Econometrica*, 80(4):1387–1432, 2012.

Liran Einav, Mark Jenkins, and Jonathan Levin. The impact of credit scoring on consumer lending. *The RAND Journal of Economics*, 44(2):249–274, 2013.

Amy Finkelstein and James Poterba. Testing for asymmetric information using "unused observables" in insurance markets: Evidence from the uk annuity market. *Journal of Risk and Insurance*, 81(4):709–734, 2014.

Laurie Goodman, Janneke Ratcliffe, Michael Neal, Jung Hyun Choi, Linna Zhu, John Walsh, Caitlin Young, Daniel Pang, Amalie Zinn, Katie Visalli, et al. Housing finance: At a glance monthly chartbook, february 2023. 2023.

Yesol Huh and You Suk Kim. The real effects of secondary market trading structure: Evidence from the mortgage market. *The Review of Financial Studies*, 35(8):3574–3616, 2022.

Yesol Huh and You Suk Kim. Cheapest-to-deliver pricing, optimal mbs securitization, and welfare implications. *Journal of Financial Economics*, 150(1):68–93, 2023.

Hemant Ishwaran, Udaya B Kogalur, Eugene H Blackstone, and Michael S Lauer. Random survival forests. 2008.

Hemant Ishwaran, Thomas A Gerds, Udaya B Kogalur, Richard D Moore, Stephen J Gange, and Bryan M Lau. Random survival forests for competing risks. *Biostatistics*, 15(4):757–773, 2014.

You Suk Kim, Steven M Laufer, Richard Stanton, Nancy Wallace, and Karen Pence. Liquidity crises in the mortgage market. *Brookings Papers on Economic Activity*, 2018(1): 347–428, 2018.

You Suk Kim, Donghoon Lee, Therese C Scharlemann, and James I Vickery. Intermediation frictions in debt relief: evidence from cares act forbearance. *FRB of New York Staff Report*, (1035), 2022a.

You Suk Kim, Karen Pence, Richard Stanton, Johan Walden, and Nancy Wallace. Nonbanks and mortgage securitization. *Annual Review of Financial Economics*, 14:137–166, 2022b.

Tom Mayock and Lan Shi. Adverse selection in the market for mortgage servicing rights. *Journal of Housing Economics*, 58:101858, 2022.

Anthony Pennington-Cross. Credit history and the performance of prime and nonprime mortgages. *The Journal of Real Estate Finance and Economics*, 27(3):279–301, 2003.

Gabriel Rodriguez-Yam, Richard A Davis, and Louis L Scharf. Efficient gibbs sampling of truncated multivariate normal with application to constrained linear regression. 2004.

Ryan Sandler. Aligning incentives: The effect of mortgage servicing rules on foreclosures and delinquency. *Regional Science and Urban Economics*, 102:103922, 2023.

Lan Shi, Yan Zhang, and Xinlei Shelly Zhao. How do banks respond to non-bank competition in the conforming residential mortgage market? the role of balance sheet. *The Role of Balance Sheet (January 1, 2023)*, 2023.

Richard Stanton, Johan Walden, and Nancy Wallace. The industrial organization of the us residential mortgage market. *Annu. Rev. Financ. Econ.*, 6(1):259–288, 2014.

James I Vickery and Joshua Wright. Tba trading and liquidity in the agency mbs market. *Economic Policy Review*, 19(1), 2013.

Dayin Zhang. Government-sponsored wholesale funding and the industrial organization of bank lending. 2020.

# A Tables and Figures

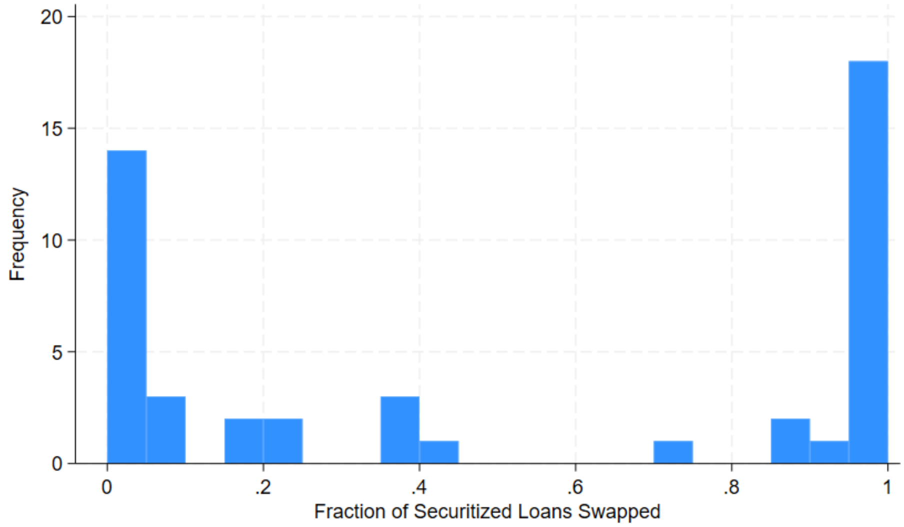
*Table 1: Estimation Sample Summary Statistics*

| | (1) | | | | |
|---|---|---|---|---|---|
| | mean | sd | p25 | p50 | p75 |
| *Loan and Borrower Characteristics* | | | | | |
| Original Loan Amount | 256590 | 117751 | 167000 | 240000 | 332800 |
| Interest Rate | 4.43 | 0.57 | 4 | 4.38 | 4.88 |
| Credit Score | 745.9 | 44.8 | 714 | 753 | 783 |
| Loan-to-Value Ratio | 78.9 | 15.4 | 72.9 | 80 | 90.9 |
| Debt-to-Income Ratio | 36.4 | 9.07 | 30.1 | 37.8 | 43.8 |
| Monthly Income | 8789.8 | 8022.9 | 4916.1 | 7388.8 | 10845 |
| *Auction Characteristics*† | | | | | |
| Fannie Mae Price | 103.0 | 1.49 | 102.0 | 102.9 | 103.8 |
| Freddie Mac Price | 102.9 | 1.61 | 102.0 | 102.8 | 103.7 |
| Highest Bulk Bid | 103.7 | 1.35 | 102.8 | 103.6 | 104.4 |
| Highest Posted Price | 103.1 | 1.24 | 102.3 | 103.1 | 103.8 |
| Number of Participants | 12.7 | 5.56 | 8 | 13 | 16 |
| *Loan Survival Characteristics*†† | | | | | |
| 1(12-Month Survival) | 0.81 | | | | |
| 1(24-Month Survival) | 0.49 | | | | |
| 1(36-Month Survival) | 0.33 | | | | |
| 12-Month Servicing Revenues | 0.23 | 0.039 | 0.22 | 0.24 | 0.25 |
| 24-Month Servicing Revenues | 0.37 | 0.13 | 0.26 | 0.44 | 0.48 |
| 36-Month Servicing Revenues | 0.46 | 0.21 | 0.27 | 0.46 | 0.70 |
| Observations | 108524 | | | | |
| | | | | | |
| *Originator Characteristics* | | | | | |
| Number of Loan Auctioned | 3724 | 3745 | 1038 | 2205 | 4839 |
| Network Size | 16 | 7.7 | 11 | 14 | 21 |
| Observations | 35 | | | | |

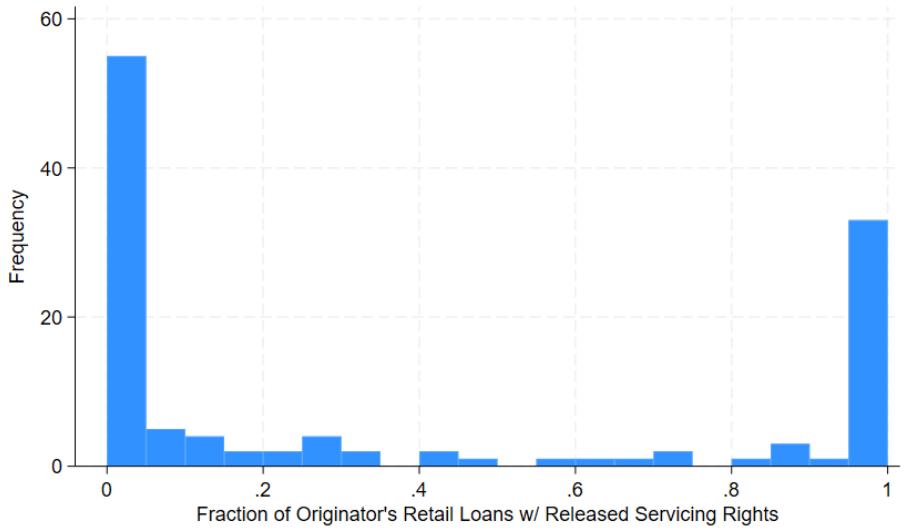
† Note that not all loans receive price offers from Fannie Mae or Freddie Mac. During this period, approximately half of loans are observed to have bids from the GSEs. As discussed in Section 5.3.1, the model of Section 5 uses predicted values for missing GSE prices where necessary. The table displays only the summary statistics for loans where a GSE bid is observed. The summary statistics for the highest bulk bid and the highest posted price do not have this problem–in the estimation sample, almost all loans have at least one observation of each type.

†† Servicing revenues are quoted in their normalized form according to the formula in Section 5.1.1

*Figure 1: Securitization Behavior of Auction Participants*



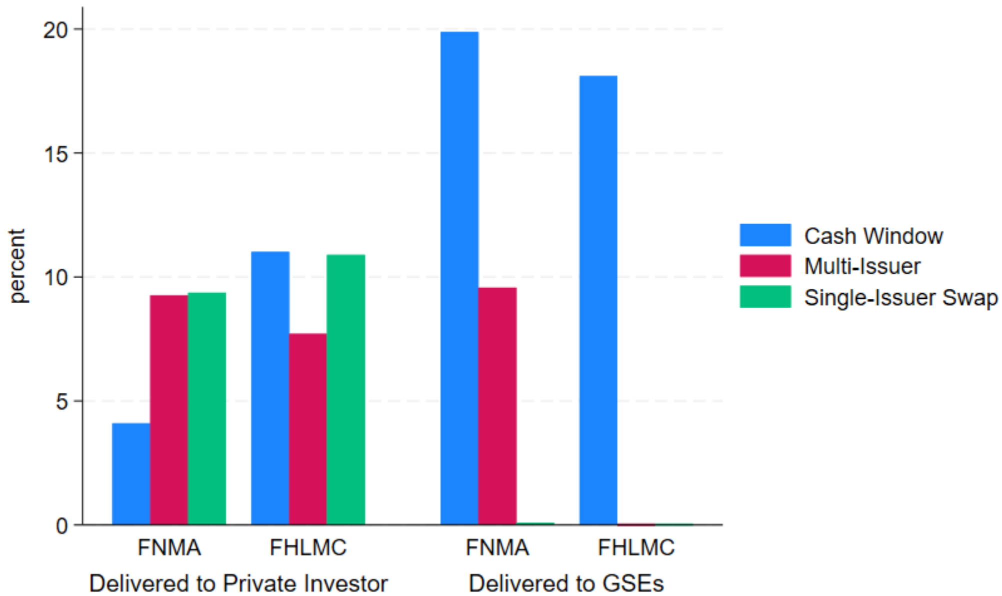

*(a) Distribution of Swap Use by Auction Investors*




*(b) Distribution of Sellers' Fraction of Servicing Released Loans*

Figure 1a shows the histogram of auction investors' fraction of lender swaps and cash window transactions. Note that securitization method is not always known–loans that end up in Fannie Majors securities can be delivered through *either* lender swaps or cash window transactions as discussed in Chaudhary (2020). Thus, the histogram only includes loans where the securitization method is known with certainty. By and large, investors specialize, using either lender swaps or cash window transactions exclusively. Some investors utilize a mix of swaps and cash window transactions. This owes largely to volume considerations–when an investor lacks the volume of loans required to assemble a pool at a given coupon rate, they may choose to sell those loans through the cash window, while assembling pools and using lender swaps for high-volume coupon rates. Figure 1b shows the histogram of auction originators' fraction of loans with retained servicing rights. Originators come in three types: (a) those who service all their own loans, (b) those who outsource all servicing, and (c) those who outsource only *some* servicing. We focus on the largest of these three groups who service all their own loans.

*Figure 2: Securitization Method by GSE and Channel*




The figure depicts the share of securitization method for all auctioned loans that can be matched to the FNMA/FHLMC securities data. Loans sold to investors are securitized using a mix of securitization methods for both Fannie Mae and Freddie Mac. Loans sold directly by sellers to the GSEs predominantly end up in cash window securities, though some loans securitized with Fannie Mae are placed into multi-issuer pools through the Fannie Majors program. However, it is reasonable to assume from the fact that virtually no such loans are placed in single issuer swap securities, that loans placed in the Fannie Majors securities were still acquired by Fannie Mae in exchange for cash.

*Figure 3: Distribution of Investor-GSE Price Spread*

*(a) Full Distribution* *(b) Distribution by Realized Channel*

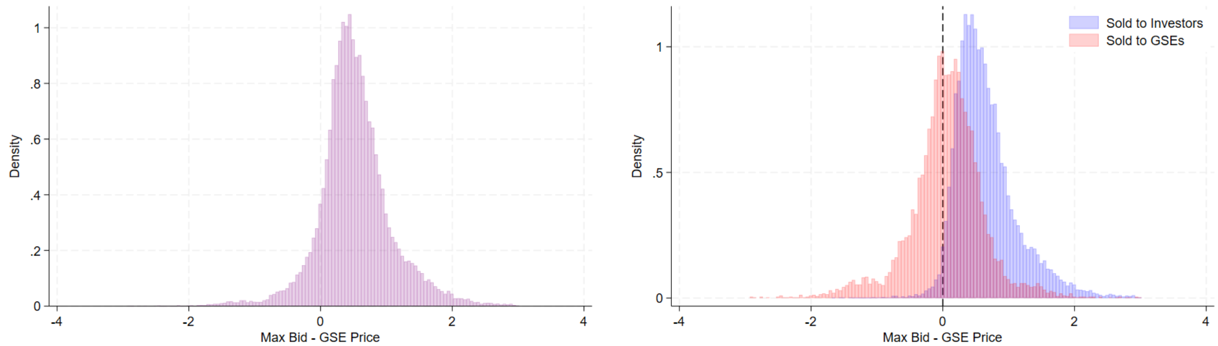



Figure 3a shows the distribution of the spread between the maximum investor bid and the maximum GSE price for all auctioned loans originated before March 2020. Figure 3b shows the distribution based on whether the auctioned loan was ultimately sold to the GSEs (red) or to third-party investors (blue). Almost all (98%) of loans sold to third-party investors have a positive bid spread, indicating that originators place positive value on servicing payments even after accounting for costs associated with servicing and securitization which are not incorporated into the GSE price offers.

*Table 2: Testing for Monotonicity in Investor-GSE Bid Spreads*

| | Model: Probability of Selling to Investors | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Bid Difference | 0.198*** | 0.204*** | 0.190*** | 0.180*** | 0.182*** |
| | (0.00204) | (0.00226) | (0.00226) | (0.00240) | (0.00243) |
| $N$ | 65265 | 65129 | 63316 | 63316 | 63316 |
| $R^2$ | 0.126 | 0.207 | 0.232 | 0.280 | 0.289 |
| Within R2 | | 0.113 | 0.135 | 0.110 | 0.110 |
| State FEs | No | Yes | Yes | Yes | Yes |
| Loan Characteristic Controls | No | No | Yes | Yes | Yes |
| Seller FEs | No | No | No | Yes | Type-Specific |

Standard errors in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Linear probability model estimates for the probability of selling to third-party investors as a function of the spread between investor bids and GSE prices. Model (5) controls for seller fixed effects by 'type' to address the bimodality of the bid spread distribution. Sellers are classified as high-spread or low-spread, corresponding to the two peaks of the bid spread distribution.

*Table 3: Performance Probits: Correlation with Choice to Sell to GSEs*

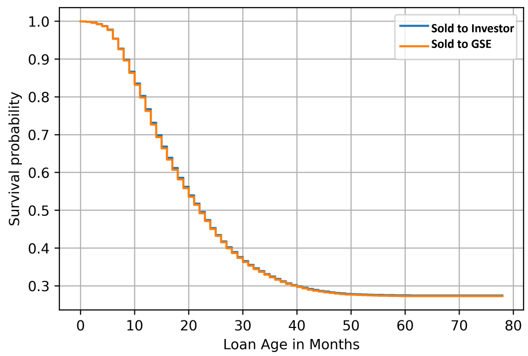
| | Performance Outcome | | | | | |
|---|---|---|---|---|---|---|
| | 1(Default) | 1(12m Surv) | 1(18m Surv) | 1(24m Surv) | 1(30m Surv) | 1(36m Surv) |
| $\rho$ | -0.0907*** | .0109 | .0128* | .0174** | .0185*** | .0214*** |
| | (0.0065) | (.0062) | (.0054) | (.0053) | (.0054) | (.0055) |
| Controls | Loan Size, Credit Score, LTV, DTI, Income, Note Rate | | | | | |
| Fixed Effects | Originator, State, Month, Occupancy, Loan Purpose, LTV-FICO Bins | | | | | |
| Observations | 108,524 | | | | | |

Standard errors in parentheses.
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

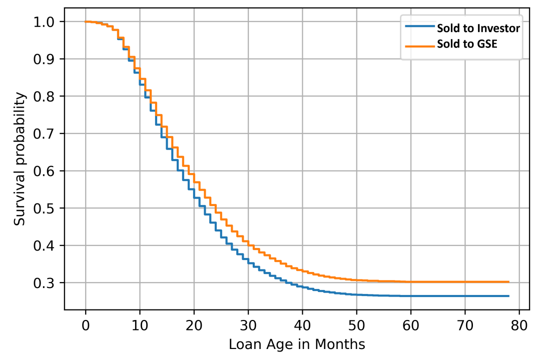
Bivariate probit regressions for binary performance outcomes. Selection outcome is choice to sell to the GSEs. All regressions control for loan and borrower characteristics. Fixed effects for binned values of loan-to-value ratio and credit score are defined according to Fannie Mae and Freddie Mac's loan level price adjustment matrices, which are fixed across the sample period. For 12- through 36-month survival, the negative outcome includes *both* default and prepayment; however, default is exceedingly rare in our conventional loan sample, meaning that the adverse survival outcomes are driven almost entirely by early prepayment.

*Figure 4: Survival Differences Between Channels*

*(a) Channel Differences w/o Decision*

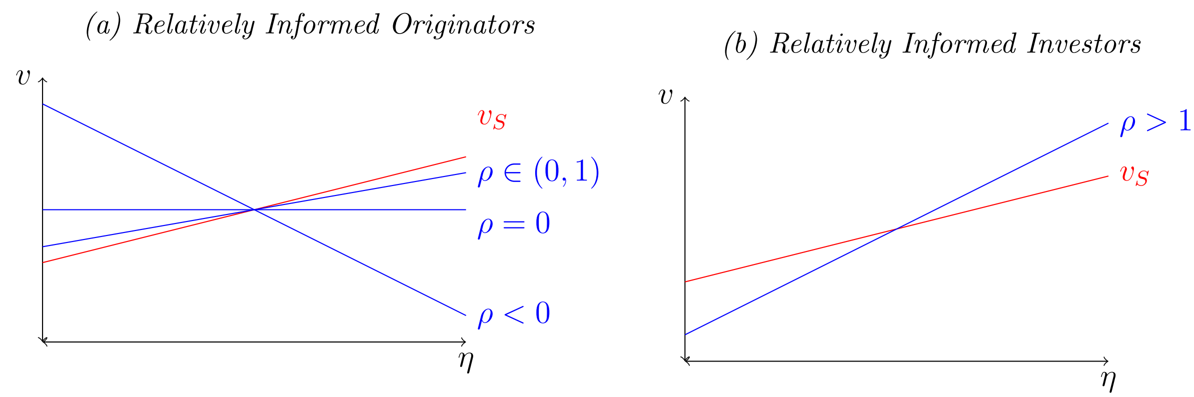



*(b) Channel Differences w/ Decision*



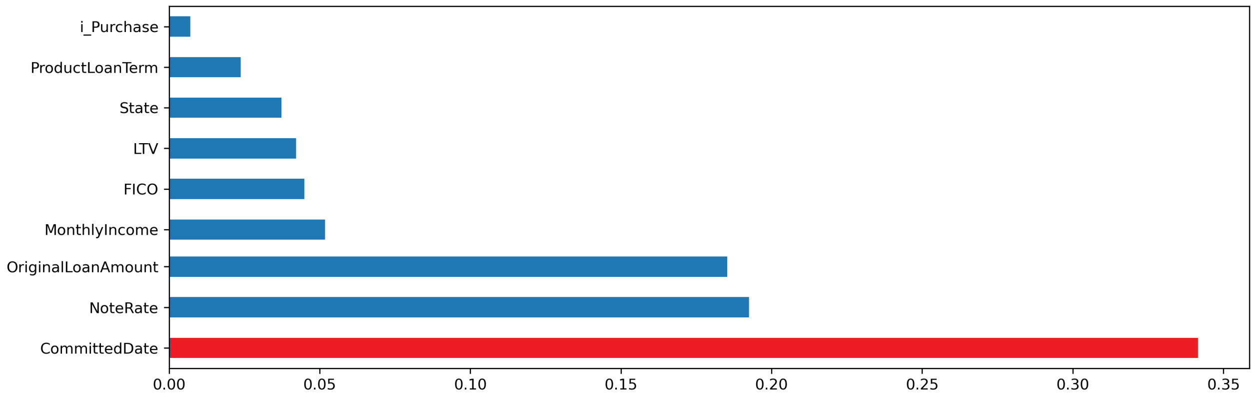

Random survival forests trained on loan and borrower characteristics and averaged over loans within group (based on sale decision). Models are trained separately without known sale decision (Figure 4a) and with known sale decision (Figure 4b). Without realized sale decision, no difference in survival is predicted between the two groups, suggesting loans are comparable in terms of the contribution of observed characteristics to expected loan performance. Knowing the realized sale decision predicts 3% difference in survival probability three years after origination and contributes to improved model fit.

*Figure 5: Expected Value of the Keep (Blue) and Sell (Red) Options*




Figures depict the value of selling to the GSEs (blue) and selling to third party investors (red) as a function of the bid residual $\eta$, conditional on $\omega$. Figure 5a shows three possible cases where the private signal is relatively more informative than investor bids. Depending on the values of the covariance terms, the value of selling to the GSEs may be increasing, constant, or decreasing in the bid residual. In all cases, originators' optimal strategy is to sell when received bids are idiosyncratically low. Figure 5b shows the counterintuitive case where bids contain are more informative than originators' private signals. In this case, the value of selling to the GSEs increases in $\eta$ with a slope greater than one, and originators' optimal strategy is to sell when received bids are idiosyncratically high.

*Figure 6: Importance Plots - GSE Cash Window Prices*




The figure depicts the importance plots from a random forest model trained on GSE cash window prices. Intuitively, this shows the share of explained variation that can be attributed to each covariate. By far, the most important covariate for explaining cash window prices is the auction date (or *committed date*). This captures ex-post realized trends in GSE pricing over time, and can be used to impute cash window prices where they are not observed.

*Figure 7: Distribution of Realized Three-Year Servicing Revenues by Cohort*

*(a) Raw Cash Flows* *(b) Normalized Cash Flows*

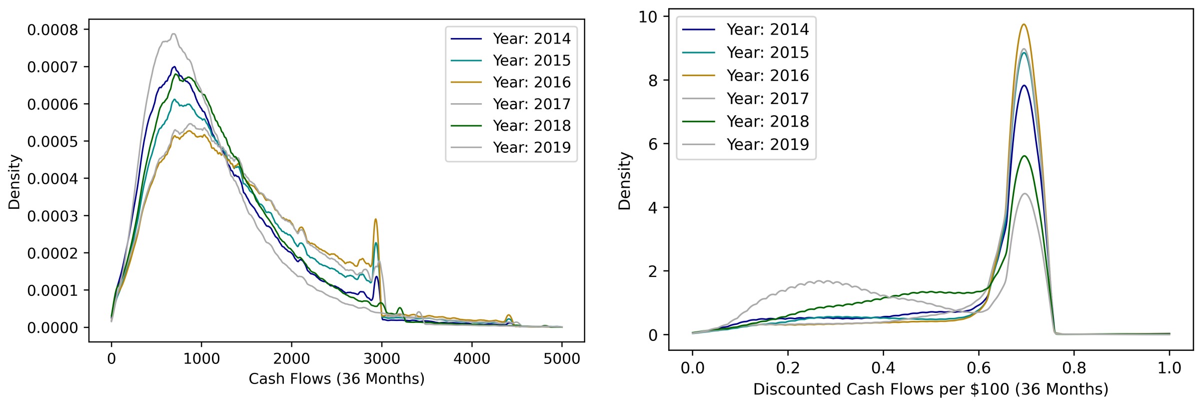



Figure 7a shows the distribution of realized cash flows, while Figure 7b shows the distribution of normalized cash flows. Realized cash flows scale with the length of loan survival and the size of the loan. Normalized cash flows are expressed in dollars per $100 of loan volume and thus reflect only the duration of loan survival and the timing of any prepayment. The mass point at $3000 in raw cash flows reflects bunching around the conforming loan limit, which was slightly over $400k during this time–a loan at this conforming loan limit pays servicing fees close to $1000 per year in early years.

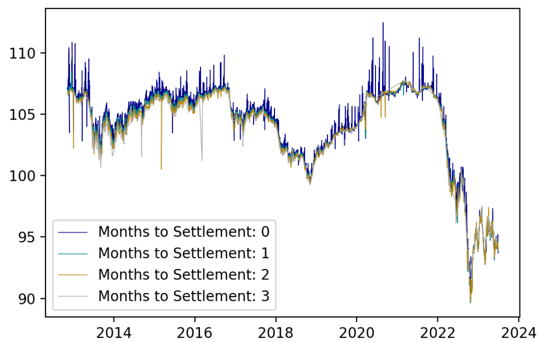
*Table 4: Loan Characteristics: Sample Comparison*

| | (1) Historical Sample | | (2) Full Auction Sample | | (3) Model Sample | |
|---|---|---|---|---|---|---|
| | mean | sd | mean | sd | mean | sd |
| Original Loan Amount | 234145 | 114225 | 254868 | 114360 | 256590 | 117751 |
| Interest Rate | 4.18 | 0.41 | 4.41 | 0.55 | 4.43 | 0.57 |
| Loan-to-Value Ratio | 80.1 | 15.6 | 80.7 | 15.0 | 80.1 | 15.4 |
| Credit Score | 746.0 | 60.5 | 746.5 | 63.5 | 745.9 | 44.8 |
| Debt-to-Income Ratio | 35.7 | 37.4 | 36.2 | 9.35 | 36.4 | 9.07 |
| Number of Borrowers | 1.49 | 0.50 | 1.45 | 0.50 | 1.44 | 0.50 |
| 1(First-Time Homebuyer) | 0.28 | 0.45 | 0.31 | 0.46 | 0.30 | 0.46 |
| 1(Home Purchase) | 0.28 | 0.45 | 0.20 | 0.40 | 0.21 | 0.41 |
| 1(Second Home) | 0.045 | 0.21 | 0.044 | 0.21 | 0.045 | 0.21 |
| Years | 2013-2017 | | 2018-2020 | | 2018-2020 | |
| Observations | 5497024 | | 210531 | | 108524 | |

Borrower characteristics are comparable between the historical sample and the model sample, with some loan characteristic differences, namely in loan size, interest rate, and home purchase fraction. Relative to the historical sample, the model sample has a slightly higher average loan size, owing to the fact that home prices increase over time. Loan interest rates are also marginally higher in the model sample, reflecting the brief five-year highs reached in late-2018. These higher interest rates led to lower refinancing rates during this time, meaning the model sample has a higher fraction of home purchases (and so also first-time homebuyers).

*Figure 8: To-Be-Announced Prices*

*(a) TBA Price by Forward Months*




*(b) TBA Price by Coupon*

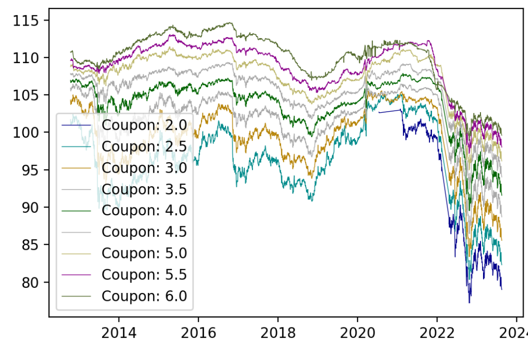



Figure 8a shows the time trend of daily average To-Be-Announced prices for a single coupon (4.0%) for securities with 0-3 remaining months to settlement. At zero months to settlement, TBA prices exhibit high variance, while TBA prices with 1-2 months until settlement show low variance. In our bid regressions, we control for TBA prices with 1-2 months until settlement both because of the low variance in average daily prices, and because loans currently being pooled into securities are not eligible for immediate sale on the TBA market, making 1-2 months a more salient time window to capture resale value. Figure 8b shows the time trend of daily average TBA prices for coupons between 2.5 and 6.0. Over time, the price difference between coupons changes considerably, compressing considerably between early 2018 and early 2020. Thus, controlling for TBA prices is important for capturing baseline price differences between loans, especially in early periods.

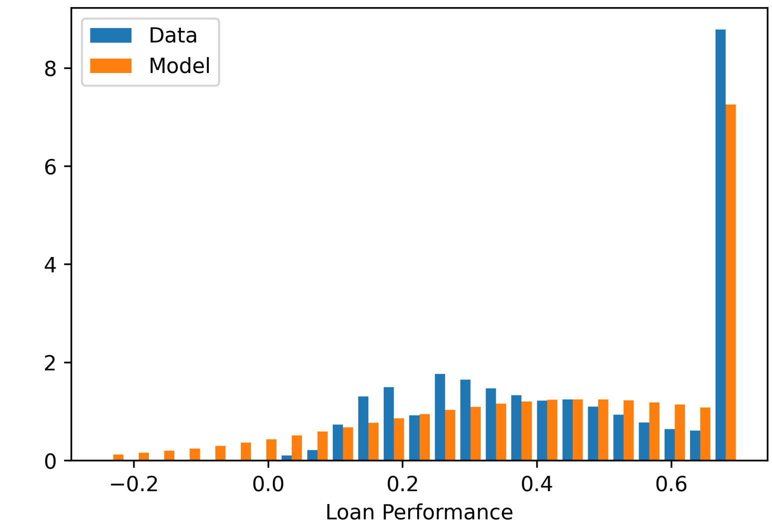
*Table 5: Model Estimates*

| | Performance Model | | Max Bid Model | |
|---|---|---|---|---|
| | Coefficient | Standard Error | Coefficient | Standard Error |
| *Borrower Characteristics* | | | | |
| Income | — | — | -0.0060*** | (0.00087) |
| Credit Score | -0.00022*** | (0.0000026) | 0.0025*** | (0.00007) |
| *Loan Characteristics* | | | | |
| Loan Size | -0.038*** | (0.00016) | -0.43*** | (0.0037) |
| Loan-to-Value Ratio | 0.0018*** | (0.000012) | -0.0025* | (0.00022) |
| *Seller Characteristics* | | | | |
| Seller Network Size | — | — | 0.028*** | (0.00036) |
| *Market Controls* | | | | |
| Yield Curve | -0.037*** | (0.00035) | | () |
| TBA Price | 0.0011*** | (0.00012) | 1.07*** | (0.0041) |
| *Fixed Effects* | State, FICO-by-LTV, Loan Size Bins | | | |
| *Model Characteristics* | | | | |
| Type | Tobit | | OLS | |
| Observations | 5485296 | | 108524 | |
| $R^2$ | 0.142 | | 0.761 | |

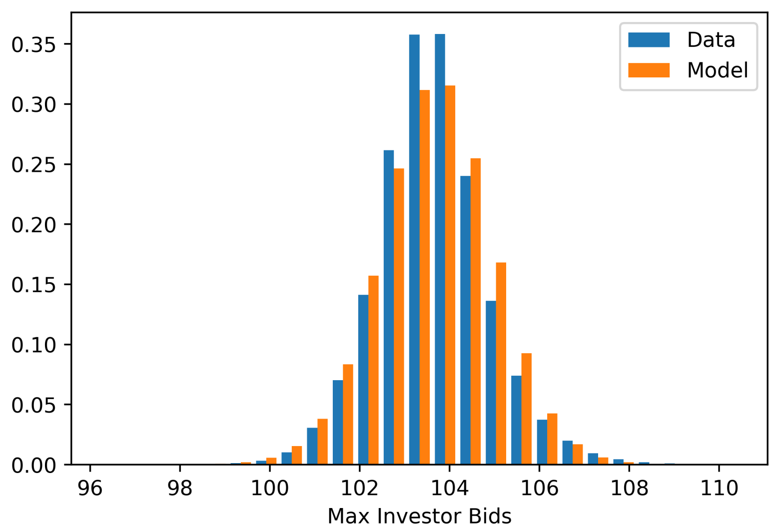
| | | | | | | |
|---|---|---|---|---|---|---|
| *Covariance Terms* | | | | | | |
| $\varepsilon$ | 0.098*** | (0.00011) | 0.212*** | (0.00073) | 0.046*** | (0.00064) |
| $\omega$ | 0.212*** | (0.00073) | 1.000 | — | 0.480*** | (0.00084) |
| $\eta$ | 0.046*** | (0.00064) | 0.480*** | (0.00084) | 0.452*** | (0.002) |
| *Servicing Costs* | | | | | | |
| $c_O$ | 0.539*** | (0.00069) | | | | |

Estimates for the model in Section 5. Linear coefficients are within-bin where applicable. For instance, FICO is binned in 20 point increments and the shown effect is for marginal increases in FICO *within* those bins. Fixed effects and additional loan and borrower characteristic estimates (e.g., DTI, occupancy type, loan purpose) are suppressed for ease of interpretation. Yield curve variable is 10-year maturity to 1-year maturity price ratio. TBA prices are selected based on the median observed coupon for each loan's note rate, which is between 0.50 – 0.875 below the note rate.

*Figure 9: Model Fit*



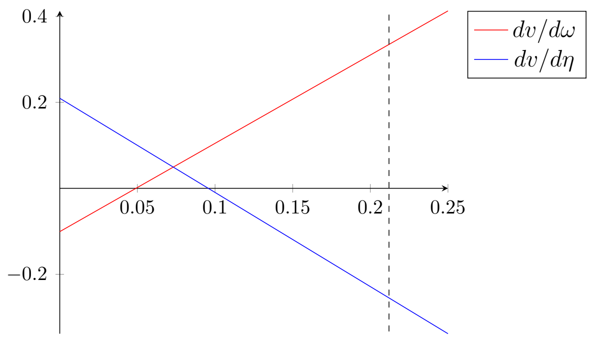

*(a) Model Fit - Loan Performance*



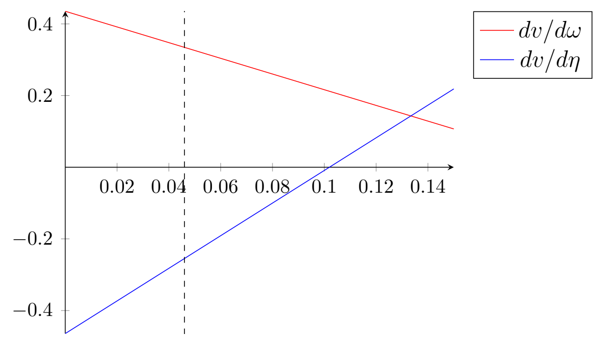

*(b) Model Fit - Maximum Bid*

Subfigure 9a shows the model fit for performance. Because the performance residual is unbounded with a non-trivial standard deviation, the model predicts a small (but still noticeable) fraction of loans with negative loan performance. Furthermore, due to ex-post shocks in economic conditions, realized loan performance can depend considerably on the time of origination. Thus, realized loan performance has occasional spikes, while modeled loan performance is smooth. This is unconcerning if shocks driving performance are unanticipated at the time of origination. Arguably, modeled loan performance *better* captures originators expectations about loan performance. Subfigure 9b shows the model performance for maximum bids. Modeled bids closely align with observed bids.

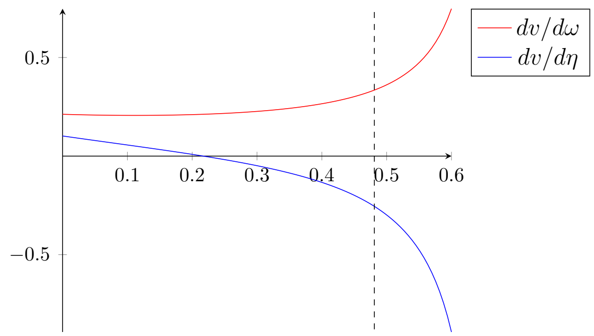
*Figure 10: Marginal Effect of Information Structure on Valuations*




*(a) Derivatives in $\sigma_{MS}$*




*(b) Derivatives in $\sigma_{MP}$*




*(c) Derivative in $\sigma_{SP}$*

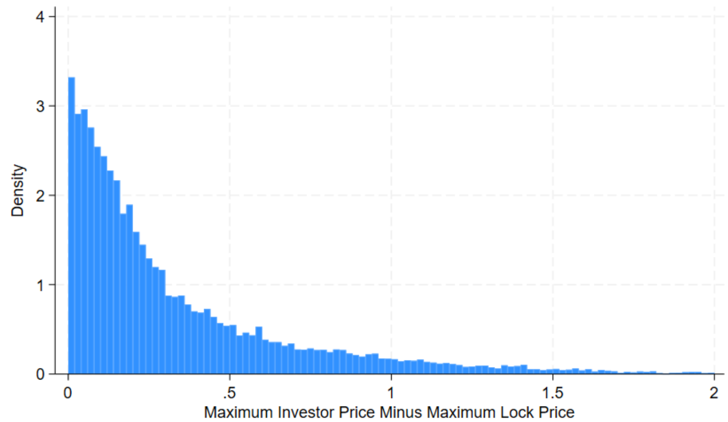
The graphs depict the marginal effects of the three information structure parameters on sellers' valuations. The plots vary one parameter, while holding fixed the parameter estimates for the other two. At current estimates, higher bids are predictive of lower performance, while higher signals are predictive of higher performance. Other things being equal, higher bids would be predictive of higher performance if any of the following were true: (a) sellers' signals were only poorly correlated with residual performance, or (b) bidders' residual bids were very strongly correlated with residual performance, or (c) sellers' signals were only very weakly correlated with residual bids. However, all of these would imply less informative signals.

*Table 6: Benchmark Model Estimates*

| | Performance Model | | Max Bid Model | |
|---|---|---|---|---|
| | Coefficient | Standard Error | Coefficient | Standard Error |
| *Borrower Characteristics* | | | | |
| Income | — | — | -0.013*** | (0.00087) |
| Credit Score | -0.00022*** | (0.0000026) | 0.0013*** | (0.00028) |
| *Loan Characteristics* | | | | |
| Loan Size | -0.038*** | (0.00016) | -0.36*** | (0.0040) |
| Loan-to-Value Ratio | 0.0018*** | (0.000012) | -0.016* | (0.00046) |
| *Seller Characteristics* | | | | |
| Seller Network Size | — | — | 0.044*** | (0.0011) |
| *Market Controls* | | | | |
| Yield Curve | -0.037*** | (0.00035) | | () |
| TBA Price | 0.0011*** | (0.00012) | 0.90*** | (0.0043) |
| *Fixed Effects* | State, FICO-by-LTV | | | |
| *Model Characteristics* | | | | |
| Type | Tobit | | OLS | |
| Observations | 5,485,296 | | 52,169 | |
| $R^2$ | 0.142 | | 0.704 | |
| *Covariance Terms* | | | | |
| $\varepsilon$ | 0.098 | (0.00011) | 0.108 | (0.0007) |
| $\eta$ | 0.108 | (0.0007) | 0.523* | (0.0032) |

Estimates are obtained and interpreted as in Table 5.

*Figure 11: Distribution of Max Investor Price Minus Max Lock Price*




For ease of visualization, the figure shows the distribution only for loans where the maximum investor price was *not* a lock price. Approximately 15% of loans auctioned have a maximum bid that is a lock price. About 80% of auctioned loans have a spread between the maximum investor price and maximum lock price between \$0-\$0.50 (per \$100 loan volume) Only about 5% of auctioned loans have a spread of greater than \$1.

*Table 7: Counterfactual Outcomes Summary*

| | | | Outcome | | | |
|---|---|---|---|---|---|---|
| Lock Outcome | Auction Outcome | Share | Average Residual | Residual Cash Flows | Bid Minus Lock (Norm) | Bid Minus Lock (Cash) |
| Securitized | Securitized | 41% | 0.032 | \$88±40 | 0.140 | \$385±175 |
| Securitized | Sold | 23% | -0.036 | -\$99±45 | 0.613 | \$1675±760 |
| Sold | Securitized | 4% | 0.100 | \$275±125 | 0.015 | \$41±19 |
| Sold | Sold | 32% | -0.027 | -\$75±34 | 0.177 | \$485±220 |

Results based on 1000 simulation draws of model residuals from the estimated joint distribution. Average residual is the average of the performance model residual within each group, which does not scale with loan size. Residual cash flows are computed by multiplying performance residuals by loan size for each loan. Bid minus lock averages observed differences between the maximum lock and the maximum bid for each group, and dollar value flows are computed by multiplying the normalized (per \$100 loan volume) amounts by the loan size.

*Table 8: Counterfactual Model Estimates: Information Structure*

| | | | | | | |
|---|---|---|---|---|---|---|
| *Covariance Terms* | | | | | | |
| $\varepsilon$ | 0.098 | (0.00011) | 0.230 | (0.00059) | 0.051 | (0.0022) |
| $\omega$ | 0.230 | (0.00059) | 1.000 | — | 0.401 | (0.0054) |
| $\eta$ | 0.051 | (0.0022) | 0.401 | (0.0054) | 0.452* | (0.002) |
| *Servicing Costs* | | | | | | |
| $c_1$ | 0.650 | (0.0022) | | | | |
| $c_2$ | 0.575 | (0.0027) | | | | |
| $c_3$ | 0.597 | (0.0049) | | | | |
| $c_4$ | 0.565 | (0.015) | | | | |
| $c_5$ | 0.500 | (0.0015) | | | | |
| $c_6$ | 0.575 | (0.0092) | | | | |
| $c_7$ | 0.609 | (0.038) | | | | |
| $c_8$ | 0.528 | (0.0089) | | | | |
| $c_9$ | 0.632 | (0.0018) | | | | |
| $c_{10}$ | 0.551 | (0.0037) | | | | |
| $c_{Small}$ | 0.481 | (0.0066) | | | | |

Estimates for the model in Section 7.2. Loan and borrower characteristic and fixed-effects estimates are identical to the baseline model and are suppressed for ease of interpretation. Seller-specific securitization costs are ordered by the time of auction platform entry for anonymity.

*Table 9: Counterfactual 2 Outcomes*

**Seller Outcomes**

| Seller | Securitization Cost | Counterfactual Keep Value | Counterfactual Sale Value | Realized Choice Value | Outcome | Welfare Change |
|---|---|---|---|---|---|---|
| *Individual Seller Outcomes* | | | | | | |
| 1 | 0.650 | 103.24 | 103.25 | 103.57 | Sell | -0.32 |
| 2 | 0.575 | 102.96 | 103.80 | 103.66 | Sell | 0.14 |
| 3 | 0.597 | 103.49 | 103.91 | 104.05 | Sell | -0.14 |
| 4 | 0.565 | 102.81 | 103.77 | 103.56 | Sell | 0.21 |
| 5 | 0.500 | 103.81 | 103.90 | 104.21 | Sell | -0.30 |
| 6 | 0.575 | 103.63 | 104.40 | 104.36 | Sell | 0.04 |
| 7 | 0.609 | 104.06 | 103.99 | 104.33 | Keep | -0.27 |
| 8 | 0.528 | 105.12 | 104.45 | 105.37 | Keep | -0.25 |
| 9 | 0.632 | 103.39 | 103.76 | 103.91 | Sell | -0.15 |
| 10 | 0.551 | 103.03 | 103.48 | 103.61 | Sell | -0.14 |
| *Small Seller Outcomes* | | | | | | |
| Agg. | .481 | 102.95 | 103.40 | 103.57 | 84% Sell | -0.11 |

# B Additional Market Details

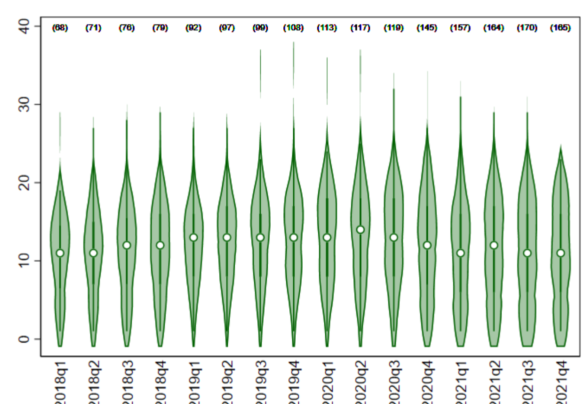
## B.1 Wholesale Connections in Optimal Blue Auctions

The advent of loan trading platforms has the potential to increase competition among wholesalers. By connecting originators with wholesalers in a centralized location, these platforms can encourage interactions with a greater number of counterparties, and bilateral relationships may become less important.

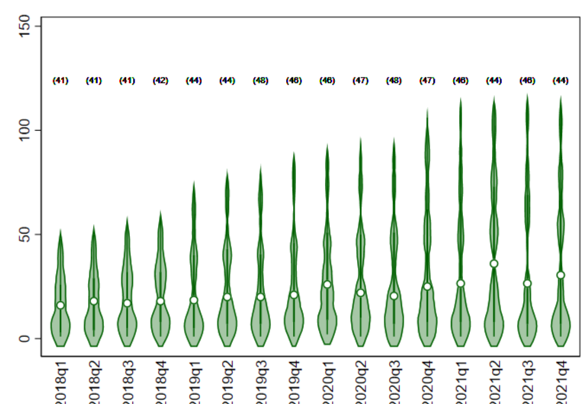
This can be seen in the specific case of Optimal Blue auctions. Figure 12 plots the distribution of network size over time for large bidders and sellers on the platform. The median seller interacts with between 10-12 unique bidders over the course of a given quarter. Further, sellers' networks remain relatively constant across time–sellers may add or drop a single bidder from their network, but large changes in network size and composition are rare. When relationships do change, this is usually the result of adding or subtracting smaller bidders. Most sellers interact around seven of the ten largest bidders on the platform, and these relationships rarely change. A network consisting of 10-12 unique bidders is larger than the median number of wholesale relationships for originators as a whole. However, it is difficult to know whether this difference is attributable to the platform or whether originators with more relationships are disproportionately likely to use the platform.

*Figure 12: Auction Network Size Over Time*

*(a) Seller Network Size*




*(b) Bidder Network Size*

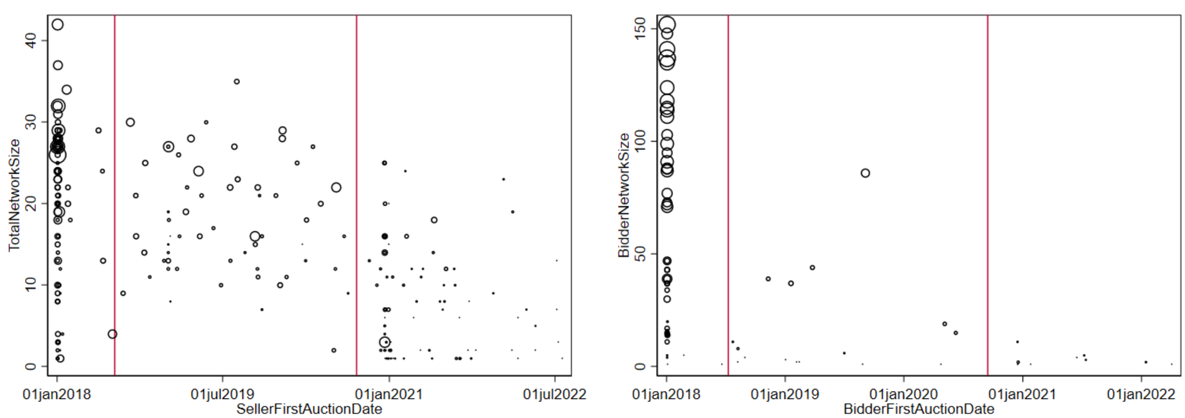



Note: The above distributions are depicted only for sellers holding auctions with more than 250 auctions involving three or more bidders and bidders participating in more than 250 unique auctions. The number of qualifying sellers/bidders is shown at the top of each figure. Networks are imputed from observed interactions between bidders and sellers. A bidder and seller are considered to have a network connection if the bidder submits at least one bid for a loan auctioned by the seller within a given quarter.

That worry notwithstanding, we present qualitative evidence for the network expansion effects of the auction platform by looking at sellers who join the platform in the middle of our data sample. As background, the loan auction platform was launched in 2017 by the company ResiTrader, and it grew steadily over the course of that year. By the beginning of 2018, the platform was used regularly by nearly 100 unique sellers and 50 unique bidders. In July 2018, ResiTrader was acquired by Optimal Blue, the provider of the largest rate locking platform. Then in September of 2020, Optimal Blue was acquired by Black Knight. Figure 13 shows the dates of platform entry for bidders and sellers relative to the two acquisition events. Both acquisitions were followed by increased adoption of the platform by numerous sellers and even some bidders, with the total number of sellers growing to over 200 and the total number of bidders growing to about 80. For sellers who entered the platform after mid-2018, most experienced an increase in the number of wholesalers who regularly purchase their loans.

*Figure 13: Use of Auctions Over Time*

*(a) Platform Entry by Sellers* *(b) Platform Entry by Investors*




Note: Observations weighted by number of auctions participated in between January 2018 and July 2022. Vertical red lines denote major events for the platform: (1) July 10, 2018: Acquisition of the ResiTrader loan trading platform by Optimal Blue; (2) September 15, 2020: Acquisition of Optimal Blue by Black Knight Financial.

## B.2 Cash Window Details

### B.2.1 Measurement and GSE Prices

This appendix discusses the measurement concerns with the GSE cash window prices reported in the auction data. Figure 14 shows the distribution of the difference between the maximum investor bid and the maximum GSE cash window price as computed using the raw data. The distribution is bimodal, and the bimodality persists even when conditioning on the outcome of the auction.

*Figure 14: Distribution of Investor-GSE Price Spread*

*(a) Full Distribution* *(b) Distribution by Realized Channel*

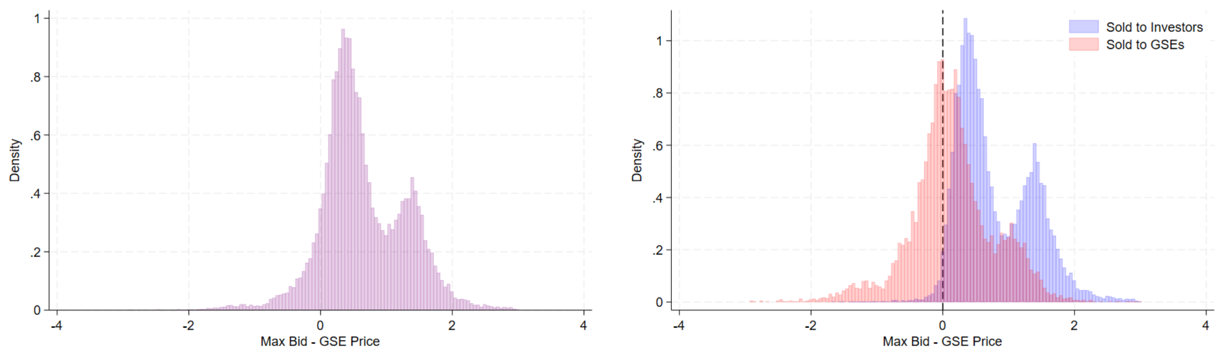
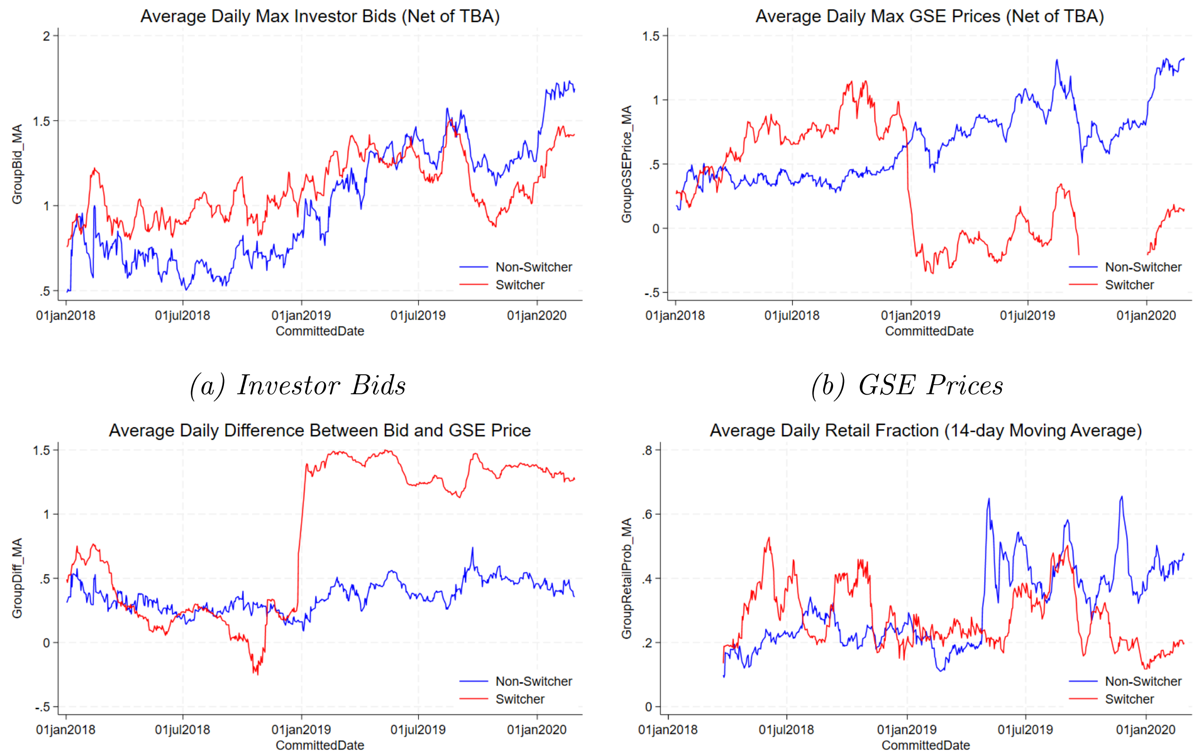



Figure 14a shows the distribution of the spread between the maximum investor bid and the maximum GSE price for all auctioned loans originated before March 2020. Figure 14b shows the distribution based on whether the auctioned loan was ultimately sold to the GSEs (red) or to third-party investors (blue). Almost all (98%) of loans sold to third-party investors have a positive bid spread, indicating that originators place positive value on servicing payments even after accounting for costs associated with servicing and securitization which are not incorporated into the GSE price offers.

To better understand the source of this bimodality, we examine the time series of average investor bids and GSE prices. As shown in Figure 15, the average investor bids move continuously over time, while the average GSE prices drops precipitously on January 1, 2019 for a subset of sellers (deemed "switchers" after this abrupt switch in GSE prices). This drop accounts fully for the observed bimodality, as it causes a discrete jump in bid spreads for these sellers. A few lines of evidence suggest that this apparent change in GSE prices is artefactual rather than genuine. First, the auction platform distinguishes between "net" and "gross" prices, allowing sellers to attach weights to particular investors or GSEs to account for the implicit cost of transacting with them. The discrete jump in prices at a focal date like January 1 could reflect a difference in the type of reported price. Furthermore, the use of these weights is optional, which would explain why it affects only a subset of sellers. Consistent with this explanation is the fact that after the change, the bid spread and GSE price averages move in rough parallel after the change. Finally, around the date of the price drop, the fraction of "retail loans" (i.e., loans securitized directly with the GSEs, rather than sold to investors), remains steady for switchers. This is extremely unlikely given the magnitude of the GSE price decrease (about 2.5 years of servicing revenues), which if genuine should lead to a sharp increase in the fraction of loans sold to third party investors. As additional confirmation, note that the drop in reported GSE prices is close in magnitude to the estimated securitization costs from Section 6.

*Figure 15: Time Series of Investor Bids and GSE Prices*




*(a) Investor Bids*

*(b) GSE Prices*

*(c) Bid − GSE Price*

*(d) Retail Fraction*

The figure depicts the 14-day moving average time trends of bids, GSE prices, and retail fractions for two types of sellers–switchers and non-switchers. Switchers are sellers for whom observed GSE prices see a sharp decline in the data on exactly January 1, 2019. We see a level shift in the GSE price reported starting on January 1, followed by GSE prices that continue to move in parallel with a consistent offset. Despite what appears to be a price decrease on the order of \$0.60 (the equivalent of nearly three full years of servicing revenues), the fraction of retail loans (i.e., loans delivered by sellers to the GSEs) does not change appreciably at this time. The best explanation of this fact is that the reported values of the GSE price are not genuine, but reflect changes only in how those prices are reported. Sellers labeled "switchers" account for only about 1/4 of loans sold during this time.

### B.2.2 Cash Window Price Bins

This appendix briefly expands on the choice of loan covariate bins in the bid and cash window price models.

First, we use FICO score by LTV bins that align with the loan level price adjustments charged by Fannie Mae and Freddie Mac. An example of Fannie Mae's pricing matrix is given in Table 10. The placement of these bins was unchanged during the sample period, and would reflect additional buy-ups or buy-downs by the GSEs.

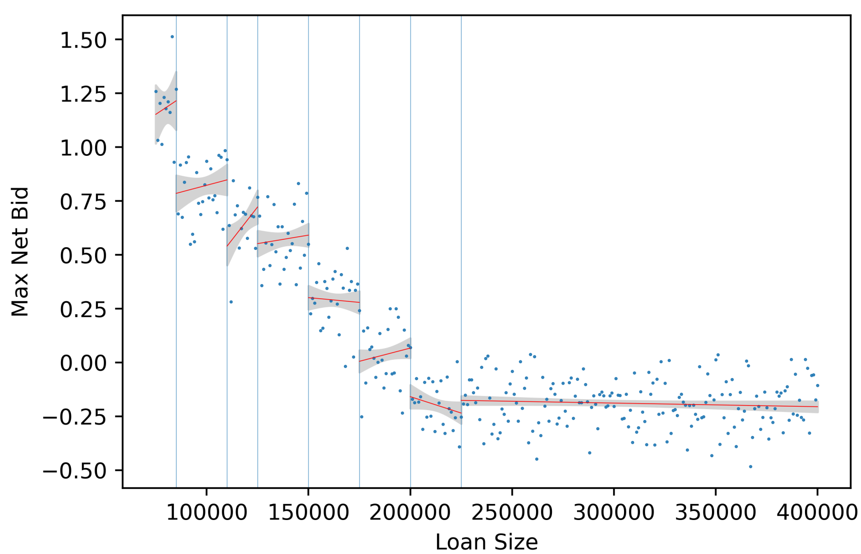
*Table 10: Fannie Mae Loan-Level Price Adjustment Matrix (FICO-by-LTV Scores)*

| | LTV Ratio | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| FICO Score | <60% | 60 – 70% | 70 – 75% | 75 – 80% | 80 – 85% | 85 – 90% | 90 – 95% | 95 – 97% | >97% |
| ≥740 | 0.000% | 0.250% | 0.250% | 0.500% | 0.250% | 0.250% | 0.250% | 0.750% | 0.750% |
| 720 – 739 | 0.000% | 0.250% | 0.500% | 0.750% | 0.500% | 0.500% | 0.500% | 1.000% | 1.000% |
| 700 – 719 | 0.000% | 0.500% | 1.000% | 1.250% | 1.000% | 1.000% | 1.000% | 1.500% | 1.500% |
| 680 – 699 | 0.000% | 0.500% | 1.250% | 1.750% | 1.500% | 1.250% | 1.250% | 1.500% | 1.500% |
| 660 – 679 | 0.000% | 1.000% | 2.250% | 2.750% | 2.750% | 2.250% | 2.250% | 2.250% | 2.250% |
| 640 – 659 | 0.500% | 1.250% | 2.750% | 3.000% | 3.250% | 2.750% | 2.750% | 2.750% | 2.750% |
| 620 – 639 | 0.500% | 1.500% | 3.000% | 3.000% | 3.250% | 3.250% | 3.250% | 3.500% | 3.500% |
| <620 | 0.500% | 1.500% | 3.000% | 3.000% | 3.250% | 3.250% | 3.250% | 3.750% | 3.750% |

Source: Before May 2023

Finally, we also incorporate loan size bins of width $\$25,000$ for loans below $\$300,000$. This is consistent motivated by documentation from the GSEs themselves.[26] Figure 16 plots the Fannie Mae cash window price averages (net of TBA) as a function of loan size. The figure strongly suggests steep discontinuities in cash window prices at conventional loan size boundaries.

*Figure 16: FNMA Cash Window Prices (Net of 2mo TBA)*




Accounting for both types of binned loan covariates considerably improves model fit, not only for GSE prices but also for investor bids.

[26]See, for example: https://sf.freddiemac.com/working-with-us/selling-delivery/delivery-options-pricing/cash-payups

## B.3 Channel Comparison Details

This section discusses differences between loan and survival characteristics between the retail and wholesale markets as a whole, in part to motivate the choice of training data for the cash flow model of Section 5.3.2.

Table 11 shows differences in loan characteristics across channels. Brokered loans differ considerably from retail loans along multiple dimensions, so we bracket them for now and compare correspondent/wholesale loans to retail loans. In terms of financial characteristics, correspondent loans are slightly 'worse' than retail loans. Borrowers have larger loan sizes, higher interest rates, lower credit scores, lower incomes, higher LTV ratios (thus lower down-payment percentages), and higher DTI ratios (thus higher relative loan payment burders). Although total loan costs are comparable between channels, up-front origination charges are higher for retail loans, and retail loans also have more purchased discount points.

*Table 11: Loan Characteristics by Lending Channel*

| | Brokered Loans | | Correspondent Loans | | Retail Loans | |
|---|---|---|---|---|---|---|
| | mean | sd | mean | sd | mean | sd |
| Loan Amount | 331997.8 | 146374.4 | 298659.9 | 132745.7 | 294373.9 | 140194.5 |
| Interest Rate | 3.390 | 0.716 | 3.755 | 0.835 | 3.415 | 0.697 |
| Credit Score | 752.8 | 42.48 | 748.6 | 43.98 | 751.0 | 44.11 |
| Annual Income | 111.6 | 76.04 | 107.6 | 73.76 | 108.1 | 74.58 |
| LTV Ratio | 73.55 | 17.15 | 78.03 | 16.03 | 75.27 | 16.62 |
| DTI Ratio | 35.42 | 9.525 | 35.56 | 9.342 | 34.55 | 9.549 |
| Borrower Age | 45.03 | 13.65 | 43.80 | 13.72 | 45.20 | 14.19 |
| 1(Black Borrower) | 0.0412 | 0.199 | 0.0497 | 0.217 | 0.0474 | 0.212 |
| Total Loan Costs | 1.684 | 1.031 | 1.677 | 1.051 | 1.678 | 1.055 |
| Origination Charges | 0.884 | 0.821 | 0.773 | 0.696 | 0.803 | 0.732 |
| Discount Points | -0.237 | 0.900 | 0.0915 | 0.643 | 0.170 | 0.715 |
| 1(12-month Survival) | 0.872 | 0.334 | 0.879 | 0.326 | 0.907 | 0.290 |
| 1(18-month Survival) | 0.766 | 0.424 | 0.755 | 0.430 | 0.815 | 0.388 |
| 1(24-month Survival) | 0.645 | 0.478 | 0.618 | 0.486 | 0.710 | 0.454 |
| 1(30-month Survival) | 0.468 | 0.499 | 0.463 | 0.499 | 0.549 | 0.498 |
| 1(36-month Survival) | 0.291 | 0.454 | 0.337 | 0.473 | 0.350 | 0.477 |
| Observations | 1472931 | | 2785406 | | 4801440 | |

Figure 12 compares loan performance across channels. As with loan characteristics, loans in the broker channel differ systematically from loans in the correspondent and retail channels in terms of loan performance. Comparing correspondent and retail channels, correspondent loans are more likely than retail loans to survive in the first $12-36$ months. These differences in loan survival are relatively large. In the first 12 months after origination, roughly 12% of correspondent loans have prepaid, compared to only 9% of retail loans. This gap grows to close to 10 percentage points by 24 months before tapering off by 36 months (this, however, reflects the fact that very few loans in our sample period survive for a full 36 months).

*Table 12: Survival Differences Across Channels (Sample Comparison)*

| | | Channel | | | | | |
|---|---|---|---|---|---|---|---|
| | | Brokered | | Correspondent | | Retail | |
| | | mean | sd | mean | sd | mean | sd |
| Full MBS Sample | 1(12-month Survival) | 0.865 | 0.342 | 0.889 | 0.314 | 0.906 | 0.292 |
| | 1(18-month Survival) | 0.755 | 0.430 | 0.777 | 0.416 | 0.808 | 0.394 |
| | 1(24-month Survival) | 0.634 | 0.482 | 0.654 | 0.476 | 0.700 | 0.458 |
| | 1(30-month Survival) | 0.456 | 0.498 | 0.499 | 0.500 | 0.549 | 0.498 |
| | 1(36-month Survival) | 0.299 | 0.458 | 0.361 | 0.480 | 0.386 | 0.487 |
| | Discounted UPB 12 | 11.14 | 1.685 | 11.22 | 1.611 | 11.29 | 1.521 |
| | Discounted UPB 18 | 15.59 | 3.659 | 15.79 | 3.458 | 16.03 | 3.238 |
| | Discounted UPB 24 | 18.97 | 6.073 | 19.34 | 5.742 | 19.92 | 5.380 |
| | Discounted UPB 30 | 20.44 | 8.693 | 21.52 | 8.203 | 22.44 | 7.873 |
| | Discounted UPB 36 | 21.14 | 10.37 | 22.73 | 10.07 | 23.73 | 9.861 |
| | Observations | 1880028 | | 2660004 | | 7434179 | |
| Matched HMDA-MBS Sample | 1(12-month Survival) | 0.862 | 0.345 | 0.886 | 0.318 | 0.903 | 0.295 |
| | 1(18-month Survival) | 0.753 | 0.431 | 0.772 | 0.420 | 0.805 | 0.396 |
| | 1(24-month Survival) | 0.630 | 0.483 | 0.643 | 0.479 | 0.695 | 0.460 |
| | 1(30-month Survival) | 0.447 | 0.497 | 0.485 | 0.500 | 0.540 | 0.498 |
| | 1(36-month Survival) | 0.284 | 0.451 | 0.346 | 0.476 | 0.370 | 0.483 |
| | Discounted UPB 12 | 11.13 | 1.696 | 11.21 | 1.617 | 11.28 | 1.532 |
| | Discounted UPB 18 | 15.58 | 3.668 | 15.77 | 3.471 | 16.01 | 3.254 |
| | Discounted UPB 24 | 18.91 | 6.097 | 19.24 | 5.776 | 19.86 | 5.413 |
| | Discounted UPB 30 | 20.25 | 8.725 | 21.31 | 8.232 | 22.27 | 7.928 |
| | Discounted UPB 36 | 20.63 | 10.35 | 22.23 | 10.07 | 23.29 | 9.899 |
| | Observations | 1713463 | | 2117319 | | 6766847 | |
| Mixed-Channel Sample | 1(12-month Survival) | 0.872 | 0.334 | 0.879 | 0.326 | 0.907 | 0.290 |
| | 1(18-month Survival) | 0.766 | 0.424 | 0.755 | 0.430 | 0.815 | 0.388 |
| | 1(24-month Survival) | 0.645 | 0.478 | 0.618 | 0.486 | 0.710 | 0.454 |
| | 1(30-month Survival) | 0.468 | 0.499 | 0.463 | 0.499 | 0.549 | 0.498 |
| | 1(36-month Survival) | 0.291 | 0.454 | 0.337 | 0.473 | 0.350 | 0.477 |
| | Discounted UPB 12 | 11.18 | 1.630 | 11.18 | 1.650 | 11.30 | 1.500 |
| | Discounted UPB 18 | 15.72 | 3.541 | 15.67 | 3.544 | 16.10 | 3.173 |
| | Discounted UPB 24 | 19.19 | 5.927 | 19.00 | 5.872 | 20.05 | 5.324 |
| | Discounted UPB 30 | 20.78 | 8.579 | 21.01 | 8.261 | 22.45 | 7.939 |
| | Discounted UPB 36 | 21.10 | 10.27 | 22.02 | 10.06 | 23.09 | 9.897 |
| | Observations | 1261113 | | 1417013 | | 3697099 | |

# C Robustness Exercises

## C.1 Bivariate Probit Robustness

Section 4.2 provided descriptive motivation for the existence and importance of seller private information using bivariate probit model and testing for positive correlation between loan performance (survival or non-default) and the decision of sellers to securitize the loan directly. Initial results provided some evidence of private information. Table 13 shows that the same pattern appears under a variety of model specifications. We control for privately-known loan characteristics (discussed further in Appendix C.3, GSE cash window prices, and spreads between investor bids and GSE prices.

*Table 13: Bivariate Probits - Alternative Models*

| | Model | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| $\rho_{12}$ | 0.0109 | 0.00548 | 0.0228* | 0.0131 | 0.00546 | 0.0153 | 0.0174 |
| | (0.00620) | (0.00670) | (0.00909) | (0.00971) | (0.00670) | (0.00993) | (0.0105) |
| $\rho_{18}$ | 0.0128* | 0.0107 | 0.0216** | 0.0179* | 0.0107 | 0.0227** | 0.0256** |
| | (0.00544) | (0.00586) | (0.00775) | (0.00825) | (0.00586) | (0.00845) | (0.00889) |
| $\rho_{24}$ | 0.0174** | 0.0175** | 0.0168* | 0.0182* | 0.0176** | 0.0283*** | 0.0303*** |
| | (0.00534) | (0.00575) | (0.00750) | (0.00799) | (0.00575) | (0.00819) | (0.00861) |
| $\rho_{30}$ | 0.0185*** | 0.0194*** | 0.0195* | 0.0214** | 0.0194*** | 0.0339*** | 0.0310*** |
| | (0.00541) | (0.00583) | (0.00759) | (0.00808) | (0.00583) | (0.00831) | (0.00874) |
| $\rho_{36}$ | 0.0214*** | 0.0228*** | 0.0215** | 0.0235** | 0.0229*** | 0.0358*** | 0.0320*** |
| | (0.00552) | (0.00595) | (0.00778) | (0.00829) | (0.00595) | (0.00853) | (0.00899) |
| $N$ | 108524 | 104206 | 61302 | 54053 | 104206 | 54032 | 54032 |
| Seller FEs | Yes | No | No | No | Yes | Yes | Yes |
| Private Vars. | No | Yes | No | Yes | Yes | Yes | Yes |
| GSE Bids | No | No | Yes | Yes | No | Yes | No |
| Bid Differences | No | No | No | No | No | No | Yes |

Standard errors in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Bivariate probit regressions for binary performance outcomes. Selection outcome is choice to sell to the GSEs. Columns indicate alternative models controlling for additional characteristics. Rows display estimated bivariate probit correlations for 12- through 36-month survival. All regressions control for loan and borrower characteristics and fixed effects as in Table 3. Results shown in Table 3 are robust to alternative models. No residual correlation is detected between 12-month survival and sale to GSEs. However, all models detect strong residual correlation between 36-month survival and sale to the GSEs.

Results are robust to all alternative specifications. That residual correlation exists after controlling for investor bids (via bid spreads), suggests that sellers possess private information over and above what is communicated by bidders through their submitted bids. That residual correlation exists after controlling for privately-known loan characteristics suggests that private sellers' private information is not merely an artefact of non-comprehensive disclosure requirements in our wholesale auctions.

## C.2 Selection into Wholesale Market and Optimal Blue Auctions

The main analysis of this paper focuses on one particular margin of selection: selection of buyers by originators conditional on taking a loan to auction. As an alternative margin of adverse selection, we can focus on originators who sell some loans via auctions and securitize other loans outside of the Optimal Blue platform. Compared to lenders who conduct a substantial fraction of their securitization operations through the Optimal Blue platform, these lenders sell most of their auctioned loans and do not turn down bids in favor of securitization.

For the selection outcome, we focus on two choice margins. The first is the choice of the wholesale versus retail channel. Not all wholesale loan trades are conducted through the Optimal Blue platform, even for routine users of the platform. Thus when asking about adverse selection into the wholesale market, we want to look at all wholesale trades, not just those conducted via auctions. Second, we look at the decision by originators to select loans into auctions. If a loan is taken to auction rather than sold to the GSEs, the originator typically experiences a delay and incurs holding costs. Thus we should expect that loans believed to have greater survival prospects would be more likely to be sold directly to the GSEs.

*Table 14: Bivariate Probit Correlation Estimates*

| | 1(12-month Survival) | | |
|---|---|---|---|
| **Outcome** | $\rho$ | $\sigma_\rho$ | z |
| Wholesale | -.024 | .0029 | -8.35 |
| Auction | -.032 | .013 | -2.47 |
| Covariates | Yes | | |
| Fixed Effects | State, Loan Purpose, Origination Month Interest Rate | | |

$\rho$ denotes correlation with 12-month survival event.
Negative correlation indicates adverse selection.

For now, we focus on short-term survival outcomes as the measure of loan performance. In both cases, adverse selection will manifest as a negative correlation between residuals in Equation 1. In both regressions, $X$ includes major covariates visible to originators and auction participants. This includes loan amount, credit score, loan-to-value ratio, and borrower's monthly income, as well as quadratic transformations of loan amount and credit score. I include dummy variables for loan purpose, property state, note rate, and month of origination.

Table 14 shows residual correlation estimates for the two selection decisions. A negative correlation can be detected along both dimensions, suggesting adverse selection into auctions and into wholesale market more broadly. However, it is unclear from the bivariate probit exercise alone whether this adverse selection if driven by "soft" or "hard" private information.

## C.3 Unused Observables

In this Appendix we test for adverse selection is the "Unused Observables" test of Finkelstein and Poterba (2014). Similar to the positive association test of Section 4.2, we let $Y_1$ be a performance outcome and $Y_2$ a coverage/choice outcome.

$$Y_{1i} = f(X_i\beta + W_i\alpha) + \varepsilon_i, \quad Y_{2i} = g(X_i\gamma + W_i\delta) + \mu_i$$

where $X$ are used covariates and $W$ are unused covariates. We reject the null of symmetric information if $\alpha \neq 0$ and $\delta \neq 0$.

Unlike in the positive association test using a bivariate probit, $Y_1$ need not be a binary survival variable. As such, the unused observables test can capture a richer model of cash flows. Furthermore, estimating the coefficients on covariates in $W$ informs us about the dimensions along which adverse selection operates

If information in $W$ is predictive of prepayment adverse selection means that, other things being equal, originators will want to sell loans to wholesalers that are likely to prepay early while retaining and servicing loans that are unlikely to prepay early. While $W$ can represent "soft" information about borrowers (e.g., information that the originator may have picked up in their interactions with the borrower), it may also include "hard" information—characteristics of the borrower/loan that are not known to buyers at the time of sale. This is important because in most wholesale market transactions, buyers only observe a limited subset of the total available loan characteristics. These include financial characteristics such as loan size, credit score, income, loan-to-value-ratio, and interest rate. However, other characteristics are not always visible. In the auction environment, for instance, we know that certain characteristics such as borrower demographics are not visible at the time of bid formation. If these characteristics are relevant to survival, we should expect loans to be selected into wholesale markets on that basis. Whether driven by "soft" or "hard" private information, these result in starkly different survival rates between loans sold on the wholesale market and those sold to GSEs directly by originators.

For unused covariates in $W$, I focus on (1) discount points and lender credits, (2) borrower race, (3) first-time homebuyer status, and (4) the presence of a co-borrower. These covariates are unobserved (or at least not readily observable) by bidders participating in an auction, thus bids can not be conditioned on them. Furthermore, all four of these variables are likely to contribute to survival in theory:

1. Discount points and lender credits are a way for borrowers to "buy up" or "buy down" the interest rate in exchange for an up-front fee/credit. Borrowers with private information suggesting they will refinance or move shortly may want to accept credits up front, accepting a higher interest rate knowing they will not be paying this premium for long. On the other hand, borrowers who expect to be paying down their mortgage for an extended period may want to pay money up-front in exchange for a lower interest rate in the long term.

2. A borrower's race is strongly predictive of refinancing behavior—black borrowers tend not to refinance as quickly as white borrowers. This can lead to large differences in expected servicing revenues from black and white borrowers. While originators are barred from discriminating on the basis of race in the origination process, there is no restriction preventing originators from reselling loans of black and white borrowers at different rates.

3. First-time homebuyers have less exposure to mortgage payments, and may be less likely to understand the long term value of refinancing.

4. Loans with a co-borrower may be more or less susceptible to refinance and prepayment. A priori, however, it is unclear which direction (if any) the effect would point.

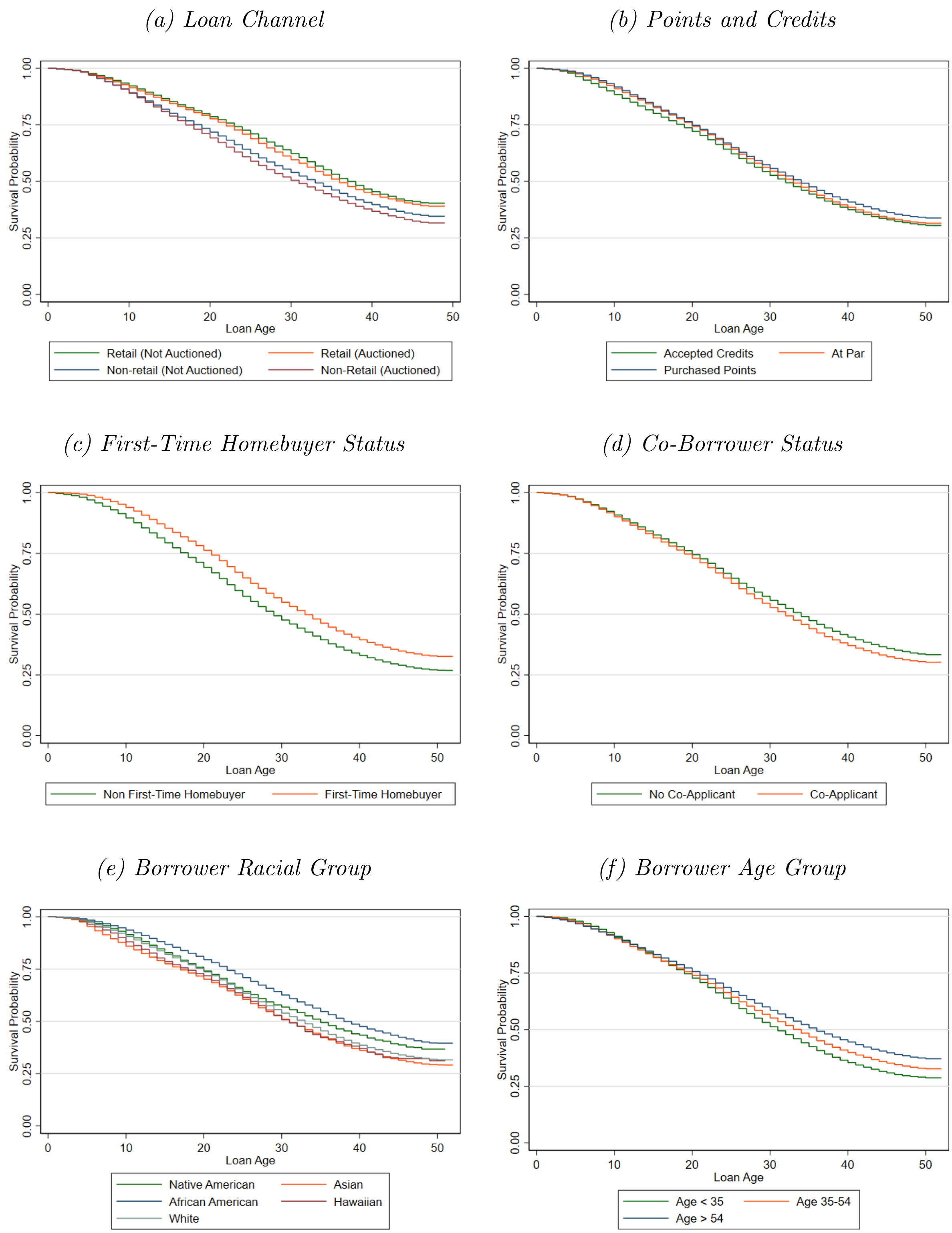
Before discussing the test results, we must argue that the privately observed covariates allow sufficient scope for adverse selection. Figure 17 shows how various characteristics which are typically unobserved to the buyer are differentially associated with survival. Mortgages for older borrowers tend to survive longer than those for younger borrowers, and mortgages for black borrowers tend to survive longer than those for white borrowers. Loans for first-time and single borrowers tend to prepay relatively slowly. And predictably loans where the borrower purchased discount points survive longer than those where the borrower accepted credits from the lender.

Finally, we show that the scope for adverse selection along the dimension of these unobserved characteristics is significant. We show this by looking at the permutation importances generated by the loan performance model. The permutation importance of feature $i$ is defined as:

$$PIMP_i = S\left(y, \hat{y}_M\left(X_1, \ldots, X_n\right)\right) - S\left(y, \hat{y}_M\left(X_1, \ldots, \pi(X_i), \ldots, X_n\right)\right)$$

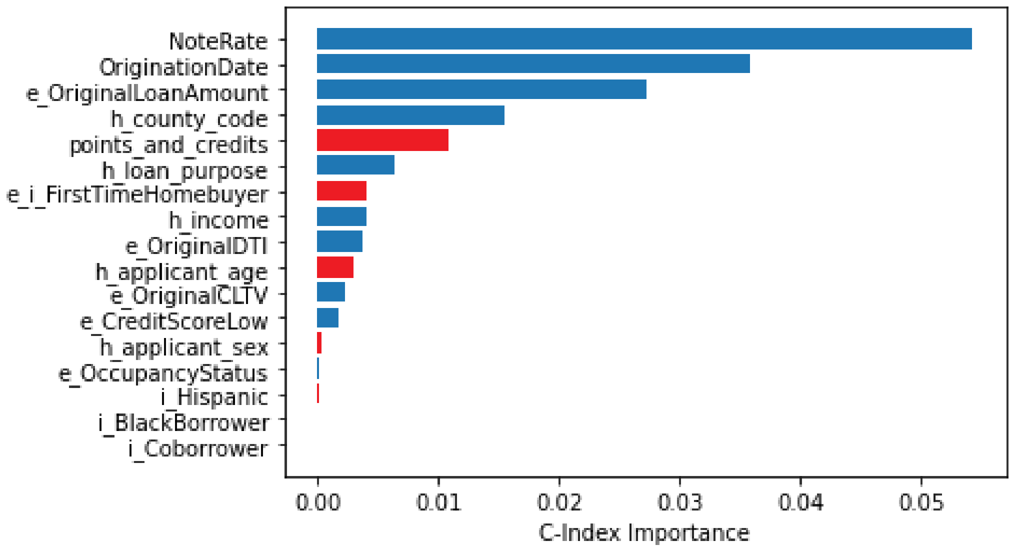
where $\hat{y}(\cdot)$ is the model prediction, $S$ is a scoring function, and $\pi$ is a permutation of the data. That is, the permutation importance measures the reduction in fit seen when randomly permuting the values of a given variable. Variables with the greatest predictive value lead to large decreases in fit when permuted, while unimportance variables cause negligible decreases. Figure 18 shows permutation importances for the random survival forest model introduced in Section 4.3, using the $C-$index of Equation 8 as the scoring function. While demographic characteristics matter little for predicting loan survival (on account of there being little variation in most borrower demographics), discount points and first-time homebuyer status contribute substantially. Discount points is the fourth most important variable for predicting loan survival, after interest rate, date, loan size, and location–four variables well-known to affect loan survival. First-time homebuyer status is roughly on par with a borrower's annual income. This suggests that insofar as wholesale markets leave these features unpriced, sophisticated sellers have additional information on which to select loans into the wholesale market.

*Figure 17: Differential Loan Survival by Unused Observables*




Note: Kaplan-Meier survival plots for distribution channel (a) and unused observables (b-f). Survival plots constructed for all loans originated by lenders participating in auctions. Survival plots for first-time homebuyer status are constructed excluding refinance loans.

*Figure 18: Permutation Importance Plot*




In this section, I focus exclusively on the choice of originators to select loans into auctions. As a first pass, I consider a simple model of binary loan survival as in Section 4.2. Table 15 shows estimates for $\alpha$ and $\delta$ using a simple linear probability model for 12-, 18-, and 24-month survival, as well as for the decision of the originator to take a loan to auction. For survival, the coefficients on points/credits, race, and first-time homebuyer status have the expected sign. The presence of a co-borrower has a positive effect on loan survival in the first year, yet this flips negative by 24-months. With the exception of discount points, the coefficients in the selection regression largely align with expectations under the hypothesis of adverse selection. Loans with a first-time homebuyer and loans with a black borrower both have longer expected survival and are predictably less likely to be auctioned and more likely to be retained for securitization directly with the GSEs.

While this model of binary survival is suggestive, it masks important information about servicing returns and doesn't adequately capture the shape of true cash flow model over time. To better account for cash flows, I model loan performance using a survival model. Survival models are standard in risk and insurance, and are frequently used to detect adverse selection. Furthermore, survival models have a long track record in the mortgage finance literature on modeling prepayment and default (see Deng et al. (2000) and Pennington-Cross (2003)). While these models often treat default and prepayment risk separately, I choose to abstract away from differences between these two "adverse" events. This assumption is not likely to be problematic in my case, since I focus on loans eligible for sale to Fannie Mae and Freddie Mac, where strict lending standards has led to extremely low overall default rates.

*Table 15: Unused Observables and Survival - Binary Model - Linear Probability*

| | (1)<br>P(12m Surv.) | (2)<br>P(18m Surv.) | (3)<br>P(24m Surv.) | (4)<br>P(Auctioned) |
|---|---|---|---|---|
| 1(Co-Applicant) | 0.312*** | -0.0845 | -1.415*** | -0.0934 |
| | (0.0505) | (0.0810) | (0.121) | (0.0736) |
| 1(First-Time Homebuyer) | 4.173*** | 5.819*** | 5.779*** | -3.600*** |
| | (0.0773) | (0.117) | (0.161) | (0.112) |
| Discount Points | 1.988*** | 3.118*** | 3.549*** | 0.845*** |
| | (0.0519) | (0.0826) | (0.119) | (0.0843) |
| Lender Credits | -0.373*** | -0.340*** | -0.175*** | 0.397*** |
| | (0.0184) | (0.0249) | (0.0284) | (0.0600) |
| 1(Black Borrower) | 2.221*** | 4.545*** | 6.646*** | -1.026*** |
| | (0.138) | (0.221) | (0.322) | (0.213) |
| Loan Characteristics | Yes | Yes | Yes | Yes |
| Originator FEs | Yes | Yes | Yes | Yes |
| State FEs | Yes | Yes | Yes | Yes |
| Rate FEs | Rate x Month | Rate x Month | Rate x Month | Rate x Month |
| $N$ | 1698343 | 1159440 | 647880 | 922442 |
| $R^2$ | 0.092 | 0.137 | 0.138 | 0.457 |

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

*Table 16: Survival Model*

| | (1) | (2) | (3) |
|---|---|---|---|
| 1(Mortgage Company) | 0.125* | 0.0898* | |
| Credit Score | 0.00150* | 0.00139* | 0.00143* |
| 1(Home Purchase) | -0.129* | -0.107* | -0.0901* |
| Lender Credits | 0.0118* | 0.0656* | 0.0628* |
| Discount Points | -0.0394* | -0.0219* | -0.0246* |
| First Time Buyer | -0.196* | -0.200* | -0.200* |
| 1(Co-Applicant) | 0.0387* | 0.0288* | 0.0292* |
| 1(Black Borrower) | -0.178* | -0.154* | -0.159* |
| Observations | 388675 | 281182 | 281182 |
| Incl. Small Origs. | Yes | No | No |
| Originator FEs | No | No | Yes |

Most covariates suppressed for interpretation.

* $p < 0.001$

I thus choose to model survival with a proportional hazard model:

$$\lambda(t, x_i, \beta, \lambda_0) = \lambda_0(t) \cdot \exp(x_i\beta)$$

where $\lambda_0$ is a non-parametric time-varying hazard function.

Table 16 shows survival estimates for three separate specifications. Negative coefficients indicate that conditional on survival at time $t$, loans with higher values of the respective covariate have a larger *instantaneous* probability of surviving to the next period. The sign of each coefficient is consonant with the estimates obtained in the binary exercise in Table 15. This result remains regardless of whether we allow for systematic differences in loan survival across originators to account for differing quality in the underlying loan pool.

To flexibly model the originator's selection decision, I use the double debiased machine learning (DML) approach of Chernozhukov et al. (2018). I model the selection decision as:

$$Y_{2i} = \delta W_i + g(X_i) + \varepsilon_{1i}$$

where

$$W_i = m(X_i) + \varepsilon_{2i}$$

and $m()$, $g()$ are high dimensional "nuisance" functions of the underlying covariates $X$ to be estimated using an appropriate choice of machine learning method.

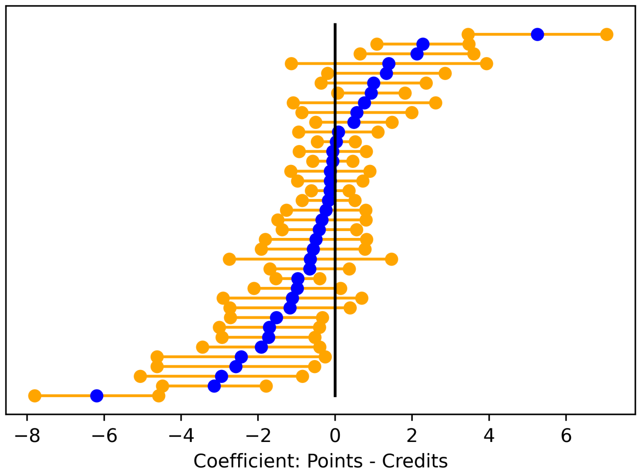
This method is capable of capturing complicated underlying relationship between loan characteristics. In the case of mortgages, this true relationship between loan characteristics and the selection decision is likely extremely complicated, since it reflects expectations about the bids of multiple bidders in a network. Moreover, the information that is visible to bidders in the auction is can be used to generate reasonable predictions about unused covariates $W$. In effect, the DML approach "nets out" the best predicted value of the unused observables, and the $\delta$ coefficient applies to the residual component of $W$.

We require the chosen machine learning method to do a "good enough" job approximating the underlying $m()$ and $g()$ functions. Due to well-known discontinuities in the mortgage market, tree based methods are a particularly appropriate choice here. Table 17 displays results from the DML estimation using random forests and XGBoosting to handle the nuisance functions. Both methods agree on the sign and magnitude of first-time homebuyer status and the presence of a co-borrower. The latter case is particularly notable, as the linear model showed no significant selection along this dimension. The two algorithms disagree strongly about selection on race and points (though as Figure 19, suggests, this might owe to heterogeneity across sellers in whether and how this variable is selected on). The random forest estimates are consistent with large degree of selection out of auctions for loans with black borrowers, yet XGBoosting is consistent with the prediction that loans with discount points are likely selected out of auctions.

*Table 17: Selection into Auctions (DML)*

| | Learning Algorithm | |
|---|---|---|
| | Random Forest | Xtreme Gradient Boosting |
| Net Points | 0.2198 | -0.1404 |
| | (0.0452) | (0.0433) |
| $\mathbb{1}$ (First-Time Homebuyer) | -0.4005 | -0.4144 |
| | (0.0791) | (0.0790) |
| $\mathbb{1}$ (Co-Borrower) | -0.1531 | -0.2261 |
| | (0.0538) | (0.0534) |
| $\mathbb{1}$ (Black Borrower) | -0.3515 | -.0340 |
| | (0.1413) | (0.1387) |
| Observations | 1231586 | 1231586 |

*Figure 19: Points and Fees Selection Coefficient by Seller*



# D Model Details: Proofs and Estimation

## D.1 A Non-parametric Model of Loan Survival

This Appendix expands on the nonparametric survival exercise on Section 4.3. My preferred method for modeling survival is the random survival forest method developed in Ishwaran et al. (2008) and extended in Ishwaran et al. (2014). The random survival forest is a non-parametric model built from an ensemble of survival trees, each of which uses a bootstrap sample and a random subset of the covariates. Each tree is grown by making successive binary splits of the data at points chosen to maximize survival differences between subsamples. Each terminal node $h \in \mathcal{H}$ in the survival tree is characterized by a tuple of survival and censoring outcomes:

$$(T_{1,h}, \delta_{1,h}), \ldots, (T_{n(h),h}, \delta_{n(h),h})$$

and each terminal node has a cumulative hazard function given by the Nelson-Ahlen estimator:

$$\hat{H}_h(t) = \sum_{t_{l,h} \leq t} \frac{d_{l,h}}{Y_{l,h}}$$

where $d_{l,h}$ is total adverse events and $Y_{l,h}$ is total at-risk population. Once an ensemble of survival trees has been grown, predictions from the trees can be combined into an overall cumulative hazard function $\hat{H}_e^o$ as:

$$\hat{H}_e^o(t|X_i) = \frac{1}{B} \sum_{b \in \mathcal{B}} \hat{H}_b(t|X_i)$$

where $\hat{H}_b(t|X_i)$ is the cumulative hazard function prediction from the appropriate terminal node for covariates $X_i$ in the tree grown on bootstrap sample $b$.

To model loan survival, we use a variety of borrower, lender, and loan characteristics. For borrower characteristics, we include the borrower's age, ethnicity, race, and sex. The identity of the lender is included via random numerical encoding. Finally, we include a rich set of loan characteristics, including loan amount, property value, DTI ratio, discount points, interest rate, credit score, occupancy status, and mortgage date.

To judge the performance of the random survival forest model of Section 4.3, I compute Harrell's concordance index (or C-index)–a measure of fit defined as:

$$C = \frac{\sum_{i \neq j} \mathbf{1}\{\eta_i < \eta_j\} \mathbf{1}\{T_i > T_j\} d_j}{\sum_{i \neq j} \mathbf{1}\{T_i > T_j\} d_j} \tag{8}$$

where $\eta_i$, $T_i$ and $d_i$ are hazard values, event times, and censoring indicators respectively. The index compares all pairs of observations in the data whose relative survival outcomes are known and calculates the fraction for which the model's hazard function correctly ranks them in accordance with their observed survival times. In a sample of the matched mortgage data, the C-index value of the random survival forest is 0.72. That is, based on the model's survival predictions, we can correctly rank survival duration for 72 percent of loan-pairs.

## D.2 Model Proofs

**Proposition:** There exists a threshold function $\bar{P}$ such that, conditional on a loan being taken to auction, $\bar{P}$ partitions the space of possible maximum bids into a region where selling is a dominant strategy and a region where keeping is a dominant strategy.

It's easier to first compute the $\eta$ threshold:

$$\begin{aligned}
& P^0 + \mathbb{E}\left[M|P^*, \omega, X\right] = P^* \\
\iff\quad & P^0 + \mu_M(X_{ij}) + \mathbb{E}\left[\varepsilon|\omega, \eta\right] = \mu_P(X_{ij}) + \eta \\
\iff\quad & P^0 + \mu_M(X_{ij}) + \omega\left(\frac{\sigma_{MS}\sigma_P^2 - \sigma_{MP}\sigma_{SP}}{\sigma_S^2\sigma_P^2 - \sigma_{SP}^2}\right) + \eta\left(\frac{\sigma_{MP}\sigma_S^2 - \sigma_{MS}\sigma_{SP}}{\sigma_S^2\sigma_P^2 - \sigma_{SP}^2}\right) = \mu_P(X_{ij}) + \eta \\
\iff\quad & P^0 + \mu_M(X_{ij}) - \mu_P(X_{ij}) + \omega\left(\frac{\sigma_{MS}\sigma_P^2 - \sigma_{MP}\sigma_{SP}}{\sigma_S^2\sigma_P^2 - \sigma_{SP}^2}\right) = \eta\left(1 - \left(\frac{\sigma_{MP}\sigma_S^2 - \sigma_{MS}\sigma_{SP}}{\sigma_S^2\sigma_P^2 - \sigma_{SP}^2}\right)\right) \\
\iff\quad & \frac{P^0 + \mu_M(X_{ij}) - \mu_P(X_{ij}) + \omega\left(\frac{\sigma_{MS}\sigma_P^2 - \sigma_{MP}\sigma_{SP}}{\sigma_S^2\sigma_P^2 - \sigma_{SP}^2}\right)}{\left(1 - \left(\frac{\sigma_{MP}\sigma_S^2 - \sigma_{MS}\sigma_{SP}}{\sigma_S^2\sigma_P^2 - \sigma_{SP}^2}\right)\right)} = \eta
\end{aligned}$$

This translates straightforwardly to a $P^*$ threshold:

$$P^*(\omega) = \frac{P^0 + \mu_M(X_{ij}) - \left(\frac{\sigma_{MP}\sigma_S^2 - \sigma_{MS}\sigma_{SP}}{\sigma_S^2\sigma_P^2 - \sigma_{SP}^2}\right)\mu_P(X_{ij}) + \omega\left(\frac{\sigma_{MS}\sigma_P^2 - \sigma_{MP}\sigma_{SP}}{\sigma_S^2\sigma_P^2 - \sigma_{SP}^2}\right)}{\left(1 - \left(\frac{\sigma_{MP}\sigma_S^2 - \sigma_{MS}\sigma_{SP}}{\sigma_S^2\sigma_P^2 - \sigma_{SP}^2}\right)\right)}$$

### D.2.1 The Value of Committal Auctions

As a preliminary remark, suppose that the seller could commit to selling at auction. In this case, the conditional expectation of selling at auction would be given by:

$$\mathbb{E}\left[P^*|\omega\right] = \mu_P(X_{ij}) + \mathbb{E}\left[\eta|\omega\right] = \mu_P(X_{ij}) + \frac{\sigma_{SP}}{\sigma_S^2}\omega$$

By equating the value of the retail and auction components, we get a $\omega$ threshold:

$$\bar{\omega} = \frac{\sigma_S^2}{\sigma_{MS} - \sigma_{SP}} \left(\mu_P - P^0 - \mu_M - c_O\right)$$

In the case where $\sigma_{SP} > \sigma_{MS}$, for $\omega$ values above this threshold, auctioning is preferred to direct retail. If $\sigma_{SP} < \sigma_{MS}$, then expected bids grow slower than expected cash flows in $\omega$, thus there is a threshold above which auctions are not used.

## D.3 Estimation Details

### D.3.1 Untruncated Densities

The untruncated density functions are given by:

$$\begin{aligned}
f_U(M, P^*, Sell|X_{ij}, \mu, \Sigma) &= \int_{e\in\Omega_{sell}} f_\omega(M - \mu_M, \omega, P^* - \mu_P) d\omega \\
f_U(M, P^*, Keep|X_{ij}, \mu, \Sigma) &= \int_{e\in\Omega_{keep}} f_\omega(M - \mu_M, \omega, P^* - \mu_P) d\omega
\end{aligned}$$

where $f$ is the multivariate normal pdf:

$$f_\omega(M - \mu_M, \omega, P^* - \mu_P) = \frac{1}{\sqrt{(2\pi)^3 |\Sigma|}} \exp\left(-\frac{1}{2}\left(M - \mu_M,\ \omega,\ P^* - \mu_P\right)\Sigma^{-1}\begin{pmatrix} M - \mu_M \\ \omega \\ P^* - \mu_P \end{pmatrix}\right)$$

### D.3.2 Truncated Densities

The truncated density functions are given by:

$$\begin{aligned}
f_T(M, P^*, Sell|X_{ij}, \mu, \Sigma) &= \int_{\underline{\varepsilon}^1}^{\infty} \int_{\omega\in\Omega_{sell}} f_e(\varepsilon, \omega, P^* - \mu_P) d\omega d\varepsilon \\
f_T(M, P^*, Keep|X_{ij}, \mu, \Sigma) &= \int_{\underline{\varepsilon}^1}^{\infty} \int_{\omega\in\Omega_{keep}} f_e(\varepsilon, \omega, P^* - \mu_P) d\omega d\varepsilon
\end{aligned}$$

which can be written as:

$$\begin{aligned} f_T(M, P^*, \cdot|X_{ij}, \mu, \Sigma) &= \int_{\omega\in\Omega.} \int_{\underline{\varepsilon}^1}^{\infty} f_e(\varepsilon, \omega, P^* - \mu_P) d\varepsilon d\omega \\ &= \int_{\omega\in\Omega.} \left[\int_{\underline{\varepsilon}^1}^{\infty} f_1(\varepsilon|\omega, P^* - \mu_P) d\varepsilon\right] f_2(\omega, P^* - \mu_P) d\omega \\ &= \int_{\omega\in\Omega.} \Phi_{\omega, P^*-\mu_P}(\varepsilon) f_2(\omega, P^* - \mu_P) d\omega \end{aligned}$$

### D.3.3 Likelihood Function

Thus the likelihood function can be written as:

$$\begin{aligned} \mathcal{L}(\{M_i, P_i^*, Y_i\}_1^n | Z, \mu, \Sigma) = \textstyle\prod_{i=1}^n \quad & f_U(M_i, P_i^*, Sell|X_{ij}, \mu, \Sigma)^{\mathbf{1}_{\{Y_i=Sell\}}} \\ \cdot \quad & f_T(M_i, P_i^*, Sell|X_{ij}, \mu, \Sigma)^{\mathbf{1}_{\{Y_i=Sell, M_i=\bar{M}\}}} \\ \cdot \quad & f_U(M_i, P_i^*, Keep|X_{ij}, \mu, \Sigma)^{\mathbf{1}_{\{Y_i=Keep\}}} \\ \cdot \quad & f_T(M_i, P_i^*, Keep|X_{ij}, \mu, \Sigma)^{\mathbf{1}_{\{Y_i=Keep, M_i=\bar{M}\}}} \end{aligned}$$

In the fully untruncated case, this reduces to:

$$\mathcal{L}(\{M_i, P_i^*, Y_i\}_1^n | Z, \mu, \Sigma) = \left(\frac{1}{\sqrt{(2\pi)^3 |\Sigma|}}\right)^n \prod_{i=1}^n \left(\int_{\omega\in\Omega_{keep}} g_e de\right)^{\mathbf{1}_{\{Y_i=Keep\}}} \left(\int_{\omega\in\Omega_{sell}} g_e de\right)^{\mathbf{1}_{\{Y_i=Sell\}}}$$

where

$$g_e(x) = \exp\left(-\frac{1}{2} x^T \Sigma^{-1} x\right)$$

Here the integrals can be estimated via quadrature. I use 81-point Gauss-Hermite quadrature to simulate the integral over $\omega$. The large number of points is needed to guarantee nice behavior of the MLE routine when taking numerical derivatives.

# E Loan Origination and Platform Use

Table 18 reports the results of a descriptive regression of quarterly origination volume on a dummy for whether or not the loan originator participates in the auction platform. Use of the auction platform is associated with 13% greater origination volume (or about 100 more originated loans per quarter). Because platform entry is not random, these estimates cannot be interpreted causally. Platform entry may enable lenders to originate more loans, or lenders who begin to originate more may be more willing to use auctions to sell their loans.

*Table 18: Quarterly Originations Before and After Platform Entry*

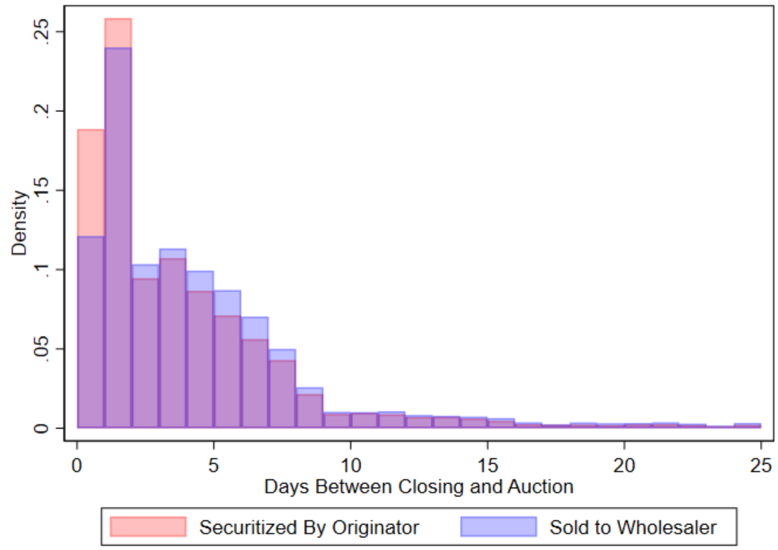
| | (1)<br>log(Quarterly Originations) | (2)<br># Quarterly Originations |
|---|---|---|
| 1(Before Auction) | -0.139$^{*}$ | -110.9$^{**}$ |
| | (0.0614) | (41.34) |
| $N$ | 1553 | 1553 |
| $R^2$ | 0.744 | 0.785 |
| SellerFE | True | True |
| QuarterFE | True | True |

Standard errors in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

One potential benefit of using Optimal Blue's auction platform is that it allows quick turnaround from loan closing to sale. Figure 20 shows the distribution of time from closing to auction for loans sold via the Optimal Blue platform. Roughly half of loans are auctioned within two days of closing, and over 90% are auctioned within one week. Loans that end up in the retail channel are more likely to be auctioned within two days. This may be due to quicker delivery times when loans are sold via the GSE cash window, or it may reflect the time required to assemble a competitive pool of bidders. Regardless of delivery method, these delivery times are quite short in comparison to the average of 15-days during which loans are funded by credit lines (see Kim et al. (2018)). If quick delivery times can free up additional resources for mortgage lending, this can partially explain why origination volumes would increase after platform entry.

*Figure 20: Distribution of Time from Closing to Auction*



# F Detailed Matching Appendix

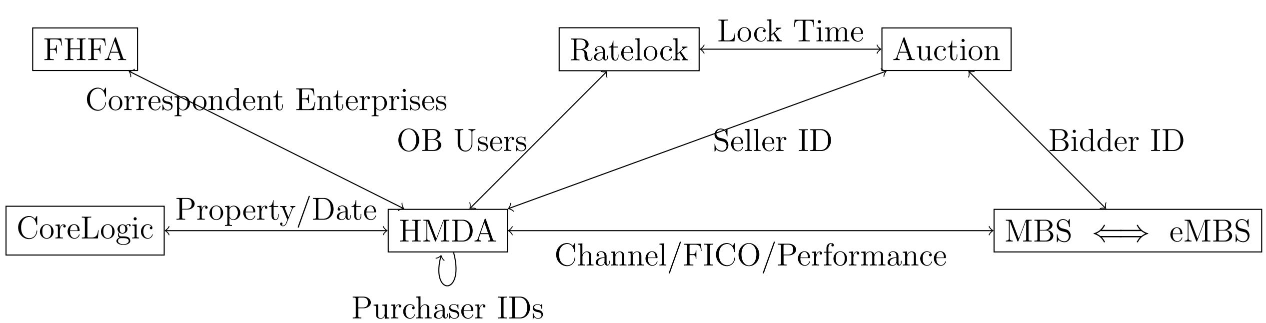
## F.1 Dataset Summary

*Table 19: Dataset Sources with Contributing Variables*

| Dataset Name | Source | Contributing Variables |
|---|---|---|
| Auction | Optimal Blue | Loan Characteristics (exact), Closing Date (exact), Property Location (exact), Lender Name, Issuer Name |
| eMBS | Black Knight | Loan Characteristics (rounded), Issuer Name, First-Time Homebuyer Status |
| HMDA | Consumer Finance Protection Bureau | Loan Characteristics (rounded), Property Location (exact), Lender Name, Borrower Demographics |
| Public MBS | Fannie Mae, Freddie Mac, Ginnie Mae | Loan Characteristics (rounded), Property Location (rounded), Closing Date (rounded) |
| CoreLogic Mortgages | CoreLogic | Loan Amount (exact), Closing Date (exact), Property Location (exact), Lender Name |
| FHA | Federal Housing Administration | Loan Amount (exact), Closing Date (rounded), Property Location (exact), Lender Name |
| FHFA | Federal Housing Finance Agency | Loan-to-Value (exact), Borrower Demographics, Property Location (exact), First-Time Homebuyer Status |
| RateLock | Optimal Blue | Lock period, Loan Characteristics (exact), Property Location (exact) |

Note: Only variables contributing to the matches below are listed. Variables denoted as "exact" are provided by the respective dataset in an unrounded form (e.g., loan amount to the nearest dollar, exact dates, LTV to the nearest .01%). Variables denoted as "rounded" may be rounded in such a way that significant information is lost (e.g., loan amount to the nearest 10000, dates to the nearest month, LTV to the nearest percentage point). When matching between datasets, exact values almost always allow for immediate unique matches in conjunction with other standard match variables such as state and year.

*Figure 21: Dataset Matching Relationships*




Note: Additional fields noted in figure are relative to baseline fields in HMDA. Details are included in Appendix F Table 19

## F.2 HMDA Sellers to HMDA Purchasers

Within the HMDA data, observations corresponding to closed mortgages are denoted as either "originations" or "purchases." Loans marked as "purchased" have been purchased after origination by an institution other than the originator. Further, all hmda observations contain a "purchaser type" variable indicating the type of institution to which the loan was first sold. Loans sold on the wholesale market have a separate value of this purchaser type variable from loans sold to one of the GSEs.

To match within HMDA, I conduct an initial strict match on loan characteristic variables. After this, I keep candidate matches which either match strictly on other variables or have missing values of those variables due to differences in reporting requirements between sellers and purchasers. Finally, I keep candidate matches which have one or more strict match on loan cost variables. Because loan cost variables are reported to the nearest cent, resulting matches are virtually certain to be genuine, given the previous strict match on geography and loan size.

## F.3 HMDA to FHFA

To match HMDA with FHFA data, we first conduct a strict match on Census tract, rounded loan amount, year, GSE, government agency, and rounded interest rate for loans HMDA identifies as securitized within the same calendar year they were originated. The remaining loans are matched again without the requirement of a strict match on year or GSE. In both rounds, we require a tight match on all non-missing characteristics common to the two datasets. Additional rounds repeat this procedure, relaxing the requirements for tight matches on all covariates, while paying special attention to perfect matches on continuous characteristics such as loan-to-value ratio and discount points.

The match rate between HMDA originations and FHFA issuances is high. Greater than 98% of originated loans in HMDA that are known to be securitized with the GSEs can be matched to a unique loan in the FHFA data. Of all loans reported by FHFA, approximately 93% have a match in HMDA. The CFPB does not require all lenders to report their origination activity, only those who meet lending volume requirements in successive years. Thus the slightly lower match rate here is consistent with industry standard wisdom that HMDA covers close to 95% of originated loans.

## F.4 HMDA to CoreLogic

To merge HMDA loans to CoreLogic mortgages, we first obtain geocoded census tracts for CoreLogic preoperty addresses. Then mortgages are matched on census tract, government agency, loan purpose, and rounded mortgage amount. When available, initial interest rates and loan terms are used to restrict matches. The resulting candidate matches can be used to tighten the match between HMDA and MBS loans by bringing in exact mortgage amounts and loan origination dates.

## F.5 HMDA to Auctions

To match HMDA with Optimal Blue auctions, we first prepare HMDA by dropping mortgages for multi-family properties ($> 4$ units) and adding zip codes for each HMDA observation by matching on a zip code to census tract crosswalk. Then we prepare the auction data by dropping duplicates of auctioned loans (approximately 12 thousand out of 1.2 million auctioned loans have strict duplicates on all characteristics except for the auction date).

Then we conduct a strict match on broad loan characteristics: sponsoring government agency, loan term, property location (i.e., zip code and county), interest rate (rounded to 1/8 of a percentage point), occupancy status, number of units, a dummy for whether the loan was for a home purchase, and a dummy for whether the loan was an adjustable rate mortgage. Then we do perform a fuzzy but tight match on continuous loan covariates, including loan amount, yearly income, DTI ratio, and CLTV ratio.

After the first round, we keep unique matches—that is, matches where the HMDA loan matches to a unique auctioned loan and where the auctioned loan matches to a unique HMDA loan. At this stage, roughly 79% of auctioned loans have a unique match. With these unique matches, we construct a crosswalk between HMDA originators and auction sellers, keeping matches where the majority of a HMDA lender's candidate loan matches occur for the same auction seller. The majority ($\sim 95\%$) of auction sellers match to a unique HMDA originator. However, a few large sellers match to multiple HMDA originators, suggesting that they acquire loans from other originators prior to selling at auction.[27] With the originator-seller crosswalk constructed, we revisit first round matches keeping the strict on discrete loan covariates, while loosening the match criteria on continuous covariates.[28] After two rounds,

[27]Consistent with this picture, sellers who appear to acquire loans from multiple originators have a longer time between the borrower's closing date and the date that loan goes to auction.

[28]For a few small lenders, the debt-to-income ratio reported in HMDA differs systematically from that listed on Optimal Blue. For these lenders we match on all covariates except for DTI.

we are able to match approximately 85% of auctioned loans with a unique loan in HMDA.

## F.6 Auctions to MBS

To match Optimal Blue auctions with MBS issuances, we first construct a crosswalk between the publicly available MBS issuance data and eMBS issuance data from Black Knight.[29] Both files must be used, as public files contain information on three-digit zip code unavailable in eMBS, while eMBS contains seller names for small sellers which are not available in the public MBS files.[30]

We conduct a strict match on broad loan characteristics: sponsoring government agency, loan term, property location (i.e., zip code and county), interest rate (rounded to 1/8 of a percentage point), occupancy status, number of units, and a dummy for whether the loan was for a home purchase. Then we perform a fuzzy but tight match on continuous loan covariates, including loan amount,[31] credit score, [32] DTI ratio, and LTV/CLTV ratios.

After a single round, we are able to match approximately 91% of auctioned conventional loans with a unique loan in eMBS. To verify matches, we look at those loans that were auctioned but then sold directly by the originator to Fannie Mae or Freddie Mac. These loans are denoted by Fannie/Freddie as "Retail Loans," and greater than 99.5% of candidate matches of retail loans have a seller in eMBS which aligns with the would-be seller of the loan at auction which is known from HMDA.[33]

## F.7 HMDA to MBS

To match HMDA originations with MBS issuances, we first conduct three supplementary matches: (1) a match between the publicly available MBS issuance data and eMBS issuance data from Black Knight, (2) a match between purchased loans and sold loans withing

[29]This crosswalk is constructed by using a strict match on a loan's first payment date and then a tight fuzzy match on all shared covariates. Approximately 99.5% of all issuances in the public files match to a unique eMBS loan.

[30]The public MBS files provide names for the largest sellers/issuers, but censors the names of smaller sellers. Approximately 45% of loans sold to Fannie Mae and 35% of those sold to Freddie Mac have censored names in the public files.

[31]Loan amounts in the MBS files are more fine grained than in HMDA, making the initial match more reliable than the initial HMDA to auction match.

[32]For loans with two or more borrowers, the auction data records the lowest of the borrowers' credit scores.

[33]While not all sellers utilize the retail channel, roughly 1/3 of loans taken to auction are ultimately securitized by the originator. Without the small issuer information contained in the eMBS data, we would be able to check this for fewer than five of 150 sellers in the auctions.

HMDA,[34] and (3) a match between originated HMDA loans and FHFA data on loans acquired by Fannie Mae and Freddie Mac.

In the first round of matches, we match known retail loans (those HMDA indicates are sold to Fannie or Freddie by the originator).[35] We conduct a strict match on agency (Fannie or Freddie), loan purpose, rounded loan amount, three-digit zip code, interest rate (rounded to 1/8 of a percentage point), occupancy status, loan term, number of units, and a dummy for whether the mortgage had a single applicant. Then we conduct a coarse match on CLTV ratio and DTI ratio.

After this initial match, we keep loans with a unique match and construct a crosswalk between HMDA originators and "sellers" in the MBS data.[36] We drop matches from the initial pool of (non-unique) candidate matches, leaving us with approximately 80% of retail loans (according to HMDA) having a unique match.

In the second round, we use matches (2) and (3) above to match loans in the correspondent channel. These matches indicate whether the loan was sold to Fannie or Freddie, cutting down on the number of possible matches drastically.[37] With the agencies known, we find matches using the same procedure from the first round.[38]

In a final round, we match remaining loans by first removing all matched loans from consideration in both HMDA and MBS and matching as in round 1, but without originator-seller information.

After these three rounds we are left with over 19 million matches between HMDA originations and MBS issuances from 2018-2021.[39]

[34] For loans in HMDA which are originated and sold to a third party with HMDA reporting requirements, both the origination and the purchase will be recorded. Using census tract, loan size, and fees (made available in 2018 and afterward), we can match roughly half of purchased loans in HMDA to an originated HMDA loan.

[35] While the HMDA purchaser variable appears to be inconsistently reported for loans sold to other institutions, initial checks reveal virtually no mis-reporting of the variable for Ginnie Mae, Fannie Mae, and Freddie Mac.

[36] At this point, small seller names from eMBS are invaluable, as they allow us to trim potential matches for loans originated by small lenders who would not be assigned a unique ID in the public MBS data.

[37] The match between HMDA sellers and HMDA purchasers tells us which of the two agencies purchased the loan from the initial purchaser in the wholesale market. The match between FHFA and HMDA tells us which of the two agencies eventually acquired the loan.

[38] For loans whose acquiring agency is only known through the match with FHFA, originator-seller information cannot be used to narrow down matches. However, such information can be used for loans where agency information is known through the internal HMDA-HMDA match.

[39] For context, Buchak and Jorring also match HMDA with MBS, and they find 6 million matches for 2018-2021. Our higher match rate is due to the use of seller information from eMBS, as well as increased information coming from HMDA-HMDA matches and HMDA-FHFA matches.