The Market Effects of Algorithms*

Lindsey Raymond
MIT

December 27, 2024

First Draft: December 2023

Please see here for latest version

Abstract

While there is excitement about the potential of algorithms to optimize individual decision-making, changes in individual behavior will, almost inevitably, impact markets. Yet little is known about these effects. In this paper, I study how the availability of algorithmic prediction changes entry, allocation, and prices in the US residential real estate market, a key driver of household wealth. I identify a *market-level* natural experiment that generates variation in the cost of using algorithms to value houses: digitization, the transition from physical to digital housing records. I show that digitization leads to entry by investors using algorithms, but does not push out investors using human judgment. Instead, human investors shift towards houses that are difficult to predict algorithmically. Algorithmic investors predominantly purchase minority-owned homes, a segment of the market where humans may be biased. Digitization increases the average sale price of minority-owned homes by 5% or $5,000 and reduces racial disparities in home prices. Algorithmic investors, via competition, affect the prices paid by humans for minority homes, which drives most of the reduction in racial disparities. This decrease in racial inequality underscores the potential of algorithms to mitigate human biases at the market level.

JEL Classifications: D22, D40, G41, H41, L11, L22, M13, O33

Keywords: Digitization, Machine Learning, Artificial Intelligence, Technology Adoption, Market Structure, Housing, Discrimination.

*Correspondence to lraymond@mit.edu. I am especially grateful to Sendhil Mullainathan, Danielle Li, Erik Brynjolfsson and Scott Stern for their invaluable feedback and advice. I thank Eric Budish, Mert Demirer, Brandon Enriquez, Robert Gibbons, Anh Nguyen, Manish Raghavan, Ashesh Rambachan, David Reshef, Daniel Rock, Frank Schilbach, Eric So, Evan Soltas, Maggie Yellen and various seminar participants for helpful comments and suggestions. I thank Max Feng for providing excellent research assistance and the Stanford Digital Economy Lab for financial support. The content is solely the responsibility of the authors and does not necessarily represent the official views of Stanford University or the Massachusetts Institute of Technology.

Prediction underlies many high-stakes decisions – hiring depends on forecasting candidates' potential productivity, extending credit on risk of default, and investment on projections of returns. The advent of machine learning (ML) and digital data has sparked interest in making such predictions algorithmically. A large and growing literature has begun to document when and how algorithms outperform human predictions. Although much progress has been made on how algorithms impact decision quality, their *market-level* impacts remain less explored.[1]

If algorithms change individual decisions, their use could also impact market-level outcomes, such as prices, or even change the nature of competition. These broader market dynamics mean that even if algorithms improve decision quality, individuals could be worse off. Studies focused on measuring the impact of algorithms on decision quality cannot account for effects beyond the individual or firm level.

To examine these market-level effects, I study how digital data and the shift to algorithmic prediction among investors impact the US residential housing market. While the use of data and algorithms is growing across many industries, the single-family housing market offers an ideal setting for these questions for three reasons. First, homes are economically significant: housing is the largest asset market and important driver of household wealth (Malone, 2023; Derenoncourt et al., 2022). Second, prediction plays a central role in investor decisions. Although the most familiar participants in the single-family housing market are owner-occupiers, investors play an important role in the market, and investment decisions depend on predictions of rental income, appreciation, and maintenance costs. Lastly, a natural experiment enables causal identification such that I can examine the consequences for market prices and allocations.

The central identification challenge is that the adoption of algorithmic prediction is not random across markets. In this paper, I develop a novel empirical strategy to identify market-wide effects by exploiting a simple yet fundamental insight: machine learning algorithms require machine-readable data. Specifically, I compare housing markets where only human judgment is feasible to those where machine-readable data availability makes algorithmic prediction less costly. In the U.S., county governments collect public records on the housing stock and transactions for routine administrative tasks. While much of this information was kept as physical records, counties began to transition to electronic database systems as part of the Open Government movement. The staggered rollout of this digitization created variation in the cost of accessing housing market data, and thus algorithmic prediction, across counties and over time. This staggered rollout enables the comparison of market-

[1] An algorithm is a computer hardware or software-based routine that carries out a list of instructions or a process (Sunstein, 2024).

level prices and allocations between counties before and after data digitization and between counties that have not yet digitized.

I have four sets of findings.

First, county digitization leads to entry by investors relying on algorithmic prediction, which I refer to as algorithmic investors.[2] Investors, who buy houses to resell or rent, appear in housing transaction data as corporate entities purchasing single-family homes. I classify firms that disclose using algorithmic methods as "algorithmic investors" and those that do not as "human investors." Before county digitization, algorithmic investment is rare, with many algorithmic investment firms not yet established. After digitization, the share of houses purchased by algorithmic investors increases sharply and persistently: algorithmic investors acquire approximately 1.8 percent of all annual housing purchases, which includes both owner-occupiers and investors. Relative to purchases made by only investors (human or algorithmic), algorithmic investors account for about 10 percent of all investor buying.

These initial findings could be misleading if unobserved county-level changes, like business policies or economic growth, drive both digitization timing and housing market activity. To address this, I exploit within-county bureaucratic inconsistencies in digitization timing: due to budget constraints, counties digitize records in batches, creating variation in digital availability across properties. Using not-yet-digitized houses as a control group, I conduct falsification tests and triple-difference analyses within counties and neighborhoods. The effects of county digitization consistently concentrate on digitized properties, suggesting algorithmic investor activity responds to data accessibility rather than unobserved county- or neighborhood-level changes.

My second set of results examines patterns in algorithmic investment. To understand this variation in investor activity, I develop a conceptual framework that highlights the distinct comparative advantages of human versus algorithmic prediction. Algorithms are machine learning procedures that identify statistical patterns from quantifiable data. In contrast, humans have access to unobservable factors not captured in datasets–from the quality of bathroom tile work to a yard's sunlight exposure to ambient neighborhood noise. Yet this advantage in information richness comes with a tradeoff: while humans can access private information, their predictions may be vulnerable to cognitive limitations and biases that algorithms, through their strict reliance on statistical relationships in data, avoid.

The first implication of this framework is that the ability for algorithmic investors to evaluate houses depends on the digitized property information. I exploit three institutional rules—lead paint disclosure requirements, data entry, and county zoning regulations—to identify variation in

[2]Algorithmic investors, described in Section 2.1.5, combine algorithmic prediction with teams of human analytsts.

data quality across properties. Following digitization, algorithmic investors concentrate their purchases in properties with high-quality digital records. Conversely, human investors appear to react strategically by increasing their investment in properties with poor data quality. To test these patterns more systematically, I develop an extreme gradient boosted tree model that generates an ex-ante measure of property predictability. The analysis confirms that the impact of record digitization on algorithmic investment is substantially larger for more predictable properties–those with comprehensive digital records and characteristics that are more amenable to algorithmic forecasting. Meanwhile, human investors increase their activity in less predictable properties, suggesting a market-wide reallocation based on prediction technology comparative advantage.

The second implication is that human biases can lead humans to undervalue houses, creating a potential opportunity for algorithmic investors. The key challenge, for both algorithmic investors and the analysis, is to distinguish areas of opportunity due to human biases from houses that are poor quality due to soft information inaccessible to algorithms. A variety of human behavioral biases could be at play, but I focus on one of the most studied – the role of race in the housing market. While the Fair Housing Act prohibits racial discrimination in the sale, rental and financing of housing, racial disparities remain and are the subject of extensive study (Elster and Zussman, 2022; Perry, Rothwell and Harshbarger, 2018; Freddie Mac Economic & Housing Research, 2021; Cutler, Glaeser and Vigdor, 1999). In support of the idea that humans may undervalue houses sold by minorities, I also show that, prior to digitization, a race penalty difference exists between observably similar houses sold by white and minority homeowners within the same neighborhood. In principle, this relative price difference, which I refer to as the *race penalty*, could be due to unobserved differences between houses, even at the neighborhood level. For instance, minority homeowners are often more cash-constrained or less wealthy, leading to differences in home maintenance or yard care (Perry, Rothwell and Harshbarger, 2018; Harris, 1999). To investigate whether this gap simply reflects omitted variables associated with home appearance, I create embedding representations of houses using a deep learning model trained on images of house exteriors, yards, and driveways. After controlling for aspects of house quality captured by house images, the race penalty persists. This suggests that differences in property maintenance or aesthetic factors do not fully explain the lower price received by minority homeowners.

Consistent with the possibility that humans may be undervaluing minority-owned homes, algorithmic investors disproportionately buy minority-owned homes. After digitization, algorithmic investors are six times more likely to buy a minority-owned home than a White-owned one in the same census block group. In other words, digitization of minority homes leads to a 250 percent increase in the probability that an algorithmic investor buys that home compared with 40 percent for a White-owned house. The difference in treatment effects is not driven by differences in the racial composition of neighborhoods where algorithmic and human investors are active; these results compare investor activity within the same neighborhood. In fact, algorithmic investors tend not to buy in neighborhoods with high minority shares and the neighborhoods where they are active have similar minority shares to those where the human investments occurs. In addition, minority-owned homes make up a disproportionate share of algorithmic investors house purchases. Together, these two facts indicate algorithmic investor behavior is not merely a function of operating in neighborhoods with higher shares of minority residents, but rather may reflects a specific strategy focused on acquiring minority-owned homes.

Having established that algorithmic investment is concentrated in minority homes and the existence of a race gap prior to digitization, I next turn to the impacts of algorithmic investment on prices, specifically the race penalty. Prior to digitization, the race penalty among houses purchased by human investors and owner-occupiers is about 5%. However, algorithmic investors, who enter the market after digitization, do not exhibit a race penalty in houses they purchase. Across all sales, the average race penalty falls by 40% after digitization. That is to say, before digitization, observably similar homes sell for different prices based on the seller's race. After digitization, observably similar houses sell closer to parity. However, this decline in the race penalty does not occur in counties without algorithmic investor entry.

Importantly, human investors and owner-occupiers drive much of this reduction in market-level racial disparities. After digitization, the race penalty among owner-occupier purchases decreases from 5 percent to 3 percent. Among human investors, the race penalty falls from 7 percent to 1.5 percent. A variety of factors could play a role in this decline. First, algorithmic investors' presence may drive up house prices through competitive bidding, affecting final sale prices even in transactions they do not win. Second, transaction prices inform listing prices for new homes on the market; higher starting prices lead to higher sale prices for minority-owned homes, regardless of algorithmic investor participation. Given that owner-occupiers represent about 80 percent of the market, these indirect effects drive the overall reduction in racial pricing disparities. As a result of market forces, the aggregate impact of digitization is a 5 percent increase in average sale prices for homes owned by minorities, compared with a 1 percent increase for White-owned homes. These findings highlight how market interactions can amplify the impacts of algorithms in ways that firm-level analysis cannot capture.

Although one explanation for the increasing prices of minority homes might be human mistakes, another possibility is that algorithmic investors are simply overpaying for unobservably bad minority-owned houses. Algorithms do not see all aspects of house quality available to humans, potentially leading to adverse selection. To test for evidence of this, I compare two proxies for gross returns, analyze the embedding representation of house images, and conduct within-house analyses. Across all analyses, I find no evidence that houses algorithmic investors buy from minority homeowners, and pay more for, and find that, on average, these houses do not appear to be systematically worse than houses bought from White homeowners.

So far, my analysis focuses on the transaction prices of the sold properties. In my fourth set of results, I consider how these shifts might influence the valuation of unsold homes—assets that constitute a substantial proportion of wealth for the median household. My estimates suggest that digitization leads to a 6 percent appreciation in the average value of unsold minority-owned homes. This appreciation is considerable when viewed in relation to median household wealth: a 6 percent rise in property values corresponds to an increase of roughly 20 percent in family wealth for Blacks and 13 percent for Hispanics. (Bhutta et al., 2020).

Together, these findings highlight how markets can amplify the effects of algorithms. Although algorithmic investors make up only 10% of the investor market, their activity induces significant changes that impact those using algorithms and those not using algorithms alike. Competition and price effects lead to a reduction in racial disparities in property values, affecting homeowners across the market, and change the behavior of human investors and owner-occupiers. My findings are similar in spirit to Becker (1957), where competition penalizes and drives out firms with discriminatory views. The magnitude and patterns of these effects raise questions about how even low levels of algorithmic adoption could be reshaping other parts of the economy.

This paper contributes to a growing empirical literature on the impacts of access to algorithmic recommendations. Comparing human decision-makers' choices with predictive models has a long history (Dawes, 1971; Dawes, Faust and Meehl, 1989; Hastie and Dawes, 2001). Modern advances in ML, increased computing power, and data availability have renewed interest in these questions. I build on prior work that shows that algorithmic recommendations can lead to improvements ranging from better heart attack diagnosis to more efficient bail and hiring decisions.[3] Other work shows

[3]For example, see Autor and Scarborough (2008); Li, Raymond and Bergman (2020); Raghavan et al. (2020); Frankel (2021); The White House (2022); OECD (2023) for applications in the labor market, Einav, Jenkins and Levin (2013); Fuster et al. (2022); Gillis and Spiess (2019); Arnold, Dobbie and Yang (2018); Blattner and Nelson (2021) for consumer finance, Mullainathan and Obermeyer (2021); Obermeyer and Emanuel (2016); Kleinberg, Mullainathan and Raghavan (2016); Chouldechova et al. (2018); Abaluck et al. (2020); **?**); Mullainathan and Rambachan (2023) for examples in the criminal justice system, health care, among other areas. See Rambachan (2022); Kleinberg et al. (2017*a*, 2015) for issues comparing human and machine predictions.

that access to algorithms translates into improved productivity or efficiency.[4] However, not all studies find positive effects.[5]

There is limited work on the effects of ML-powered algorithms at the market level. Calvano et al. (2020); Clark et al. (2023); Calder-Wang and Kim (2023); Brown and MacKay (2023) focus on the impact of ML-powered pricing algorithms on collusive behavior and price levels. Other studies focus on the impacts of automated algorithmic trading on the liquidity and pricing efficiency of financial markets (Hendershott, Jones and Menkveld, 2011; Chaboud et al., 2014; Upson and Van Ness, 2017). I examine the market-level impacts of algorithmic prediction outside of financial market trading.

This paper is also closely related to a large literature on racial disparities in the housing market. Although institutional discrimination has declined over time, audit studies, surveys, and empirical work continue to find evidence consistent with racial discrimination in the housing market.[6] Racial disparities in house values contribute to the large and persistent racial wealth gap.

Initially, academics and policy makers hoped that the use of algorithms could help mitigate human biases. For example, Kleinberg et al. (2018) show that reliance on algorithms to grant bail could simultaneously reduce crime, jail populations, and racial disparities. However, there are many examples of algorithmic bias, or algorithms that disparately direct fewer opportunities or resources toward minorities.[7] This paper is the first to show the indirect effects of algorithms on racial bias that work via market competition.

Understanding the impact of investors on the housing market is an important policy question. For instance, in December 2023, Democrats introduced legislation in the House and Senate that would ban hedge fund ownership of single-family homes (Kaysen, 2023; Merkley and Smith, 2023). A growing interdisciplinary body of work has examined the impacts of large single-family investors

[4]See Brynjolfsson, Raymond and Li (2023) for the impacts of generative AI on productivity in customer service, Harris and Yellen (2023) for the impact of the adoption of predictive maintenance on repair costs in a trucking company. See Bubeck et al. (2023); Choi and Schwarcz (2023); Peng et al. (2023); Noy and Zhang (2023) for additional effects of AI access on productivity, writing, and test taking capabilities.

[5]For instance, Acemoglu et al. (2022) finds no detectable relationship between AI investments and firm performance, while Babina et al. (2022) finds a positive relationship.

[6]For example, see Elster and Zussman (2022); Perry, Rothwell and Harshbarger (2018); Perry (2021); Bayer et al. (2017); Kim (2000); Freddie Mac Economic & Housing Research (2021); Zhang and Leonard (2021); Kermani and Wong (2021); Lewis, Emerson and Klineberg (2011). See Cutler, Glaeser and Vigdor (1999) for a summary of centralized discrimination.

[7]See Smith (2021) for a summary of empirical work on algorithmic bias. See Rambachan and Roth (2019), Rambachan et al. (2020), Bakalar et al. (2021), Kleinberg, Mullainathan and Raghavan (2016) and Cowgill and Tucker (2019) for theoretical work.

in the United States and, more recently, Europe.[8] This paper speaks to policy discussions around algorithms in the housing market and access to public records.

1 The Economic Impacts of Algorithms

Cognitive biases and constraints hinder human judgment. While the human brain employs mental shortcuts, or heuristics, to manage complexity, these mechanisms can introduce systematic biases that compromise judgment quality (Tversky and Kahneman, 1974; Halford et al., 2005). Common biases include confirmation bias, where people selectively favor evidence supporting existing beliefs; anchoring effects, where initial information disproportionately shapes subsequent judgments; and recency bias, which leads to overweighting of recent events (Wason, 1960; Ebbinghaus, 1913; Tversky and Kahneman, 1974). People also systematically under-respond to new information, particularly negative feedback (Möbius et al., 2022). Beyond these general cognitive limitations, implicit biases based on race, gender, age, appearance, and language can unconsciously influence judgments (Wistrich and Rachlinski, 2017; Hamermesh and Biddle, 1994; Mobius and Rosenblat, 2006; Deprez-Sims and Morris, 2010). Decision quality further deteriorates under cognitive strain from factors like fatigue, stress, and resource scarcity (Mani et al., 2013; Levi et al., 2017).

These limitations of human judgment spurred early interest in comparing human and algorithmic decision-making, though technological constraints limited applicability. Early work, synthesized in Hastie and Dawes (2001), evaluated predictive accuracy across diverse domains and reached a striking conclusion: "expert judgments are rarely impressively accurate and virtually never better than a mechanical judgment rule." Focusing on prediction decisions, Tetlock (2005) studies political forecasters over decades, found that human forecasting ability was low, and that even simple statistical approaches outperformed both expert and non-expert human forecasters. However, the potential for algorithmic alternatives remained constrained by limited data and the available computational methods.

Advances in machine learning sparked rapid adoption of algorithmic tools and a renewed academic interest in algorithms. These computational methods automatically discover patterns in data and generate predictions without explicit programming (Google, n.d.). Modern techniques like deep learning, gradient boosting, and ensemble methods, powered by vast datasets and en-

[8] Fields (2018, 2022) examine how technology-driven "calculative agency" enabled the financialization of the single-family housing market. Raymond et al. (2016, 2018, 2021) study the impacts of institutional investors in Georgia and housing insecurity. Mills, Molloy and Zarutskie (2019) provides some empirical early-stage analysis of the activities of these firms. Gurun et al. (2023) study the increase in institutional investor ownership and the impacts of investor mergers on rent and neighborhood safety. Buchak et al. (2022) studies the "i-buyer" firms (e.g. Zillow, Offerpad, Redfin and Opendoor) and their impacts on liquidity in the housing market. Francke et al. (2023) examine the impact of a ban on large institutional buyers of housing in the Netherlands.

hanced computing power, now consistently outperform human experts across domains from medical diagnosis to financial prediction (McKinney et al., 2020; Silver et al., 2016; Barboza, Kimura and Altman, 2017). These performance gains have driven widespread adoption, with 62 percent of IT professionals reporting significant or moderate increases in AI investments (CompTIA, 2024).

As machine learning tools have become widespread, researchers have rigorously evaluated their performance in high-stakes settings, revealing both substantial improvements over human judgment and challenges in human-AI collaboration. In a landmark study of bail decisions, Kleinberg et al. (2017*b*) demonstrate that machine learning algorithms can simultaneously improve accuracy and increase equity in predicting criminal recidivism. Their econometric approach addresses two key challenges previously overlooked: humans may optimize different objectives than the algorithm (omitted-payoffs), and they may have access to additional information during in-person bail hearings (private information). Building on these insights, researchers have extended this analytical framework to evaluate human and algorithmic decisions across domains including labor markets, healthcare, and public policy (Li, Raymond and Bergman, 2020; Mullainathan and Obermeyer, 2022; Athey, 2017; Rambachan, 2022; Brynjolfsson, Raymond and Li, 2023).

Despite these advances, significant challenges in human-AI collaboration persist. A meta-analysis of over 100 experimental studies finds that human-AI collaborations typically underperform both AI systems alone and expert human decision-makers (Vaccaro, Almaatouq and Malone, 2024). Challenges to human-computer interactions including human resistance to algorithmic recommendations (algorithmic aversion) in some contexts, while also being too willing to adopt the AI recommendations in others, contribute to this underperformance Helander (2014); Agarwal et al. (n.d.); Dietvorst, Simmons and Massey (2014); Steyvers and Kumar (2024).

Yet, even when algorithms can improve decision quality, their wider impacts remain poorly understood. The advent of algorithmic predictions could alter competitive dynamics or enable collusion across firms (Clark et al., 2023; Calder-Wang and Kim, 2023; Fish, Gonczarowski and Shorrer, 2024). Evidence from the German retail gasoline market shows that stations adopting algorithmic pricing increased their average margins by 28 percent, suggesting algorithmic coordination (Clark et al., 2023). Moreover, algorithmic improvements may not enhance overall welfare; in financial markets, Budish, Cramton and Shim (2015) demonstrate how individual algorithmic trading advancements can spark an arms race that merely raises barriers to market participation without improving market efficiency. Beyond market dynamics, algorithmic adoption introduces organizational frictions, creates data management challenges, and can lead to correlated decision-making across firms (Saxena and Guha, 2024; Kleinberg and Raghavan, 2021; Lepri et al., 2016).

Given these conflicting forces, the overall economic impact of widespread adoption of algorithmic prediction remains uncertain.

2 Algorithms and Prediction in the Housing Market

2.1 Real Estate Investment

2.1.1 Single-family Homes

I study the market effects of algorithms in the US housing market. Residential real estate, properties intended for living accommodations, is the largest asset class in the United States, with a total value of $43 trillion (Malone, 2023). Single-family detached houses comprise 86 percent of the value of all residential real estate and 66 percent of the entire housing stock (Malone, 2023; Neal, Goodman and Young, 2020). In my sample, single-family houses make up 66 percent of the occupied housing in urban areas and 72 percent in rural areas (U.S. Census Bureau, 2021). Single-family houses are purchased by two types of buyers: owner-occupiers, who buy houses to live in, and investors, who buy for rental income or to flip for profit.

Although the majority of single-family homes are owned by owner-occupiers, but the balance are occupied by renters and make up a particularly important part of the rental market. Single-family homes make up the largest single segment of rental housing (41 percent) and are more prevalent in areas that are less urban or less affluent (Census, 2023; Neal, Goodman and Young, 2020; Freddie Mac Economic & Housing Research, 2018).[9]

2.1.2 Investing is a Prediction Problem

Investing is fundamentally about predicting future income, expenses, and property values. Real estate investors must estimate the potential value of an asset alongside possible rental income and costs like upgrades, repairs, and ongoing maintenance. To determine a home's value, they examine its physical condition–structural integrity, kitchen and bathroom fixtures, electrical and plumbing systems, and landscaping. Accurate square footage and layout configuration, such as whether a non-conforming bedroom is included, are also critical. Investors also evaluate how efficiently the layout uses space and the condition of neighboring properties. [10]

[9] In the US, single-family homes are detached dwellings built to be occupied by one household on their own plot of land.

[10] A key metric in real estate for comparing properties is the capitalization rate, which is calculated as net income divided by asset value. Appendix Figure A.1 shows an example of capitalization rate information provided for a multifamily property.

2.1.3 The Human Informational Advantage and "Mom and Pop" Investor

Local entrepreneurs are best positioned to solve this prediction problem. Their intimate knowledge of the market led to a widespread belief that these "mom-and-pop" investors would maintain permanent dominance in the single-family home sector (Fields, 2018; American Homes 4 Rent, 2013, 2018). This presumed advantage stems from their deep local expertise and the importance of non-quantifiable information: they understand neighborhood traffic patterns, can assess neighborhood amenities like parks, recognize emerging gentrification trends, and track the status of local businesses and industrial facilities. Moreover, these investors frequently bring relevant professional experience in construction and real estate, enabling them to accurately estimate renovation costs and timelines.

In theory, a single company could employ a large staff to evaluate and acquire property. However, the substantial costs of individually assessing scattered, structurally diverse single-family homes presented a significant barrier (Amherst, 2016). This challenge was exemplified by Redbrick Partners' attempt in the early 2000's to build a large single-family rental portfolio with a large staff of human investors. Despite the favorable conditions of a rapidly appreciating housing market, the firm managed to acquire only 1,000 homes over four years, ultimately finding their approach unsustainable without technological solutions (Mills, Molloy and Zarutskie, 2019). These operational challenges led to the company's eventual closure (Fields, 2018).

2.1.4 Investing is a Challenging Cognitive Task

Although human investors have access to huge amounts of information about a particular property, synthesizing the qualitative and quantitative data into a precise valuation is challenging. For example, how would you weigh the relative value of a one-car versus two-car garage or access to an attractive local park? Is it worth replacing outdated bathroom fixtures and cutting down overgrown trees? The complexity and mental load of complex decisions can exceed human cognitive capabilities, leading to oversimplification and reliance on heuristics. Human cognitive limitations can hinder accuracy. This tendency toward heuristic-based decision making is evident in other domains; for instance, Mullainathan and Obermeyer (2021) demonstrate how physicians often default to simplified diagnostic models that overemphasize obvious symptoms like chest pain when predicting heart attacks. Real estate investors face similar constraints. Unlike algorithmic approaches that can learn from vast datasets, individual investors are also constrained to learn from their relatively limited personal experience.

Investing in houses is challenging, and human investors expend substantial effort to develop structured decision processes that minimize costly errors. The stakes are significant; Redfin estimates that investors lose money on one in seven homes (Redfin, 2023*a*). To mitigate cognitive biases and avoid over-reliance on intuition, investors commonly employ a "buy box" framework to focus on properties where they possess clear valuation advantages.[11] A typical buy box might target specific parameters – for instance, two or three-bedroom houses in select Fresno, California zip codes that appeal to middle-class buyers from diverse employment sectors. Complementing this approach, investors also often maintain exclusion criteria that *eliminates* properties with high-risk characteristics. These might include houses requiring repairs to critical systems like electrical, roofing, or septic, all of which present challenging cost forecasting scenarios. These systematic approaches help investors avoid common psychological traps, such as pursuing properties that intuitively "feel like a great deal" or those with aesthetic appeal but hidden structural deficiencies.

2.1.5 Using Algorithms to Value Houses

The complexity of real estate valuation makes it an ideal candidate for machine learning (ML) algorithms, which can detect statistical patterns that humans might miss. These algorithms can process an extensive array of variables about properties, neighborhoods, identifying complex interactions while optimizing for out-of-sample prediction accuracy. The algorithms analyze diverse data sources including population demographics, homeownership rates, vacancy patterns, income levels, crime statistics, school rankings, recent transactions, construction specifications, maintenance requirements, and employment trends, among others (Amherst, 2016; Invitation Homes, 2017). Although building these technology platforms is expensive and requires specialized teams of data scientists and software engineers, algorithmic investors emphasize their necessity: "[w]ithout using technology to filter and deliver automated valuations... it would be extremely time-consuming and inefficient to review and bid on these properties... The entire process uses a vast amount of data that is impossible to distill into actionable information without the use of technology" (Amherst, 2016; Christophers, 2023).

The algorithms guide acquisition decisions through a hybrid human-machine process, filtering available properties and estimating potential returns to create a queue of promising investments for human review. Acquisitions teams, who build and monitor the algorithmic recommendations, often based in financial centers like New York, California, and Texas, determine how to incorporate the algorithmic recommendations and making an offer.[12] This approach approach enables making offers on large numbers of properties across wide geographic areas, based on statistical relation-

[11]This is also an algorithm, but not a ML algorithm. For example, see New Investors Must Start with a Buy Box or they are wasting time and money.

[12]Many algorithmic firms employ their own internal real estate agents to make offers on properties. Offers are made primarily by real estate agents to homeowners.

ships in the data. For instance, Amherst Residential reports that its algorithm, Amherst Explorer, automatically evaluates approximately five hundred new listings daily within its target markets, calculating automated valuations by estimating potential rents, refurbishing costs, taxes, insurance, and other expenses to determine net operating income. Each morning, before the acquisition team arrives at work, the algorithm has already generated a list of targeted properties with projected returns Amherst (2016).[13] However, while these algorithms leverage vast amounts of quantified data, they may miss insights available to traditional investors with deep neighborhood expertise. I will formalize this trade-off more explicitly in Section 5.1.

2.2 County Housing Records

Algorithms depend on housing data produced by county governments. In this section, I describe county records, the digitization process, and the impacts of digitization on investors in the housing market.

2.2.1 The Process of County Record Digitization

County governments' records are the most accurate and up-to-date sources of housing market activity and characteristics of the housing stock.[14] These records, which were kept in paper books or on microfilm, are used in such day-to-day activities as dividing property in a divorce proceeding, building and engineering planning, genealogy research, and verifying property ownership. However, paper and microfilm records are expensive to maintain, susceptible to physical damage, and difficult to search. In 2009, spurred by the "Open Government" movement efforts to promote digital and transparent government, many counties began to digitize. Counties scanned records into searchable databases that could be accessed from the Internet (The White House, 2009).[15] Panel A of Figure 1 shows the share of counties with publicly available digitized records. Across the four states I study, the share of counties with digitized records doubled from 40 percent to 80 percent over the next five years.

First, each state had to ensure that county recorders could legally store their records digitally.[16] Then, each county needed to allocate funding. Digitization required scanning and indexing each

[13] According to its IPO prospectus, Invitation Homes, one of the largest single-family investors, analyzed about a million homes to assemble its portfolio of 50,000 properties.

[14] By law, County governments are responsible for maintaining public records of property; the Recorder's office maintains and preserves all legal documents affecting title to real property, and the Assessor's office determines the value of real property to collect property taxes. Deed records are public records that date back to county founding; some land records date back to the 1600's.

[15] Because this information is public data, digitizing these records also required making them accessible online.

[16] I use the year county recorder deed records are first available. In practice, property characteristics data also generally become available at this time.

paper or microfilm record, a time-consuming and expensive process. Next, each county needed to construct a software databases that could be linked to their websites that was integrated into common standards for county computer systems.

The timeline for completing digitization varied widely across counties. Panel B of Figure 1 illustrates the share of counties with publicly accessible digitized record systems by state over time. In most cases, state-level coordination led to sharp increases in the share of digitized counties within a state. However, the specific year each county completed digitization varied due to unforeseen challenges, such as database setup issues, record digitization delays, or funding constraints, resulting in idiosyncratic variation

2.2.2 How County Digitization Changes the Housing Market

Digitization affects the housing market through three channels: providing real-time data availability, supplying historical transaction data for training machine learning algorithms, and offering detailed information on the characteristics of each house. Once a county transitions to digital records, all new housing sales become immediately available online. This real-time, reliable, and accurate information is vital in enabling algorithms to update promptly, learning which houses are on sale and which have recently sold. Investors and county officials consistently highlighted the real-time availability of digital data as the most impactful change, reshaping how they access and utilize market information.[17] Digitization not only simplifies the process of downloading historical transaction data but also provides it in a reasonably clean format, offering free access that eliminates the cost barriers.

To predict the value of a house, algorithms require a digital representations of such characteristics as the number of bedrooms, bathrooms, and levels. When a house is in the database, digital records of the characteristics of the house are readily available and easy to assess algorithmically. If the house has not yet been added, investors would need to manually collect these data to estimate the value, making it harder to value these houses with an algorithm.[18] I leverage the bureaucratic variation in when each house was digitized to perform robustness checks and estimate house-level effects.

[17] MLS data and Zillow data are considered unreliable because they depend on accurate data entry from real estate agents and are generally not updated in real time. Private data providers, especially in the early 2010s, either did not have data or failed to provide real-time data updates.

[18] Collecting this data by hand is possible, but significantly more costly.

3 Data and Empirical Strategy

3.1 Empirical Strategy: Digitization

My analysis uses a difference-in-difference (DiD) analysis. I use a dynamic event study with differential timing to isolate the causal impact of digitization on the entry of algorithmic investors and on market-level and house-level outcomes:

$$y_{ct} = \delta_t + \alpha_c + \sum_{j \neq -1}^{J} \beta^j \times \mathbb{1}[t = j] \times D_{ct} + \gamma X_{ct} + \epsilon_{ct} \tag{1}$$

The outcome variables y_{ct} capture the results for county c and year t. First, I examine the impact of digitization on algorithmic investor entry. The outcome is $y_{ct} = \frac{q_{ct}^{algo}}{q_{ct}}$), the number of houses purchased by algorithmic investors (q_{ct}^{algo}) in county c and year t over the total number of houses sold. I estimate the impact of digitization on price using $y_{ct} = ln(price_{ct})$ or the natural log of the county average sale price of houses in year t. D_{ct} is an indicator equal to one if county c has digitized in year t and 0 otherwise. Digitization is an absorbing state; once a county builds a database system, it does not return to paper records. Counties that did not been digitized by 2017 are used as controls. All regressions include year fixed effects (δ_t) to account for factors that vary over time such as interest rates, housing market policy, and other macroeconomic variables. I also account for time-invariant factors specific to each county, such as size, income levels, and geography (α_c). Standard errors are clustered at the county level. The β^j vector is the parameter of interest that captures the time-varying treatment effect of digitization. At the county level, I weight the regressions based on the number of property transactions in each county-year.

I use a series of dynamic differences in difference estimators that are robust to the effects of digitization varying over time. The treatment effects of digitization could increase over time as algorithms may become more accurate and organizational processes are established. On the other hand, the effects of treatment could also decrease as competition in the housing market intensifies. To address time-varying treatment effects, I use the Sun and Abraham (2021) interaction weighted estimator (IW) that is robust to the correlation over time and across adoption cohorts. I also present results using a series of additional robust estimators introduced by de Chaisemartin and D'Haultfœuille (2020), Borusyak, Jaravel and Spiess (2022), Callaway and Sant'Anna (2021) as well as using traditional two-way fixed effects regression analysis. In general, estimates from robust estimators are larger and more stable because they avoid comparisons between already-treated counties.

These estimators require three assumptions: no anticipation, no spillovers between treated and not-yet-treated counties, and parallel trends. First, participants should not change their behavior in anticipation of future treatment. Second, digitization in one county should not impact the housing market in a county that has not yet been digitized. Third, in the absence of treatment, the treatment and control groups would have evolved similarly.[19] For example, there should be no changes in county economic policy that differentially impact treatment and controls. In Section 4.2, I examine the robustness to a series of alternative explanations.

In Figure 1, I plot the share of counties with accessible and digitized county Recorder databases over time. The sharp nature of digitization patterns is important to my empirical strategy. The discrete change in digitization will generate discrete changes in algorithm availability, while other unobservables should evolve smoothly around the threshold.

I also estimate a series of cross-sectional hedonic regressions at the house level. This complements the county-level analysis and allows me to explore the impact of house-level digitization (D_{ict}) on house-level outcomes, accounting for differences in observable house characteristics. I examine the likelihood that an algorithmic investor purchases an available house, denoted $\mathbb{1}[q_{ict}^{algo} = 1]$, and a natural logarithmic transformation of the sale price. At the house level, algorithmic purchase could be correlated with unobserved aspects of the house, the number of bidders, time on the market, or the tech-savviness of the listing real estate agent. To address this, I also perform a two-stage least squares (2SLS) regression, where purchase by an algorithmic investor is instrumented with digitization (Angrist, Imbens and Rubin, 1996).

$$\mathbb{1}[q_{ict}^{algo} = 1] = \delta_t + \alpha_g + \beta D_{ict} + \gamma X_{ict} + \epsilon_{ict} \tag{2}$$

The second stage of the relevant house-level regression, run using 2SLS to obtain correct standard errors, is:

$$y_{ict} = \delta_t + \alpha_g + \beta \times \mathbb{1}[\widehat{q_{ict}^{algo} = 1}] + \gamma X_{ict} + \epsilon_{ict} \tag{3}$$

3.2 Data

My sample includes data from 400 counties in Georgia, North Carolina, South Carolina, and Tennessee, spanning the period between 2009 and 2021. Information on property records comes from the county governments. I use detailed property-level house characteristics and sales information from ATTOM Data and Zillow. I also rely on aggregated rental and listing data from Zillow and demographic and socioeconomic data from the US Census.

[19] In another way of saying the same thing, the timing of digitization is not correlated with first stage or reduced form outcomes.

3.2.1 Digitization Data

I hand-collected data on county record digitization from county recorders' offices, the Internet Archive, and ATTOM Data. The primary source of information was direct interviews with county officials. County officials provided the year when their transaction records first became publicly available online. Once counties switched to electronic records, all future property transactions were automatically digitized, and information on recent transactions became immediately available online. Database systems also enabled easy download of historical data and house information.

I supplemented these interviews with snapshots of county websites from the Internet Archive. These snapshots verify when the county websites first provide remote access to the county records. Counties did not keep systematic records when each house record was digitized. Instead, I collect this information from ATTOM Data, which tracked when each record was added. I discuss further details on digitization in Sections 2.2 and 2.2.2.

A central concern with hand-collected data is the potential for measurement error. To address this, I use the digitization year provided by ATTOM to corroborate the county information. Although these two series do not align perfectly—-since houses are not all digitized at once and new houses are continually added—the two are similar. To validate the year of digitization of the ATTOM house record, I compared the year provided by ATTOM with the year of digitization from a subset of Georgia counties that maintained more detailed records of house-level digitization. Since these records are no longer updated, I collected copies of information stored by the Internet Archive. Once again, the ATTOM Data year of digitization closely corresponds to county records.

3.2.2 Identifying Investors

Investors are corporate entities that buy houses to rent out or resell (Redfin, 2023*b*). I exclude government entities, banks, credit unions, timeshare operators, securitized mortgage trusts, homeowner associations, churches, corporate relocation services, hotels, vacation rentals, farms, builders, and property owner associations. This definition follows other work on investors in the single-family market (Redfin, 2023*b*; Mills, Molloy and Zarutskie, 2019).

After identifying all investors, I hand-classify each firm as human or algorithmic. I identify algorithmic investors and their properties using business registration information, public filings, and personnel records. I start with properties owned by corporate entities and identify corporate mailing addresses (Gurun et al., 2023; Mills, Molloy and Zarutskie, 2019). To match subsidiaries to the parent firm, I perform two rounds of fuzzy clustering, first on the mailing address and then on public business registration data, properties listed on landlord websites, and known lists of corporate subsidiaries from SEC filings. After this two-round matching procedure, I determine whether each firm's investment strategy is algorithmic using SEC filings, news articles and interviews, company websites, and personnel records. If companies use an algorithmic acquisition engine or automated valuation platform, or employ a data science or software engineering team, I code them as algorithmic. Consistent with previous studies, I find about 40 algorithmic investors in my data, which own about 130,000 houses (Mills, Molloy and Zarutskie, 2019; Gurun et al., 2023).[20] Although not all companies using algorithmic valuation conduct interviews or file with the SEC, all of them do have business registration data, websites, and personnel records available on LinkedIn.[21]

I identify human investors as those using non-algorithmic acquisition strategies. I rely on news articles, interviews, company websites, and personnel records to determine whether a firm relies primarily on human judgment to evaluate investments. However, most of the human investors have websites or personnel records available on LinkedIn, and all have business registration data. Due to the time-intensive nature of this search process, I only hand-classify firms with at least 80 purchases in my sample. Of the entities with less than 80 purchases over the decade in my sample, of those that are not categorized as algorithmic, I assume that these are investors using human judgment.

3.2.3 Housing Market Data

Residential housing market data comes from ATTOM Data and Zillow's ZTRAX database and includes about 8.3 million housing transactions in North Carolina, South Carolina, Georgia, and Tennessee from January 2009 to December 2021. Both sources draw on county recorder offices and county property tax assessor records. The recorder office data include detailed property transaction information, such as sale price, date, identities of buyers and sellers, the corporate structure of the buyer or seller, any relationship between the two, and indicators for arms-length transactions and sales of newly constructed houses.

Tax assessor records provide additional property details, including property type, geographic coordinates (latitude and longitude), year built, architectural style, number of bedrooms and bathrooms, air conditioning type, roof construction material, and historical estimates of market value, land, and improvements. As of 2023, all housing records in this sample of counties dating back to the early 2000's, have been digitized, enabling historical analysis. I exclude non-arms-length and multi-parcel transactions and geocode each house using latitude and longitude to census-defined geographies, including county, tract, block group, and block. While not all houses can be geocoded to the census block level, all are geocodable to the block group level.

[20] There are a series of consolidations between the algorithmic investors in the dataset such that at the end of the sample, the total number of firms is smaller.

[21] Only algorithmic investors that are publicly traded REITS or involved sale of securities to investors, must submit SEC filings.

To supplement this transaction data, I incorporate zip and county-level measures of housing market dynamics from Zillow. These estimates , include measures such as the average sale price-to-list price ratio, the share of listings with price cuts, median sale prices, and the share of sales over the list price.

Additionally, I scrape interior and exterior house images from Zillow and investor websites, yielding images for a subset of 50,000 houses. These images are processed into vector embeddings for analysis using a deep learning model, described in Section 5.6.3.

Finally, I integrate socioeconomic and demographic variables from the American Community Survey (ACS) and the 2010 and 2020 Decennial Census. Because many counties in my sample have populations below the 65,000 threshold required for the one-year ACS estimates, I rely on five-year ACS data estimates. These data include variables such as median income, median age, racial composition, education levels, the fraction of the population that is rent burdened, median rent, household size, labor force participation, and unemployment rates at the county, census tract, block group, and block levels.

3.2.4 Identifying Homeowner Race

I use the Bayesian Improved Surname Geocoding (BISG) proxy method to infer race and ethnicity from publicly available homeowner names. The BISG model predicts race and ethnicity based on owners' surnames and census block addresses using Bayes' theorem. This approach is widely adopted in fair lending analysis (Elliott et al., 2009). The Consumer Financial Protection Bureau, which uses this algorithm for fair lending analysis, has conducted BISG validation tests in mortgage lending, a setting that closely mirrors my own (Consumer Financial Protection Bureau, 2023). Using census block geocoding, BISG exhibits Area Under the Curve (AUC) scores of 0.94 or higher across classifications, including Hispanic, Black, non-Hispanic White, and Asian borrowers. These validation find suggests the model can accurately categorize races and ethnicity from geography and surname information.[22]

3.3 Summary Statistics

I present summary statistics on the houses purchased by owner-occupiers, human investors, and algorithmic investors.

[22] AUC scores range from 0 to 1 and represent the model's classification accuracy. A score of 0.5 indicates that the model performs no better than random guessing, while 1 indicates perfect classification.

3.3.1 House-Level Descriptive Statistics

Owner-occupiers make up the bulk of the market. Owner occupiers buy 86 percent of all houses as shown in column 1 of Table 1. The average house they buy is 30 years old; has 2.12-bedrooms, 2.14-bathrooms, a garage, a parking space, and a fireplace; and sells for $194,270.

Human investors purchase on average less expensive, older homes. However, these houses are not significantly different from the overall population of sold houses. Column 2 of Table 1 shows that human investors pay an average of $127,755 for homes that are slightly smaller and less likely to have a garage and a parking space.

Algorithmic investors tend to purchase newer, larger, and more expensive homes. As shown in column 3 of Table 1, their average house is 21 years old and sells for $219,130. It has 2.76 bedrooms, 2.47 bathrooms, and almost always at least a parking space. It was last remodeled 18 years before.

The most striking difference between human and algorithmic investors is the very low variation in characteristics of houses purchased by algorithms and the very large standard deviations among houses purchased by human investors. The standard deviations on all house characteristics in column 3 of Table 1 are much smaller than column 2. These differences are illustrated even more clearly in Panels A through D of Appendix Figure A.5. The distribution of houses purchased by algorithmic investors, relative to human investors, is much more concentrated in terms of bedrooms, bathrooms, age, and sale price. I will return to this in more detail in Section 5.

Table 2 shows the county-level characteristics of the houses purchased by human and algorithmic investors. Algorithmic investors are active in slightly larger and wealthier counties with a higher Hispanic population. Otherwise, the characteristics of the counties are relatively similar.

3.3.2 Firm-Level Descriptive Statistics

Compared to human investors, algorithmic firms are significantly larger and operate across much wider geographic areas. As shown in Panel E of Appendix Figure A.5, the market prior to digitization is dominated by numerous small firms. On average, human investors participating in the market purchase just one house each. In contrast, algorithmic firms purchase an average of 2,000 houses annually. Consequently, digitization leads to a significant increase in the scale of the largest firms in the market. Panel F of Appendix Figure A.5 highlights the geographic breadth of algorithmic investors, who are active in nearly 300 different zip codes annually. Notably, less than 5 percent of their purchases occur in the same zip code as their corporate mailing address. By comparison, 40 percent of houses purchased by human investors are located in the same zip code as their corporate address, illustrating the localized nature of human investment versus the broader reach of algorithmic firms.

4 Digitization Leads to Algorithmic Investor Entry

4.1 County Digitization and Entry

The raw data clearly demonstrates the impact of digitization on the buying behavior of algorithmic investors. Panel A of Figure 2 shows the natural log transformation of the number of houses purchased by algorithmic investors in each county, by time to digitization. Panel B of Figure 2 presents the analysis of the accompanying event study that shows similar large and persistent increases in the share of homes bought by algorithmic investors. This increase persists and remains stable until the end of the sample period. Panel B shows that digitization is associated with a 2 percentage point increase in market share of algorithmic investors. All regressions are adjusted for county- and year-fixed effects and weighted by the number of transactions and control for county minority share and population. Standard errors are clustered at the county level.

Alternative estimators show similar results. In Appendix Figure A.3, I show the results are similar using alternative event study estimators: Borusyak, Jaravel and Spiess (2022), Sun and Abraham (2021), de Chaisemartin and D'Haultfœuille (2020) and the traditional fixed two-way effects model. Robust estimators avoid comparing newly treated units with already treated units, thus delivering larger and more stable estimates than the two-way fixed effects model.[23]

Table 3 presents the corresponding DiD estimates on the natural logarithm of the number of houses purchased by algorithmic investors. Across estimates, I find digitization leads to large increases in the number of houses purchased by algorithmic investors. The Callaway and Sant'Anna (2021) estimates of a 100 log point increase are lower because this estimator cannot be weighted by county size. Taken together, I interpret these results to suggest that county digitization and the subsequent increase in data availability, on average, lead to a sharp and sustained increase in home purchases by algorithmic investors.

The timing of county digitization is not strongly related to observable county characteristics. Table A.1 shows that early and late digitizing counties are balanced in unemployment, income, other demographics, rent, and vacancy rates. Early digitizing counties are larger and have a higher Hispanic population than late digitizing counties, but are otherwise similar in socioeconomic and demographics. Appendix Table A.4 shows how estimates from the standard DiD vary with additional controls to county size and minority share. Column 1 shows that, controlling for county- and year-fixed effects, digitization increases the number of houses purchased by algorithmic investors by 113 log points, controlling for county size, minority share and county- and year-fixed effects.

[23] Callaway and Sant'Anna (2021) cannot be weighted with the number of transactions, so I only plot the other estimators.

Column 2 shows how estimates vary with additional controls for pre-digitization county socioeconomic status, including demographics, poverty, unemployment, share with young children, and educational attainment. In column 3, I add controls for the pre-digitization number of housing units and rent burden. In general, the estimates fall slightly, but remain stable. I interpret these results to suggest that my estimates of the impact of digitization are not driven by systematic differences in observables between counties.

I also see similar strong impacts at the house level. Column 1 of Table 4 shows that digitization results in a 17-fold increase in the likelihood that an algorithmic investor purchases a home compared with a non-digitized house in the same census tract controlling for house characteristics and tract- and year-fixed effects. Column 2 indicates a 16-fold increase compared with a non-digitized house in the same census block group. Column 3 demonstrates a 7-fold increase within the same census block. These results suggest that algorithmic investors are sensitive to the availability of digital information when valuing houses.

4.2 Within County Triple Difference and Falsification Tests

I next address if there are unobservable factors that affect both algorithmic investor activity and the timing of digitization. For instance, county officials might be working to attract business investment and modernize government processes. To investigate this, I leverage house-level variation in the cost of algorithmically valuing houses to test for evidence of unobserved shocks.

The timing of house-level digitization is not related to house or neighborhood attributes. Appendix Table A.2 reveals that the early- and late-digitized houses are evenly matched in features such as the number of bedrooms and bathrooms and the presence of a basement or other structures.[24] While houses that are digitized later show a more recent last sale date, newly constructed houses are also digitized later such that there are no substantial disparities. Appendix Table A.3 shows that houses are also similar on neighborhood characteristics.

To investigate common county shocks, I compare digitized and not-yet-digitized houses within the same county before and after digitization. Panel A of Figure 3 plots the raw data, showing the number of houses bought by algorithmic investors separately for digitized and non-digitized houses. Algorithmic investors mostly purchase houses that exist in the county's database. These investors buy very small numbers of non-digitized houses. These outcomes could potentially be attributed to measurement errors in county record keeping, misclassification of algorithmic investors, transactions

[24]For this analysis, "early digitized" refers to those digitized before the county's median digitization year, and "late digitized" were digitized after.

involving the purchase of multiple houses, or scenarios where algorithmic investors supplement county databases with additional data.

Panel B of Figure 3 displays the corresponding interaction-weighted event study at the county level separately for digitized and non-digitized houses. This panel illustrates that the increase in the number of homes purchased by algorithmic investors is mostly confined to houses with digital records. Unobserved county-level shocks, such as changes in housing or foreclosure policy, should impact all houses in a county, regardless of digitization status. County shocks are not consistent with an impact that is concentrated in digitized houses.

I perform a falsification exercise to assess if county digitization impacts non-digitized houses after adjusting for house-level characteristics, using the regression in Equation 4 run at the census block group level. In the equation, β^{Digit} captures the impact of market digitization on algorithmic investor purchases of digitized houses. $\beta^{NoDigit}$ measures the impacts on non-digitized houses. I also include controls for neighborhood and house characteristics.

$$\begin{aligned} 1[q_{igt}^{algo} = 1] = \delta_t + \alpha_g + \beta^{Digit} \times \mathbb{1}[HouseDigitized_{ict} = 1] \times CountyDigit_{ct} + \\ \beta^{NoDigit} \times \mathbb{1}[HouseDigitized_{ict} = 0] \times CountyDigit_{ct} + \gamma X_{igt} + \epsilon_{igt} \end{aligned} \tag{4}$$

Column 4 of Table 4 shows that county digitization does not impact non-digitized houses; the impact is solely on digitized houses. These results are not consistent with unobserved, neighborhood-level shocks driving investor activity.

However, suppose that the existence in the county database simplifies the house discovery process for *all* investors. Human investors should then also be more likely to buy digitized houses. I test this by examining the effect of house-level digitization on the propensity to purchase by individual investors in column 5 of Table 4. $\beta^{NoDigit} = 0.0017$ and $\beta^{Digit} = -0.0699$. Digitization reduces the probability of human investment purchase and has no impact on non-digitized houses. I interpret these results to show that digitization affects algorithmic investors differently than human investors.

Together, these results build confidence that digitization and changing data availability drive algorithmic activity. First, if algorithmic investors had some influence on the digitization process, early digitized houses might look different from those that are digitized later. Second, if algorithmic investors were not relying on algorithms to purchase houses, we would not expect their purchases to be so heavily concentrated in digitized houses. Third, if localized neighborhood shocks were driving our results, we would expect both digitized and non-digitized houses in the same area to be impacted in a similar manner. Lastly, I show that the impact of digitization on the likelihood of purchase is specific to algorithmic investors. Thus, the evidence suggests that algorithmic investment is indeed driven by changes in data availability due to digitization.

5 Allocation and Specialization

In this section, I consider the possibility that algorithms and humans have distinct comparative advantages in prediction problems. I begin with a conceptual framework that illustrates the trade-off between humans and machines.

5.1 Conceptual Framework

Houses are characterized by an observable X and an unobserved Z and a common value component Y. Although the underlying data is multidimensional, I will use two unidimensional variables $x(X) = E[Y|X]$ and $z(X,Z) = E[Y|X,Z] - E[Y|X]$ and $E[y|X,Z] = E[y|x,z] = x + z$.

Human investors generate a prediction using both x and z, but may be biased ($\delta(x,z) \geqslant 0$). Humans may be biased on some houses, but not on others, or may not make systematic errors. For instance, humans seem to overvalue houses with pools and air conditioning during warm weather (Busse et al., 2012).

$$h(x,z) = E[y|x,z] + \delta(x,z)$$

Machine learning algorithms use patterns in data to make predictions. Algorithms look for patterns in thousands of examples, while humans are limited to their own experience. Algorithms are not subject to the same cognitive limitations as humans. For example, algorithms can quantify the specific value of a two-car garage versus a one-car garage, which is likely outside the scope of most humans. Algorithms are affected by subjective and transitory factors such as warm weather, mood swings, and prejudices. However, an algorithm cannot see z.

$$m(x) = E[y|x]$$

For any given house, would an algorithm do better or would a human do better at predicting true y? Given a house with true value y, if $|E[y|x] - E[y|x,z]| >> 0$ or $|E[y|x] - y| >> 0$, then the human-accessible private information could be crucial. For instance, an algorithm may not be able to capture the aesthetic appeal of an architecturally complex house or of a beautiful view of neighboring farmland. On the other hand, proximity to a farm could have drawbacks not captured by an algorithm, such as loud mechanical noises, the smell of manure, the presence of pesticides, and rodent infestation. A human has the advantage of being able to walk through a house, estimating repairs and maintenance costs. However, if $\delta(x,z) > 0$, or human decision making is systematically biased, the value of an algorithm might outweigh the importance of private information.

$$m(x) - h(x,z) = \underbrace{E[y|x] - E[y|x,z]}_{\text{informational advantage, } \mu} - \underbrace{\delta(x,z)}_{\text{human bias, } \delta} \tag{5}$$

Equation 5 highlights the trade-off between human and algorithmic valuation. If private information z is important, the human informational advantage can outweigh human error. If $\delta(x,z) = 0$ or humans are not biased, humans will do better. If humans make mistakes, the benefits of an algorithm can outweigh the importance of private information.

5.2 Measuring House Predictability

Before showing how human investors respond to digitization, I categorize houses by their degree of algorithmic predictability. I construct this measure from commonly available observables. I refer to the difficulty of predicting a house from observables as *predictability*.

I use the Extreme Gradient Boosting (XGBoost) algorithm to predict the transaction price (Chen and Guestrin, 2016). Given the high-dimensionality of the data and significance of non-linear relationships, non-parametric models outperform linear models when modeling houses. For example, even a slight increase in square footage could have a significant impact on price in a densely populated neighborhood, while the same would not be true in a rural area. In these cases, non-parametric models, such as tree-based algorithms, are able to capture nuanced, nonlinear relationships, particularly among the many variables that can influence house pricing like location, size, design, age, and local amenities.

The XGBoost algorithm operates on a gradient boosting framework in which new models are generated to correct the errors of pre-existing ones. In essence, it creates a robust overall model by combining multiple weak models. Doing so improves the accuracy of the prediction according to the regularized objective shown in Equation 6. l is the differentiable convex loss function; T is the number of leaves in each tree; and w is the leaf weights (Chen and Guestrin, 2016). Intuitively, this objective function balances training loss $l(\hat{y_i}, y_i)$ with L1 regularization (γT) and L2 regularization ($\lambda||w||^2$) components, encouraging both simpler and more generalizable models.

$$\mathcal{L}(\phi) = \sum_i l(\hat{y_i}, y_i) + \sum_k \gamma T + \frac{1}{2}\lambda||w||^2 \tag{6}$$

The model is built using pre-digitization data for each county to exclude any impacts from algorithmic investors. I randomly split the data into a training set and a 25 percent held-out test set. Using the training data set, I perform a grid search through the XGBoost hyper-parameter space, using 3-fold cross validation with early stopping (Shen, Gao and Ma, 2022; LaValle, Branicky and Lindemann, 2004).

In panel A of Figure 4, I plot the predicted versus actual log price for the held-out sample. The average out-of-sample root mean squared error is 0.903. Thus, 40 percent of the houses in the test set are within 10 percent of the price. The same measures computed for Zillow's Zestimate, which incorporates demand information from user interactions with its website, reveal that 59 percent of houses are priced within 10 percent of the sales price in areas with Zillow coverage (Zillow, 2023).

For each house, I calculate the out-of-sample average—the difference between the actual and predicted price—to capture how easy or hard it is to predict each house. The variation in prediction error is enormous: For some houses, it is close to 50 percent, while for others it is close to zero. Houses in neighborhoods constructed by the same builder and in the same style are easier to model, but older houses are much less standardized because they were constructed before the introduction of modern building codes and in a wider variety of architectural styles. Other factors making prediction difficult are features of sentimental, aesthetic, or historical value; and proximity to a noisy highway or pungent agricultural property.

Further complicating algorithmic prediction are differences among counties in how they collect data. Counties vary in the frequency with which they update their housing records, the quality of their data control, and the thoroughness of the information they collect on each home or transaction. Together, the less informative or accurate the observable information, the more important human private information becomes.

5.3 Human Investor Shift Towards Hard-to-Predict Houses

Human investors purchase houses across the entire distribution of model error; in some instances, the predicted price is far lower than the actual price, while in other cases the predicted price is significantly higher. In Panel A of Figure 4, I show the predicted versus actual prices of the gradient-boosted tree model described in Section 4.2. In general, the model is best at predicting houses in the middle of the distribution. In Panel B of Figure 4, I plot the actual price versus the predicted price for houses in a held-out test set from 2012 and 2013. The houses colored in light blue will be purchased by human investors and the purple ones by algorithmic investors. Human investors purchase houses across the entire distribution of model error; in some instances, the predicted price is far lower than the actual price, while in other cases the predicted price significantly exceeds the actual price. While some of these may reflect poor human decision-making, on average, large differences between predicted and actual prices may reflect private information. Unlike human investors, algorithmic investors only purchase houses where the model-predicted price closely approximates the actual price. In other words, they buy houses where the scope for adverse selection is small.

In Figure 5, I show how human investors react to digitization. Human investors become 50 percent less likely to buy houses in the lowest decile of model error, where algorithms are most effective. They become almost twice as likely to buy houses that are most difficult to predict.[25] Human investors become less likely to purchase houses in deciles 1 to 7 and more likely to purchase houses in the top two deciles of average model error. These results are consistent with human investors specializing where human comparative advantage is highest.

5.4 Discontinuities: Data Errors, Zoning Rules, and Lead Paint

A testable implication of my conceptual framework is that characteristics that increase the importance of private information should limit algorithmic investor buying. I show three examples where institutional factors deter algorithmic investors, providing further evidence of human investors enjoying a comparative advantage because of algorithmic investors' dependence on quantifiable information.

5.4.1 Zoning Rules

Unusual zoning rules in Wilson County, Tennessee, make it difficult to discern the number of bedrooms. Panel A of Figure 4 shows a distinct group of houses in the county where the predicted model price is much higher than the actual price. To qualify as a bedroom under the zoning rules, a room must include a specific type of closet. As a result, tax assessor records list most houses as having zero bedrooms, although the "true" number of bedrooms is much higher. Without access to the actual number of bedrooms, algorithms tend to have large errors in predicting prices in the county. While algorithmic investment is prevalent in Nashville and other nearby areas, it is limited in Wilson County. By contrast, human investors are a much greater presence there.

5.4.2 Lead Paint

Houses constructed before lead paint was banned are more difficult to value algorithmically. In the early 1900s, lead was a commonly used additive in paint and other building materials. During the 1960s and 1970s, detailed studies on the effects of lead poisoning led to concerns about health effects in residential structures. The Consumer Product Safety Commission banned lead paint in residential construction in 1978 (The Department of Housing and Urban Development, 2023). Human investors can conduct physical inspections to determine whether lead is present in pre-1978

[25] In the pre period, likelihood of purchase by an human investor is .12.

houses and accurately forecast remediation costs. Uncertain about lead exposure in those older homes, algorithmic investors cannot predict potential renovation expenses.[26]

In panel A of Figure 6, I test for a discontinuity in the density of houses bought by algorithmic investors, using a local polynomial density estimator (Cattaneo, Jansson and Ma, 2018, 2019). As seen visually in Panel A, the null hypothesis of no discontinuity around 1979 is rejected, with a p-value 0.000. In Panel B, I plot the density of houses purchased by human investors. In this case, with a p-value of 0.295, the null hypothesis that the density shows no evidence of manipulation cannot be rejected. I interpret these results to suggest that algorithmic investors appear to respond to the imposition of lead paint, while human investors do not.

5.4.3 Data Errors

The sampled databases show nearly 220,000 houses as having 15 bedrooms or 15 bedrooms. While there may be a few supersized mansions, the vast majority of these houses reflect data entry errors. Such errors are especially confounding for algorithmic investors since the number of bedrooms and bathrooms is essential to pricing a home.[27] Panel C of Figure 6 shows the number of houses with data errors sold over time; the series is relatively spiky but without any clear trends, indicating that the availability of houses with data errors does not strongly vary over time. Panel D of Figure 6 shows the natural log transformation of the number of houses with data errors purchased by human and algorithmic investors. Algorithmic investors avoid purchasing houses with data errors, unlike human investors, who do not rely exclusively on databases.

In all of these instances, institutional details produce variation in the value of private information and create opportunities for human investors.

5.5 Algorithmic Investors Specialize in Minority-Owned Homes

After illustrating human comparative advantage and where human investors focus their efforts after digitization, I now turn to algorithmic investors. I first establish the existence and robustness of a race penalty, suggesting the possibility of human bias, and then show that algorithmic investors disproportionately buy minority-owned homes.

[26]Lead-paint remains the most significant source of lead exposure in the US because many houses were built before 1978 (US EPA, 2014). Any renovation, repair or painting project in a pre-1978 home can easily create dangerous lead dust, requiring special lead-safe contracting procedures and contractors (US EPA, 2013).

[27]In principle, they could collect this information manually, but algorithmic firms are not organizationally set up to do this.

5.6 Racial and Ethnic Price Differences in the Housing Market

Prior to the passage of the Fair Housing Act, race was explicitly taken into account when estimating the value of a house. For example, *The Valuation of Real Estate*, a widely used textbook in the mid-20th century for real estate appraisal, claimed that neighborhood decline inevitably results from occupation by "...the poorest, most incompetent, and least desirable groups in the city," and described how "... racial heritage and tendencies seem to be of paramount importance" in determining property values (Babcock, 1932; Wheaton, 2023). While it is now illegal to discriminate on the basis of race in renting, buying, financing, advertising of housing, racial disparities remain in house values. In Section 5.6.2, I examine racial disparities in my sample.

5.6.1 Evidence on Racial Disparities in House Prices

Racial disparities exist in house prices. Harris (1999) documents that moving from a neighborhood that is under 10 percent Black to one that is between 10 percent and 60 percent black is accompanied by a 2.3 percent drop in house value, accounting for house and neighborhood characteristics. Perry, Rothwell and Harshbarger (2018) estimate that comparable homes lose 23 percent of their value when moving from a census tract with no Black residents to one that is 50 percent Black. Elster and Zussman (2022) find that house prices decrease 2 to 3 percent after minorities move in.

These price differences could reflect preferences for neighborhood composition, neighborhood characteristics and racial biases. White homebuyers exhibit a strong bias against living in areas with Black and Hispanic neighbors; Minority home buyers tend to be willing to live in a variety of places, including majority White neighborhoods (Lewis, Emerson and Klineberg, 2011). Price differences could reflect omitted variables correlated with the race of the homeowner, such as differences in neighborhood amenities or house characteristics. For example, levels of pollution and noise are typically higher in minority neighborhoods (Casey et al., 2017; Tessum et al., 2021). However, this could also

Yet, neighborhood characteristics do not fully explain price disparities. Based on widespread anecdotal evidence, appraisers may value the same house differently depending on the race of the owner. After receiving a low appraisal, some minority homeowners "whitewash" their homes by removing family photos and asking a white friend to stand in for them. Doing so, they say, results in higher subsequent appraisals (Kamin, 2023; Lilien, 2023; Howell and Korver-Glenn, 2018). A very small audit study found that, on average, a White homeowner received a 7 percent higher appraisal than a minority couple for the same house (Lilien, 2023). In general, minority-owned homes are more likely to receive appraisal estimates below what a buyer has offered to pay, even when considering the characteristics of the house and the neighborhood (Freddie Mac Economic &

Housing Research, 2021; Perry, 2021; Howell and Korver-Glenn, 2018). In Section 5.6.2, I examine racial disparities in prices in my data.

5.6.2 The Race Penalty

Before county digitization, I calculate that, everything else being equal, minority homeowners receive a sales price 5 percent lower than their White counterparts. (Elster and Zussman, 2022; Perry, Rothwell and Harshbarger, 2018; Perry, 2021; Bayer et al., 2017; Kim, 2000; Freddie Mac Economic & Housing Research, 2021; Quillian, Lee and Honoré, 2020; Zhang and Leonard, 2021; Kermani and Wong, 2021; Lewis, Emerson and Klineberg, 2011). In Figure 7, I plot this race penalty in period *prior* to digitization, controlling for various levels of neighborhood characteristics and observable characteristics of the house. All specifications include year and geography fixed effects.[28]

The first bar in Figure 7 shows that minority homes sell at a 14 percent discount relative to White homes in the same county, adjusting for house characteristics. This gap drops to 7 percent when adding census tract fixed effects. The 50 percent decrease in the race penalty suggests considerable unobserved heterogeneity among houses in the same county.[29] At the census block group level, the implied race penalty is 5 percent. At the census block level, minority-owned homes sell for 3 percent or $4,700 less. All of these numbers are calculated in the years before digitization.

5.6.3 Deep Learning Image Analysis

Houses are structurally unique three-dimensional objects that derive their value from their size, the color of the paint, the landscape, the maintenance, and the cleanliness of the windows, among many other factors. Two houses in the same neighborhood may have completely different architectural styles or states of disrepair. (Pinto and Peter, 2021; Harris, 1999; Choi et al., 2019). Minority homeowners are less wealthy and may invest less in house maintenance and aesthetics (Harris, 1999). The race penalty could simply reflect these differences in house appearance.

I use a deep learning model to calculate the race penalty adjusting for house images using house images scraped from Zillow. Appendix Figure A.6 shows an example house image. Images are not available for all houses in my sample. I rely on images for a total of 50,000 houses. I use AutoGluon, a deep learning model designed for unstructured data such as images, to convert each exterior image into a high-dimensional embedding vector (Erickson et al., 2020). The position of each image within this vector space indicates its visual features or content, ensuring that similar

[28] Include year by geography fixed effects yields very similar results.

[29] In our sample, census tracts encompass $4,517$ people or $2,006$ housing units. Census block groups average around $1,610$ people in our sample or 716 housing units. A census block contains around 65 people. I used the 2010 population to calculate these averages.

images are close to each other in the embedding space. Adding these deep learning embeddings to the race gap regression will control for previously omitted variables such as the aesthetic features of the house and the yard.

5.6.4 Race Penalty with Deep Learning

Incorporating house exterior images does not significantly change the race penalty. Table 5 shows the race penalty coefficients, controlling for the quality and appearance of the house with deep learning-generated embeddings. These race penalty estimates are similar to the estimates from Section 5.6.2. For example, at the census block level, the race penalty is 2.1 percent with the image embeddings and 3.3 percent without images, including block by year fixed effects. The existence and persistence of this race penalty suggests more than differences in house quality. For instance, consistent with other evidence that individuals associate lower values with the same home when they perceive it to be owned by a minority, humans could be undervaluing minority-owned homes (Lilien, 2023).

5.6.5 Algorithmic Investors Buy Minority-Owned Homes

Algorithmic investors disproportionately buy minority-owned homes. As shown in Table 6, the impact of digitization of house records is twice as strong for minority homeowners than for White homeowners. In column 1 of Table 6, the impact of digitization on a minority-owned home is twice as large relative to a White-owned home in the same census tract or census block group. However, the impact of digitization is six times as large compared with a White-owned house in the same census block. These results suggest that algorithmic investors do not just focus on minority neighborhoods, but specifically target minority-owned houses.

These results are not simply driven by all investors targeting minority-owned homes. In Appendix Table A.5, I demonstrate the effect of digitization by homeowner race within a sample exclusively consisting of investor transactions (human and algorithm), excluding the owner-occupiers (those buying houses to live in). These regressions illustrate the impact of digitization on the likelihood of purchase by algorithmic investors compared with human investors. Column 1 of Appendix Table A.5 reveals that the effect of digitization on minority-owned homes is five times stronger than on White-owned homes. The impact of digitization is five times as large at the census block group level (column 2) and nine times larger at the census block level (column 3). These results suggest that algorithmic investors are disproportionately likely to buy minority-owned homes, even compared with human investors.

6 Prices, Spillovers and Racial Disparities

In this section, I explore the consequences for market-level prices and racial inequalities.

6.1 Digitization Shrinks the Race Penalty

First, I explore how digitization affects the race penalty. Panel A of Figure 8 plots the race penalty coefficient by time to digitization, including census block group controls.[30] In the year following digitization, the race penalty shrinks by half from 8 percent to 4 percent.

In Panel B of Figure 8, I investigate the mechanism behind this change. The first two bars in Panel B of Figure 8 plot the pre-digitization race penalty for purchases of owner-occupiers and human investors. Both pay 5 percent less for the observably similar house in the same census block group with a minority owner compared with one who is White. However, as shown in the blue bar, algorithmic investors, who appear after county digitization, do not exhibit any race penalty. Algorithmic investors pay the same price for an observably similar house regardless of the race of the homeowner.

Interestingly, digitization also reduces the race penalty among owner-occupiers and human investors. The fourth bar in Panel B of Figure 8 shows that, after digitization, owner occupiers pay only 3 percent less for minority-owned homes. The last bar in Panel B of Figure 8 plots the post-digitization race penalty with human investors. After digitization, human investors pay only 1.5 percent less for minority-owned homes. In Appendix Table A.6, I use a two stage least squares (2SLS) analysis is used to address endogeneity concerns around other factors related to bidding behavior that could drive these results. The results are much noisier, but qualitatively similar.

Figure 9 shows that digitization leads to a 5 percent increase in the prices of minority homes. Among White homeowners, it is possible that algorithms may not raise prices. If homeowners are willing to sell their homes at a discount in exchange for a prompt offer, could lead to a decline in prices. Instead, we also see an increase; digitization leads to a 1.5% increase in the average sale price of White-owned homes.

6.2 Adverse Selection or Human Error?

A natural question is whether algorithms are taking advantage of human mistakes or simply overpaying for unobservably worse homes. Adverse selection has been widely cited as a barrier to the use of algorithms in the housing market and has been widely discussed as the reason why Zillow, an algorithmic investor, decided to stop buying houses (Economist, 2021).

[30] Digitization varies at the county by year level, so we cannot include geography by year controls.

I disentangle adverse selection from human error with two complementary approaches. In the first, I calculate the gross margin on each house sold: the difference between the resale and transaction price. If algorithmic firms overvalue minority homes relative to White-owned homes, then the gross margin on minority homes should be lower compared with White-owned homes. I also calculate the gross margin with the estimates of the house market value from tax assessors. Unlike resale price, which is only available for resold homes, these estimates are available for all homes. However, these are estimates made by the human tax assessor rather than actual transaction prices.[31]

Using my two measures of gross margin, I estimate the following regression for house i bought in year t, resold/assessed in year r in census block c, including purchase year by census tract, block group or block and resale or assessment year fixed effects:[32]

$$log(price^{resale}_{irc}) - log(price^{sale}_{itc}) = \gamma X_{irtc} + \beta_1^{algo} \times SellerMinority_{itrc} + \epsilon_{itrc} \tag{7}$$

The coefficient β_1 indicates if the margin is systematically different for minority homeowners. If houses purchased from minority homeowners are adversely selected, the margin should be lower, or $\beta_1^{algo} < 0$. However, if the higher prices paid by algorithmic investors for minority homes reflect the true value, then $\beta_1^{algo} \approx 0$ or $\beta_1^{algo} > 0$.

I find no significant differences in the gross margin by race of the homeowner. Columns 1 through 3 of Table 7 include census tract, block group, and block by year fixed effects, respectively, among houses bought by algorithmic investors.[33] Columns 1 through 3 of Appendix Table A.7 show similar results using the assessment margin. Among houses purchased by algorithmic investors, the margin on minority-owned homes is not systematically different from that on White-owned homes. In column 4 of Table 7 and column 4 of Appendix Table A.7, I explore whether the resale margin differs for homes bought in neighborhoods with greater minority shares. In both cases, I find no strong relationship. These results suggest that the gross margin of the algorithmic investor does not vary with the composition of the neighborhood.

Next, I show that the gross margin on minority-owned investors is *higher* than that on White homes. If minority-owned homes are priced too low, the gross margin should be higher due to the discounted acquisition price. Columns 5 through 7 in Table 7 show that the gross margin on minority homes is 10 percent higher among purchases by human investors. In column 8, I explore whether the margin varies by neighborhood composition. The margin may be higher in neighborhoods that contain a higher share of minority residents, but the estimate is noisy. Column 8 of Appendix Table

[31]Note that if tax assessor evaluations are also biased, our results will be more conservative.

[32]The time between purchase and resale or assessment is a linear combination of purchase and resale year, so this would drop from any regression.

[33]Not all houses can be geocoded to a census Block level, but I show all three results.

A.7 shows that the assessment margin is 9 percent higher in neighborhoods with a higher share of minority residents.

If these results are due to racial preferences or heuristics, the differences may be more pronounced in neighborhoods with more minority residents, where humans may have more trouble accurately valuing houses or biases. These results suggest that the higher prices algorithmic investors pay for minority-owned homes are not driven by adverse selection and may, in fact, reflect algorithm comparative advantage in valuing houses where human biases, prejudices, or cognitive limitations may cloud judgment.

These results are also not consistent with adverse selection among algorithmic firms. If minority-owned houses are unobservably worse, humans should not be willing to pay more for these houses after digitization. Humans can access unobservable aspects of house quality that are not apparent to algorithms and should not be subject to the same adverse selection concerns.

6.3 Spillovers

Thus far, my analysis has focused on the prices of sold homes. Changes in the real estate market affect home values in general, a key driver of household wealth and credit worthiness (Guren et al., 2020). If algorithmic investors purchase minority-owned houses similar to other minority-owned houses not on the market, their activity could have large indirect impacts on minority homeowner house values and household wealth. Such a spillover effect, however, would not necessarily be the case if the minority-owned houses are predominantly located in majority White areas and are not structurally similar to other minority-owned houses.

Following Hirano, Imbens and Ridder (2003), using the estimate of the expected price of sold minority homes, $E[P|S=1]$, I write the inverse propensity weighted unsold minority-owned homes house price impact $E[P|U=1]$ as:

$$
\begin{aligned}
E[P|U=1] &= \sum_X p(X|U=1)E[P|U=1,X] \\
&= \sum_X \frac{p(U=1|X)p(X)}{p(U=1)} E[P|U=1,X] \\
&= \frac{1}{p(U=1)} \sum_X p(U=1|X)p(X)E[P|U=1,X]\frac{p(X|S=1)p(S=1)}{p(S=1|X)p(X)} \\
&= \frac{p(S=1)}{p(U=1)} \sum_X E[P|S=1,X]\frac{p(U=1|X)p(X|S=1)}{p(S=1|X)} \\
&= \frac{p(S=1)}{p(U=1)} E\left[\frac{p(U=1|X)}{p(S=1|X)}P|S=1\right] \qquad (8)
\end{aligned}
$$

Equation (8) says that I can recover the average impact on the price of minority-owned homes by reweighting the prices of sold minority-owned homes, using a ratio of propensity scores to account for differences in house characteristics. After re-weighting based on the characteristics of the observable houses and the census block, I find an average increase of 6 percent in the value of minority homes. It is important to emphasize that this analysis relies on a selection-on-observables assumption when reweighting. Although algorithmic investors may not be able to see unobservable characteristics, part of the impact comes from human investors and owner-occupiers, who have access to unobserved information. If unsold minority houses are very different on unobservables than sold minority houses, this estimate may overstate the impacts.

7 Conclusion

Progress in machine learning and the widespread availability of digitized data open up extensive economic possibilities. This work illustrates how the availability of algorithmic prediction not only influences individual decisions, but also precipitates a range of changes at the market level, affecting participation, firm organization, and equilibrium outcomes. In the housing market, the availability of machine-generated predictions leads to new entrants using algorithms to value houses. Human investors react by moving toward parts of the market where algorithms are least effective. Algorithmic investors buy disproportionately where human decisions are biased, causing large price increases. Six years after digitization, the race penalty disappears. In large part, these consequences stem from the indirect effects of algorithmic investors that manifest through the nature of competition. These findings suggest numerous avenues for future research.

First, when algorithms and humans disagree, we cannot assume that the algorithm is correct: Unobserved information can lead to algorithm errors. At the same time, we cannot assume that the human is always correct. Instead, the value of the trade-off depends on the importance of private information and the degree of bias in human decisions. A growing number of papers show that human errors can be sufficiently systematic to outweigh the value of private information (Kleinberg et al., 2017*a*; Mullainathan and Obermeyer, 2019; Rambachan, 2022; Kahneman, Sibony and Sunstein, 2021). More work is needed to better understand how the value of this trade-off varies across people and prediction problems.

Second, the machine learning tools used by companies employing algorithmic prediction strategies are rapidly changing. An ecosystem of companies is attempting to curate detailed and increasingly accurate datasets, from comprehensive house surveys that measure construction quality to mobile phone data that track neighborhood activities. As data quality improves, algorithmic investors may be able to target a larger percentage of houses.

Further, the efficacy of algorithmic prediction may depend on legal and institutional structures, which vary widely among states. This study examines Georgia, North Carolina, South Carolina, and Tennessee, where housing market transactions and prices are part of the public record. However, in a dozen states, property transaction prices are not automatically included in the public record, thus potentially undermining algorithmic prediction. More work is needed to explore how institutional structures affect the use of algorithms.

Finally, this study does not dwell on the potential implications of organizational differences between algorithmic and human investors. Algorithmic investors typically operate as large, formal, arms-length organizations, while human investors often manage their rental properties more informally. Unlike human investors, who frequently choose tenants personally or through their social networks, algorithmic firms may rely more heavily on automated screening procedures for tenant selection. Together, these changes could have lasting effects on the local labor market and communities. Given the rapid adoption of algorithms, these consequences merit further study.

References

Abaluck, Jason, Leila Agha, David C. Chan Jr, Daniel Singer, and Diana Zhu. 2020. "Fixing Misallocation with Guidelines: Awareness vs. Adherence." National Bureau of Economic Research.

Acemoglu, Daron, Gary Anderson, David Beede, Catherine Buffington, Eric Childress, Emin Dinlersoz, Lucia Foster, Nathan Goldschlag, John Haltiwanger, Zachary Kroff, Pascual Restrepo, and Nikolas Zolas. 2022. "Automation and the Workforce: A Firm-Level View from the 2019 Annual Business Survey."

Agarwal, Nikhil, Alex Moehring, Pranav Rajpurkar, and Tobias Salz. n.d.. "Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology."

American Homes 4 Rent. 2013. "Form S-11." *https://www.sec.gov/Archives/edgar/data/1562401/00011931251313*

American Homes 4 Rent. 2018. "10K-2017-Q4.Pdf." *https://s29.q4cdn.com/671712101/files/doc_financial/quar 2017-Q4.pdf.*

Amherst. 2016. "U.S. Single Family Rental - An Emerging Institutional Asset Class." *https://www.amherst.com/insights/u-s-single-family-rental-an-emerging-institutional-asset-class/.*

Angrist, Joshua D, Guido W Imbens, and Donald B Rubin. 1996. "Identification of causal effects using instrumental variables." *Journal of the American statistical Association*, 91(434): 444–455.

Arnold, David, Will Dobbie, and Crystal S Yang. 2018. "Racial Bias in Bail Decisions." *The Quarterly Journal of Economics*, qjy012.

Athey, Susan. 2017. "Beyond prediction: Using big data for policy problems." *Science*, 355(6324): 483–485.

Autor, David H, and David Scarborough. 2008. "Does job testing harm minority workers? Evidence from retail establishments." *The Quarterly Journal of Economics*, 123(1): 219–277.

Babcock, Frederick M. 1932. *The Valuation of Real Estate.* New York:McGraw Hill Book Company.

Babina, Tania, Anastassia Fedyk, Alex Xi He, and James Hodson. 2022. "Artificial Intelligence, Firm Growth, and Product Innovation."

Bakalar, Chloé, Renata Barreto, Stevie Bergman, Miranda Bogen, Bobbie Chern, Sam Corbett-Davies, Melissa Hall, Isabel Kloumann, Michelle Lam, Joaquin Quiñonero Candela, Manish Raghavan, Joshua Simons, Jonathan Tannen, Edmund Tong, Kate Vredenburgh, and Jiejing Zhao. 2021. "Fairness On The Ground: Applying Algorithmic Fairness Approaches to Production Systems."

Barboza, Flavio, Herbert Kimura, and Edward Altman. 2017. "Machine learning models and bankruptcy prediction." *Expert Systems with Applications*, 83: 405–417.

Bayer, Patrick, Marcus Casey, Fernando Ferreira, and Robert McMillan. 2017. "Racial and Ethnic Price Differentials in the Housing Market." *Journal of Urban Economics*, 102: 91–105.

Becker, Gary S. (Gary Stanley). 1957. *The Economics of Discrimination. Studies in Economics of the Economics Research Center of the University of Chicago*, Chicago:University of Chicago Press.

Bhutta, Neil, Andrew C. Chang, Lisa J. Dettling, and Joanne W. Hsu with assistance from Julia Hewitt. 2020. "Disparities in Wealth by Race and Ethnicity in the 2019 Survey of Consumer Finances." *FEDS Notes.*

Blattner, Laura, and Scott Nelson. 2021. "How Costly Is Noise? Data and Disparities in Consumer Credit." Comment: 86 pages, 17 figures.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2022. "Revisiting Event Study Designs: Robust and Efficient Estimation."

Brown, Zach Y., and Alexander MacKay. 2023. "Competition in Pricing Algorithms." *American Economic Journal: Microeconomics*, 15(2): 109–56.

Brynjolfsson, Erik, Lindsey Raymond, and Danielle Li. 2023. "Generative AI at Work." *NBER Working Paper No. 31161.*

Bubeck, Sebastien, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, et al. 2023. "Sparks of artificial general intelligence: Early experiments with gpt-4." *arXiv preprint arXiv:2303.12712.*

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2022. "Why Is Intermediating Houses so Difficult? Evidence from iBuyers."

Budish, Eric, Peter Cramton, and John Shim. 2015. "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response *." *The Quarterly Journal of Economics*, 130(4): 1547–1621.

Busse, Meghan R, Devin G Pope, Jaren C Pope, and Jorge Silva-Risso. 2012. "Projection Bias in the Car and Housing Markets." National Bureau of Economic Research Working Paper 18212.

Calder-Wang, Sophie, and Gi Heung Kim. 2023. "Coordinated vs Efficient Prices: The Impact of Algorithmic Pricing on Multifamily Rental Markets."

Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with multiple time periods." *Journal of Econometrics*, 225(2): 200–230.

Calvano, Emilio, Giacomo Calzolari, Vincenzo Denicolò, and Sergio Pastorello. 2020. "Artificial Intelligence, Algorithmic Pricing, and Collusion." *American Economic Review*, 110(10): 3267–97.

Casey, Joan A., Frosch Rachel Morello, Daniel J. Mennitt, Kurt Fristrup, Elizabeth L. Ogburn, and Peter James. 2017. "Race/Ethnicity, Socioeconomic Status, Residential Segregation, and Spatial Variation in Noise Exposure in the Contiguous United States." *Environmental Health Perspectives*, 125(7): 077017.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma. 2018. "Manipulation Testing Based on Density Discontinuity." *The Stata Journal: Promoting communications on statistics and Stata*, 18(1): 234–261.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma. 2019. "Simple Local Polynomial Density Estimators."

Census. 2023. "Census Tabulation Detail: Tenure by Units in Structure." *https://censusreporter.org/tables/B25032/.*

Chaboud, Alain P., Benjamin Chiquoine, Erik Hjalmarsson, and Clara Vega. 2014. "Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market." *The Journal of Finance*, 69(5): 2045–2084.

Chen, Tianqi, and Carlos Guestrin. 2016. "XGBoost." ACM.

Choi, Jonathan H., and Daniel Schwarcz. 2023. "AI Assistance in Legal Analysis: An Empirical Study."

Choi, Jung Hyun, Caitlin Young, Alanna McCargo, Michael Neal, and Laurie Goodman. 2019. "Explaining the Black-White Homeownership Gap." *The Urban Institute.*

Chouldechova, Alexandra, Diana Benavides-Prado, Oleksandr Fialko, and Rhema Vaithianathan. 2018. "A Case Study of Algorithm-Assisted Decision Making in Child Maltreatment Hotline Screening Decisions." 134–148. PMLR.

Christophers, Brett. 2023. "How and Why U.S. Single-Family Housing Became an Investor Asset Class." *Journal of Urban History*, 49(2): 430–449.

Clark, Robert, Stephanie Assad, Daniel Ershov, and Lei Xu. 2023. "Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market." *Journal of Political Economy*, 0(ja): null.

CompTIA. 2024. "IT Industry Outlook 2024."

Consumer Financial Protection Bureau. 2023. "Using Publicly Available Information to Proxy for Unidentified Race and Ethnicity." *https://www.consumerfinance.gov/data-research/research-reports/using-publicly-available-information-to-proxy-for-unidentified-race-and-ethnicity/.*

Cowgill, Bo, and Catherine E Tucker. 2019. "Economics, fairness and algorithmic bias." *preparation for: Journal of Economic Perspectives.*

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. 1999. "The Rise and Decline of the American Ghetto." *Journal of Political Economy*, 107(3): 455–506.

Dawes, Robyn M. 1971. "A case study of graduate admissions: Application of three principles of human decision making." *American Psychologist*, 26(2): 180–188.

Dawes, Robyn M., David Faust, and Paul E. Meehl. 1989. "Clinical Versus Actuarial Judgment." *Science*, 243(4899): 1668–1674.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review*, 110(9): 2964–96.

Deprez-Sims, Anne-Sophie, and Scott B. Morris. 2010. "Accents in the workplace: their effects during a job interview." *International journal of psychology : Journal international de psychologie*, 45 6: 417–26.

Derenoncourt, Ellora, Chi Hyun Kim, Moritz Kuhn, and Moritz Schularick. 2022. "Wealth of Two Nations: The U.S. Racial Wealth Gap, 1860-2020." National Bureau of Economic Research Working Paper 30101.

Dietvorst, Berkeley J., Joseph P. Simmons, and Cade Massey. 2014. "Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err."

Ebbinghaus, Hermann. 1913. *Memory; a contribution to experimental psychology.* New York city, Teachers college, Columbia university.

Economist, The. 2021. "A Whodunnit on Zillow." *The Economist.*

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

Elliott, Marc N., Peter A. Morrison, Allen Fremont, Daniel F. McCaffrey, Philip Pantoja, and Nicole Lurie. 2009. "Using the Census Bureau's Surname List to Improve Estimates of Race/Ethnicity and Associated Disparities." *Health Services and Outcomes Research Methodology*, 9(2): 69–83.

Elster, Yael, and Noam Zussman. 2022. "Minorities and Property Values: Evidence from Residential Buildings in Israel." *Journal of Urban Economics*, 103525.

Erickson, Nick, Jonas Mueller, Alexander Shirkov, Hang Zhang, Pedro Larroy, Mu Li, and Alexander Smola. 2020. "AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data."

Fields, Desiree. 2018. "Constructing a New Asset Class: Property-led Financial Accumulation after the Crisis." *Economic Geography*, 94(2): 118–140.

Fields, Desiree. 2022. "Automated Landlord: Digital Technologies and Post-Crisis Financial Accumulation." *Environment and Planning A Economy and Space*, 54(1): 160–181.

Fish, Sara, Yannai A. Gonczarowski, and Ran I. Shorrer. 2024. "Algorithmic Collusion by Large Language Models."

Francke, Marc, Lianne Hans, Matthijs Korevaar, and Sjoerd van Bekkum. 2023. "Buy-to-Live vs. Buy-to-Let: The Impact of Real Estate Investors on Housing Costs and Neighborhoods."

Frankel, Alex. 2021. "Selecting Applicants." *Econometrica*, 89(2): 615–645.

Freddie Mac Economic & Housing Research. 2018. "Single-Family Rental: An Evolving Market." Federal Home Loan Mortgage Corporation.

Freddie Mac Economic & Housing Research. 2021. "Racial and Ethnic Valuation Gaps in Home Purchase Appraisals." Federal Home Loan Mortgage Corporation.

Fuster, Andres, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther. 2022. "Predictably Unequal? The Effects of Machine Learning on Credit Markets." *The Journal of Finance*, 77(1): 5–47.

Gillis, Talia B, and Jann L Spiess. 2019. "Big Data and Discrimination." *The University of Chicago Law Review*, 29.

Google. n.d.. "AI vs. Machine Learning: How Do They Differ?"

Guren, Adam M, Alisdair McKay, Emi Nakamura, and Jón Steinsson. 2020. "Housing Wealth Effects: The Long View." *The Review of Economic Studies*, 88(2): 669–707.

Gurun, Umit G, Jiabin Wu, Steven Chong Xiao, and Serena Wenjing Xiao. 2023. "Do Wall Street Landlords Undermine Renters' Welfare?" *The Review of Financial Studies*, 36(1): 70–121.

Halford, Graeme S., Rosemary Baker, Julie E. McCredden, and John D. Bain. 2005. "How many variables can humans process?" *Psychological Science*, 16(1): 70–76.

Hamermesh, Daniel S., and Jeff E. Biddle. 1994. "Beauty and the Labor Market." *The American Economic Review*, 84(5): 1174–1194.

Harris, Adam, and Maggie Yellen. 2023. "Human Decision-Making with Machine Prediction: Evidence from Predictive Maintenance in Trucking."

Harris, David. 1999. ""Property Values Drop When Blacks Move in, Because...": Racial and Socioeconomic Determinants of Neighborhood Desirability." *American Sociological Review*, 64(3): 461–479.

Hastie, R., and R. M. Dawes. 2001. *Rational choice in an uncertain world: The psychology of judgment and decision making.* Sage Publications.

Helander, M. G. 2014. *Handbook of Human-Computer Interaction.* Elsevier.

Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld. 2011. "Does Algorithmic Trading Improve Liquidity?" *The Journal of Finance*, 66(1): 1–33.

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder. 2003. "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *Econometrica*, 71(4): 1161–1189. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-0262.00442.

Howell, Junia, and Elizabeth Korver-Glenn. 2018. "Neighborhoods, Race, and the Twenty-first-century Housing Appraisal Industry." *Sociology of Race and Ethnicity*, 4: 473 – 490.

Invitation Homes. 2017. "2017-Annual-Report.Pdf." *https://s28.q4cdn.com/264003623/files/doc_financials/2017 Annual-Report.pdf.*

Kahneman, Daniel, Olivier Sibony, and Cass R. Sunstein. 2021. *Noise: A Flaw in Human Judgment.* . First edition ed., New York:Little, Brown Spark.

Kamin, Debra. 2023. "Home Appraised With a Black Owner: $472,000. With a White Owner: $750,000. - The New York Times." *New York Times.*

Kaysen, Ronda. 2023. "New Legislation Proposes to Take Wall Street Out of the Housing Market." *The New York Times.*

Kermani, Amir, and Francis Wong. 2021. "Racial Disparities in Housing Returns."

Kim, Sunwoong. 2000. "Race and Home Price Appreciation in Urban Neighborhoods: Evidence from Milwaukee, Wisconsin." *The Review of Black Political Economy*, 28(2): 9–28.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017*a*. "Human decisions and machine predictions*." *The Quarterly Journal of Economics*, 133(1): 237–293. tex.eprint: https://academic.oup.com/qje/article-pdf/133/1/237/30636517/qjx032.pdf.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017*b*. "Human Decisions and Machine Predictions*." *The Quarterly Journal of Economics*, 133(1): 237–293.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Cass R Sunstein. 2018. "Discrimination in the Age of Algorithms." *Journal of Legal Analysis*, 10.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. 2015. "Prediction Policy Problems." *American Economic Review*, 105(5): 491–495.

Kleinberg, Jon M., and Manish Raghavan. 2021. "Algorithmic monoculture and social welfare." *Proceedings of the National Academy of Sciences*, 118.

Kleinberg, Jon, Sendhil Mullainathan, and Manish Raghavan. 2016. "Inherent Trade-Offs in the Fair Determination of Risk Scores." *arXiv:1609.05807 [cs, stat]*. arXiv: 1609.05807.

LaValle, Steven M, Michael S Branicky, and Stephen R Lindemann. 2004. "On the relationship between classical grid search and probabilistic roadmaps." *The International Journal of Robotics Research*, 23(7-8): 673–692.

Lepri, B., Jacopo Staiano, David Sangokoya, Emmanuel Letouzé, and Nuria Oliver. 2016. "The Tyranny of Data? The Bright and Dark Sides of Data-Driven Decision-Making for Social Good." *ArXiv*, abs/1612.00323.

Levi, Yaron, David Hirshleifer, Siew Hong Teoh, and Ben Lourie. 2017. "Decision Fatigue and Heuristic Analyst Forecasts." *DecisionSciRN: Non-Rational Decision-Making (Topic).*

Lewis, Valerie A., Michael O. Emerson, and Stephen L. Klineberg. 2011. "Who We'll Live With: Neighborhood Racial Composition Preferences of Whites, Blacks and Latinos." *Social Forces*, 89(4): 1385–1407.

Li, Danielle, Lindsey Raymond, and Peter Bergman. 2020. "Hiring as Exploration." *NBER Working Paper No. 27736.*

Lilien, Jake. 2023. "Faulty Foundations: Mystery-Shopper Testing In Home Appraisals Exposes Racial Bias Undermining Black Wealth."

Malone, Thomas. 2023. "Residential Real Estate: Largest US Asset Class but Not Biggest Economic Driver." *https://www.corelogic.com/intelligence/why-the-uss-largest-asset-class-residential-real-estate-does-not-substantially-contribute-to-the-economic-output/.*

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao. 2013. "Poverty Impedes Cognitive Function." *Science*, 341(6149): 976–980. Publisher: American Association for the Advancement of Science.

McKinney, Scott Mayer, Marcin Sieniek, Varun Godbole, Jonathan Godwin, Natasha Antropova, Hutan Ashrafian, Trevor Back, Mary Chesus, Greg S. Corrado, Ara Darzi, Mozziyar Etemadi, Florencia Garcia-Vicente, Fiona J. Gilbert, Mark Halling-Brown, Demis Hassabis, Sunny Jansen, Alan Karthikesalingam, Christopher J. Kelly, Dominic King, Joseph R. Ledsam, David Melnick, Hormuz Mostofi, Lily Peng, Joshua Jay Reicher, Bernardino Romera-Paredes, Richard Sidebottom, Mustafa Suleyman, Daniel Tse, Kenneth C. Young, Jeffrey De Fauw, and Shravya Shetty. 2020. "International Evaluation of an AI System for Breast Cancer Screening." *Nature*, 577(7788): 89–94.

Merkley, Mr, and Adam Smith. 2023. "End Hedge Fund Control of American Homes Act of 2023."

Mills, James, Raven Molloy, and Rebecca Zarutskie. 2019. "Large-Scale Buy-to-Rent Investors in the Single-Family Housing Market: The Emergence of a New Asset Class." *Real Estate Economics*, 47(2): 399–430.

Mobius, Markus M., and Tanya S. Rosenblat. 2006. "Why Beauty Matters." *American Economic Review*, 96(1): 222–235.

Möbius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat. 2022. "Managing Self-Confidence: Theory and Experimental Evidence." *Management Science*, 68(11): 7793–7817.

Mullainathan, Sendhil, and Ashesh Rambachan. 2023. "From Predictive Algorithms to Automatic Generation of Anomalies."

Mullainathan, Sendhil, and Ziad Obermeyer. 2019. "Who is Tested for Heart Attack and Who Should Be: Predicting Patient Risk and Physician Error." *NBER WP*.

Mullainathan, Sendhil, and Ziad Obermeyer. 2021. "Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care*." *The Quarterly Journal of Economics*, 137(2): 679–727.

Mullainathan, Sendhil, and Ziad Obermeyer. 2022. "Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care*." *The Quarterly Journal of Economics*, 137(2): 679–727.

Neal, Michael, Laurie Goodman, and Caitlin Young. 2020. "Housing Supply Chartbook." *The Urban Institute.*

Noy, Shakked, and Whitney Zhang. 2023. "Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence." *Available at SSRN 4375283.*

Obermeyer, Ziad, and Ezekiel J. Emanuel. 2016. "Predicting the Future — Big Data, Machine Learning, and Clinical Medicine." *The New England journal of medicine*, 375(13): 1216–1219.

OECD. 2023. *OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market.* Paris:Organisation for Economic Co-operation and Development.

Peng, Baolin, Michel Galley, Pengcheng He, Hao Cheng, Yujia Xie, Yu Hu, Qiuyuan Huang, Lars Liden, Zhou Yu, Weizhu Chen, and Jianfeng Gao. 2023. "Check Your Facts and Try Again: Improving Large Language Models with External Knowledge and Automated Feedback."

Perry, Andre, Jonathan Rothwell, and David Harshbarger. 2018. “The Devaluation of Assets in Black Neighborhoods.”

Perry, Jonathan Rothwell and Andre M. 2021. “Biased Appraisals and the Devaluation of Housing in Black Neighborhoods.”

Pinto, Edward, and Tobias Peter. 2021. “AEI Housing Center Response to Perry and Rothwell (2021).” *American Enterprise Institute Housing Center.*

Quillian, Lincoln, John J. Lee, and Brandon Honoré. 2020. “Racial Discrimination in the U.S. Housing and Mortgage Lending Markets: A Quantitative Review of Trends, 1976–2016.” *Race and Social Problems*, 12(1): 13–28.

Raghavan, Manish, Solon Barocas, Jon Kleinberg, and Karen Levy. 2020. “Mitigating bias in algorithmic hiring.” ACM.

Rambachan, Ashesh. 2022. “Identifying Prediction Mistakes in Observational Data.” , (27111).

Rambachan, Ashesh, and Jonathan Roth. 2019. “Bias In, Bias Out? Evaluating the Folk Wisdom.”

Rambachan, Ashesh, Jon Kleinberg, Sendhil Mullainathan, and Jens Ludwig. 2020. “An Economic Approach to Regulating Algorithms.” National Bureau of Economic Research Working Paper 27111.

Raymond, Elora Lee, Ben Miller, Michaela McKinney, and Jonathan Braun. 2021. “Gentrifying Atlanta: Investor Purchases of Rental Housing, Evictions, and the Displacement of Black Residents.” *Housing Policy Debate*, 31(3-5): 818–834.

Raymond, Elora Lee, Richard Duckworth, Benjamin Miller, Michael Lucas, and Shiraj Pokharel. 2018. “From Foreclosure to Eviction: Housing Insecurity in Corporate-Owned Single-Family Rentals.” *Cityscape*, 20(3): 159–188.

Raymond, Elora Lee, Richard Duckworth, Benjmain Miller, Michael Lucas, and Shiraj Pokharel. 2016. “Corporate Landlords, Institutional Investors, and Displacement: Eviction Rates in Singlefamily Rentals.” *FRB Atlanta community and economic development discussion paper*, , (2016-4).

Redfin. 2023*a*. “Housing Investors Sell 1 in 7 Homes at a Loss—Highest Share Since 2016.” *https://www.redfin.com/news/investor-homes-sold-at-a-loss/*.

Redfin. 2023*b*. “Investor Home Purchases Fell a Record 49% in the First Quarter.” *https://www.redfin.com/news/investor-home-purchases-q1-2023/*.

Saxena, Devansh, and Shion Guha. 2024. “Algorithmic Harms in Child Welfare: Uncertainties in Practice, Organization, and Street-level Decision-making.” *ACM J. Responsib. Comput.*, 1(1).

Shen, Ruoqi, Liyao Gao, and Yi-An Ma. 2022. “On Optimal Early Stopping: Over-informative versus Under-informative Parametrization.”

Silver, David, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. 2016. “Mastering the game of Go with deep neural networks and tree search.” *Nature*, 529(7587): 484–489. Publisher: Nature Publishing Group.

Smith, Julie M. 2021. “Algorithms and Bias.” *Encyclopedia of Organizational Knowledge, Administration, and Technology*, 918–932.

Steyvers, Mark, and Aakriti Kumar. 2024. “Three Challenges for AI-Assisted Decision-Making.” *Perspectives on Psychological Science*, 19(5): 722–734.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics*, 225(2): 175–199.

Sunstein, Cass R. 2024. "The use of algorithms in society." *The Review of Austrian Economics*, 37(4): 399–420.

Tessum, Christopher W., David A. Paolella, Sarah E. Chambliss, Joshua S. Apte, Jason D. Hill, and Julian D. Marshall. 2021. "PM_{2.5} polluters disproportionately and systemically affect people of color in the United States." *Science Advances*, 7(18): eabf4491.

Tetlock, Philip E. 2005. *Expert political judgement: How good is it? How can we know? Expert political judgement: How good is it? How can we know?*, Princeton, NJ, US:Princeton University Press. Pages: xvi, 321.

The Department of Housing and Urban Development. 2023. "Legislative History of Lead-Based Paint."

The White House. 2009. "Memorandum on Transparency and Open Government."

The White House. 2022. "The Impact of Artificial Intelligence on the Future of Workforces in the European Union and the United States of America." The White House.

Tversky, Amos, and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science*, 185(4157): 1124–1131.

Upson, James, and Robert A. Van Ness. 2017. "Multiple Markets, Algorithmic Trading, and Market Liquidity." *Journal of Financial Markets*, 32: 49–68.

U.S. Census Bureau. 2021. "S2504 Physical Housing Characteristics for Occupied Housing Units."

US EPA, OAR. 2014. "Lead's Impact on Indoor Air Quality." *https://www.epa.gov/indoor-air-quality-iaq/leads-impact-indoor-air-quality.*

US EPA, OCSPP. 2013. "Lead Renovation, Repair and Painting Program." *https://www.epa.gov/lead/lead-renovation-repair-and-painting-program.*

Vaccaro, Michelle, Abdullah Almaatouq, and Thomas Malone. 2024. "When Combinations of Humans and AI are Useful: A Systematic Review and Meta-Analysis." *Nature Human Behaviour.*

Wason, P. C. 1960. "On the Failure to Eliminate Hypotheses in a Conceptual Task." *Quarterly Journal of Experimental Psychology*, 12(3): 129–140. Publisher: SAGE Publications.

Wheaton, David. 2023. "Fighting Appraisal Bias: How the Government and Housing Industry Can Better Address This Discriminatory Practice." *https://www.naacpldf.org/appraisal-algorithmic-bias/.*

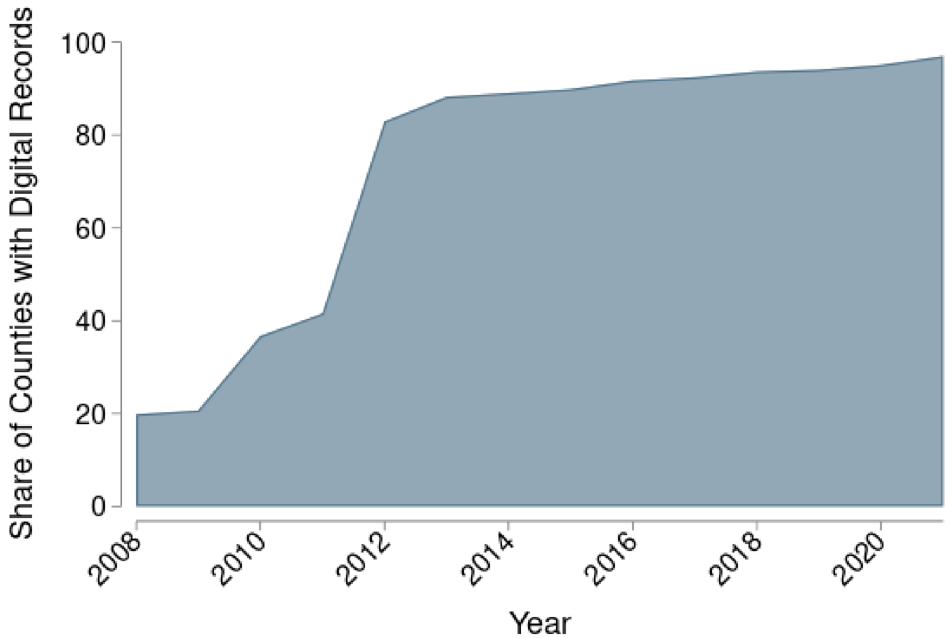
Wistrich, Andrew J., and Jeffrey John Rachlinski. 2017. "Implicit Bias in Judicial Decision Making How It Affects Judgment and What Judges Can Do About It." *Social Science Research Network.*

Zhang, Lei, and Tammy Leonard. 2021. "External Validity of Hedonic Price Estimates: Heterogeneity in the Price Discount Associated with Having Black and Hispanic Neighbors." *Journal of Regional Science*, 61(1): 62–85.

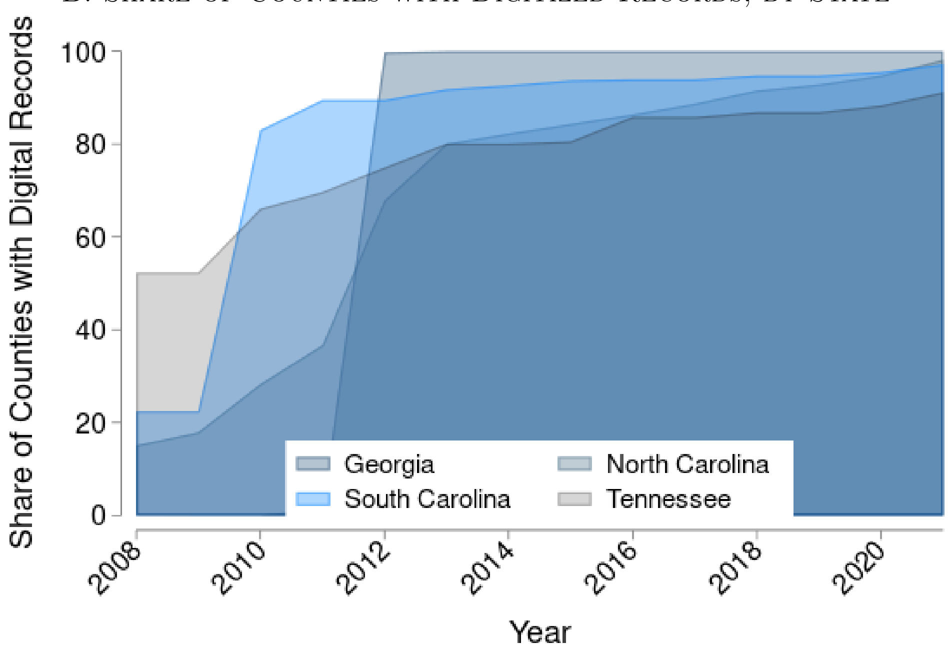
Zillow. 2023. "What Is a Zestimate? Zillow's Zestimate Accuracy."

Figure 1: County Record Digitization

A. Share of Counties with Digitized Records



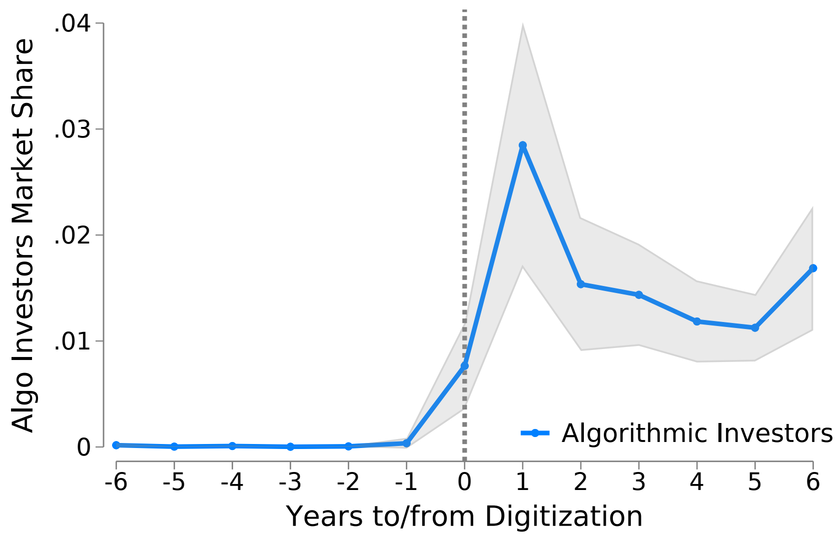
B. Share of Counties with Digitized Records, by State



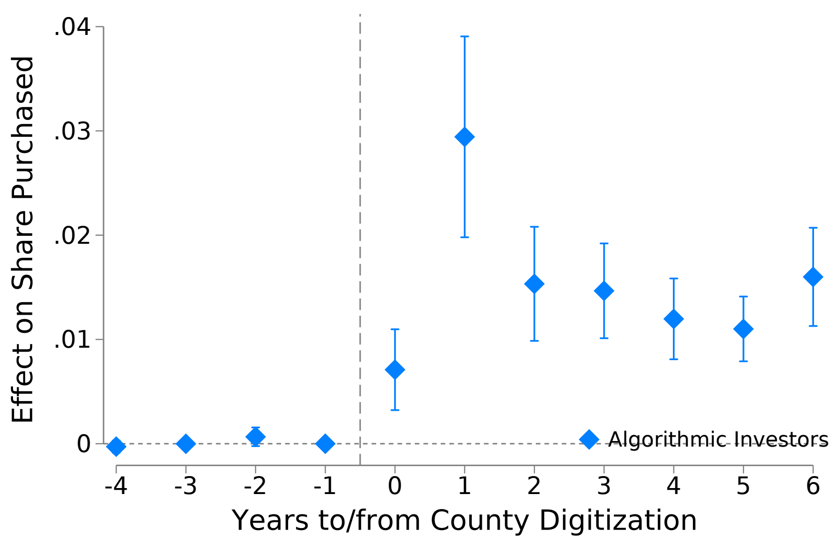
Notes: This figure shows the share of counties in the sample with digitized and publicly accessible recorder data over time. Panel B shows the share by state. The graphs are weighted by the number of housing transactions. All data comes from county governments.

Figure 2: Algorithmic investors buying, by time to digitization

A. Houses Purchased by Algorithmic Investors



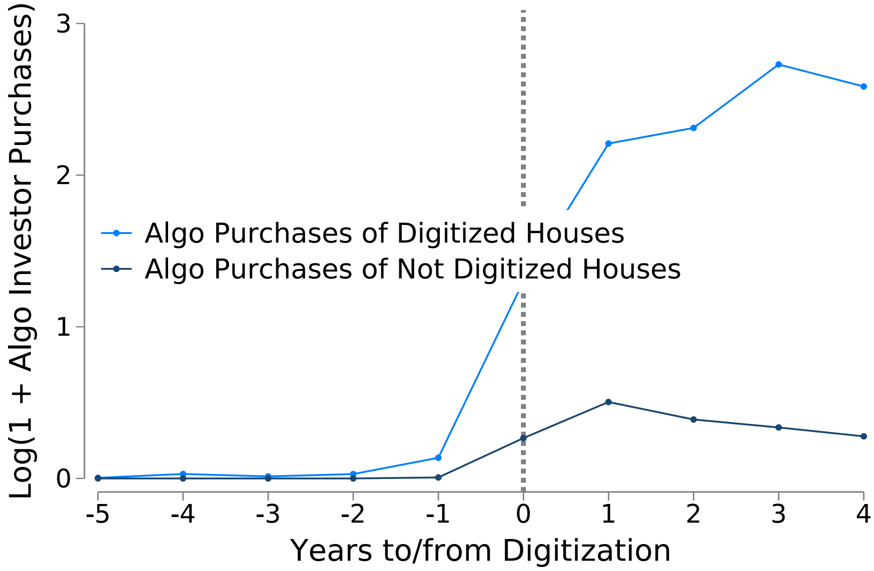
B. Impact of County Digitization on Algorithmic Investment)



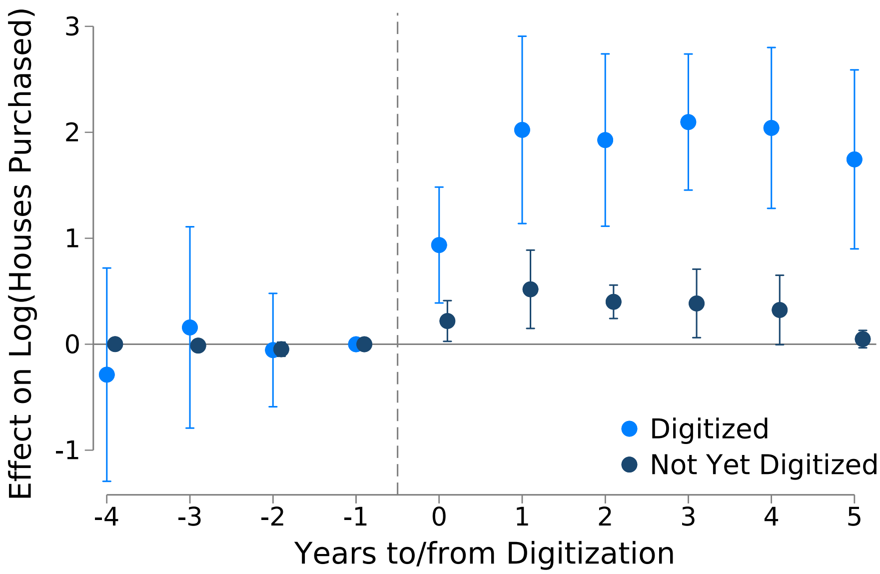
Notes: Panel A of this figure shows the number of houses purchased by algorithmic investors in county c and in year t as a share of totla house sales in that county-year, by time to digitization. Panel C shows the event-study estimates of the impact of county digitization on the share of houses purchased by algorithmic investors, including county- and year-fixed effects and controlling for county population and minority share. Standard errors are clustered at the county level and estimates are weighted by county size. All data come from ATTOM Data, Zillow, and county digitization records.

Figure 3: Houses Purchased by Algorithmic Investors, by House Digitization Status

A. Log(Houses Purchased), by House Digitization, Raw Data



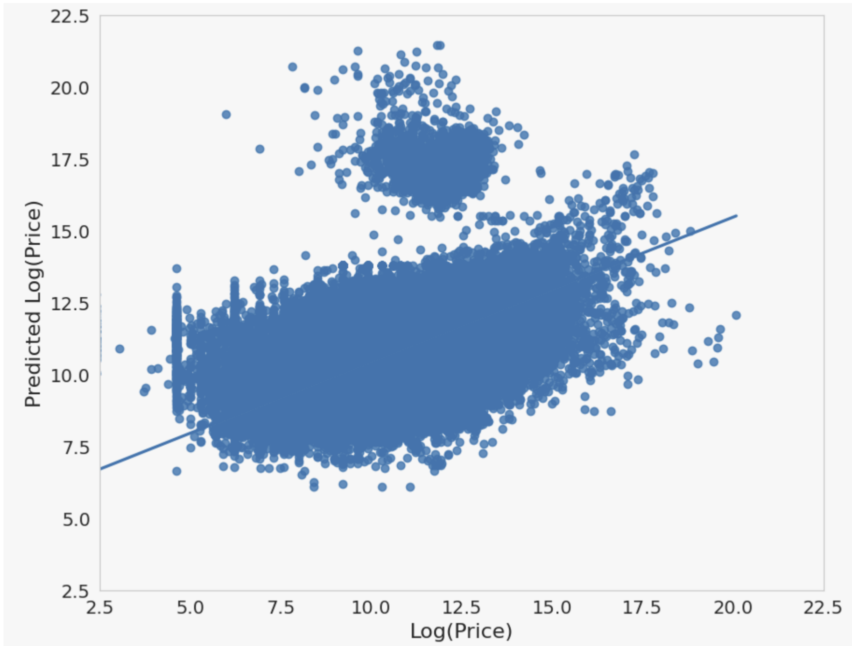
B. Log(Houses Purchased), by House Digitization, Event Study



Notes: These figures show the impact of county digitization on the number of homes purchased by algorithmic firms separately estimated for *digitized houses*, houses that have been digitized, and *non-digitized houses*, houses that have not been digitized and only have paper records. Panel A shows the raw natural log of the number of homes purchased by algorithmic firms, and Panel B plots the coefficients and 95 percent confidence intervals from Sun and Abraham (2021) interaction-weighted event study regressions. All specifications include state- and year-fixed effects, county population size and minority share; standard errors are clustered at the county level and are weighted by the number of transactions. All data come from ATTOM Data, Zillow, and county digitization records.

Figure 4: Model Predicted vs. Actual Price

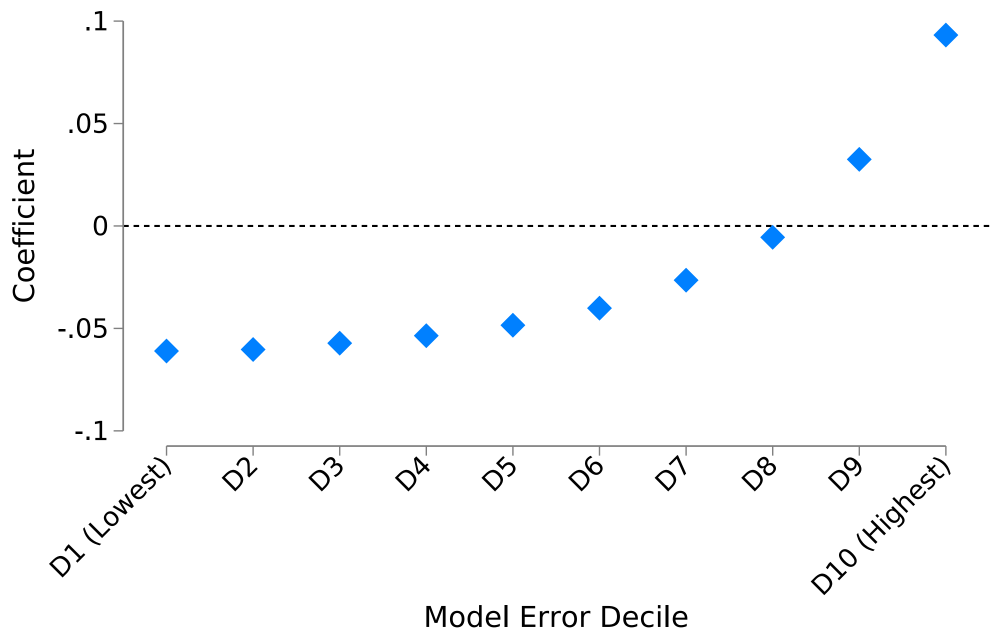
A. Predicted vs. Actual Prices, Out of Sample



B. Predicted vs. Actual Prices, by Future Investor Purchase

Notes: Panel A plots the model-predicted natural log of sales price and actual sale price on a held-out sample of housing transactions. Panel B shows the same results separately for houses that will be purchased in the future by human investors and those that will be purchased by algorithmic investors. All data comes from ATTOM Data and Zillow.

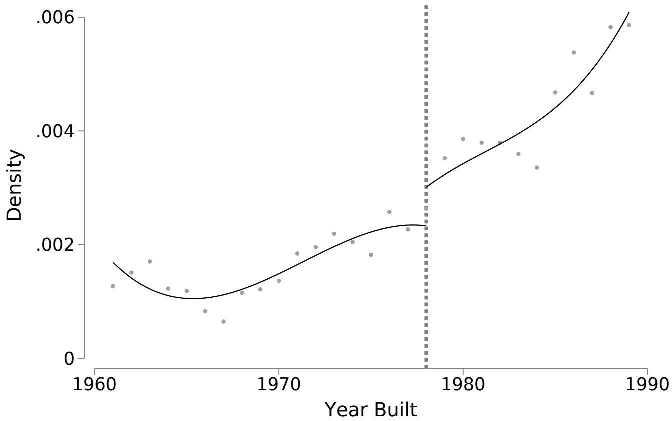
Figure 5: Impact of Digitization by House Predictability



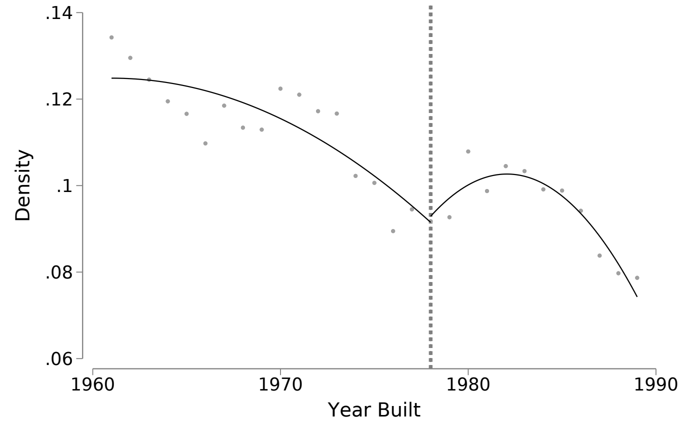
Notes: These figures plot the impact of house-level digitization on the likelihood of a purchase by a human investor. The model error is calculated as the average difference between the actual and predicted prices for each house. Errors are residualized to account for year-specific fixed effects. Every house is grouped into a decile of model error, with the houses with the lowest mean absolute error in decile 1 and the houses with the largest error in decile 10. All specifications include census block group and year-fixed effects. All data come from ATTOM Data and Zillow.

Figure 6: Discontinuities

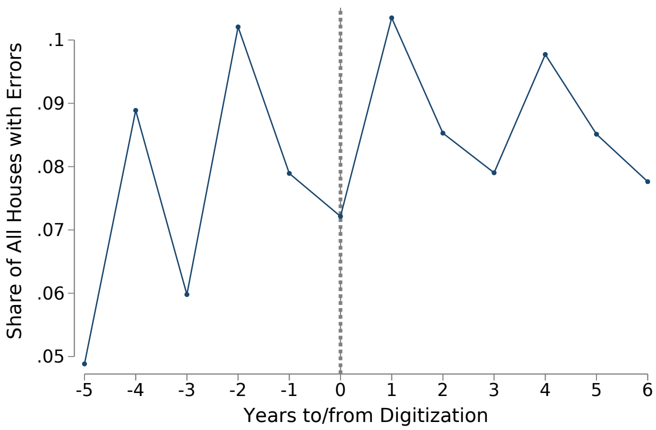
A. Algorithmic Investors and the Lead Paint Ban



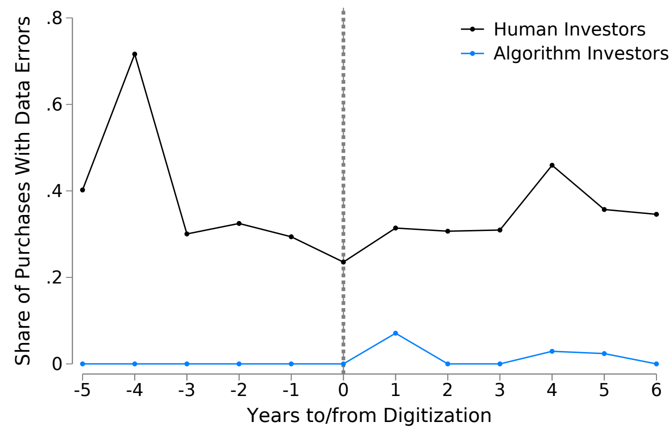
B. Human Investors and the Lead Paint Ban



C. Share of All Houses with Data Errors

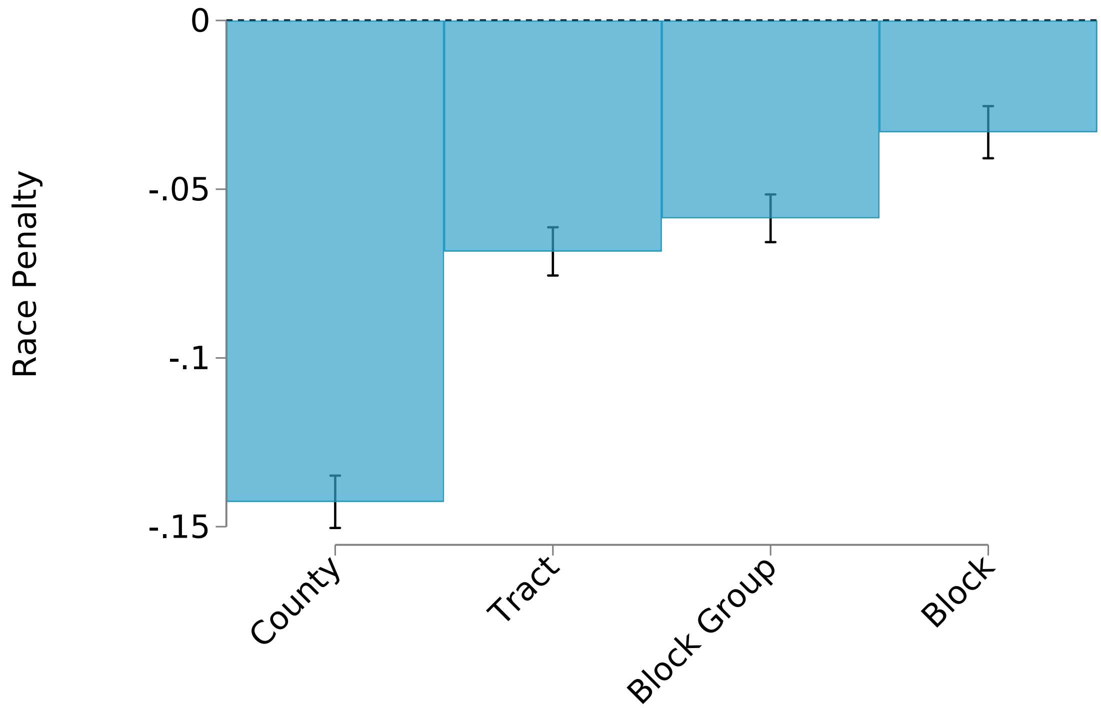


B. Investor Purchases of Houses with Data Errors



Notes: Panel A plots the distribution of houses purchased by algorithmic investors by year of construction. Panel B plots the same for human investors. Panel C shows the share of houses sold every year with data errors. Panel D plots the share of houses purchased by algorithmic and human investors with data errors. All data comes from ATTOM Data and county digitization records.

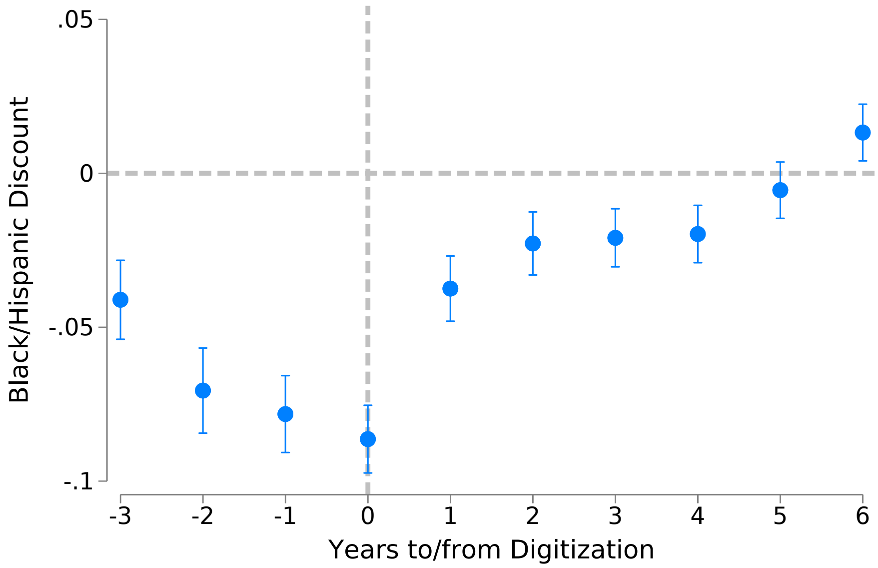
Figure 7: Race Penalty before Digitization, by Geography



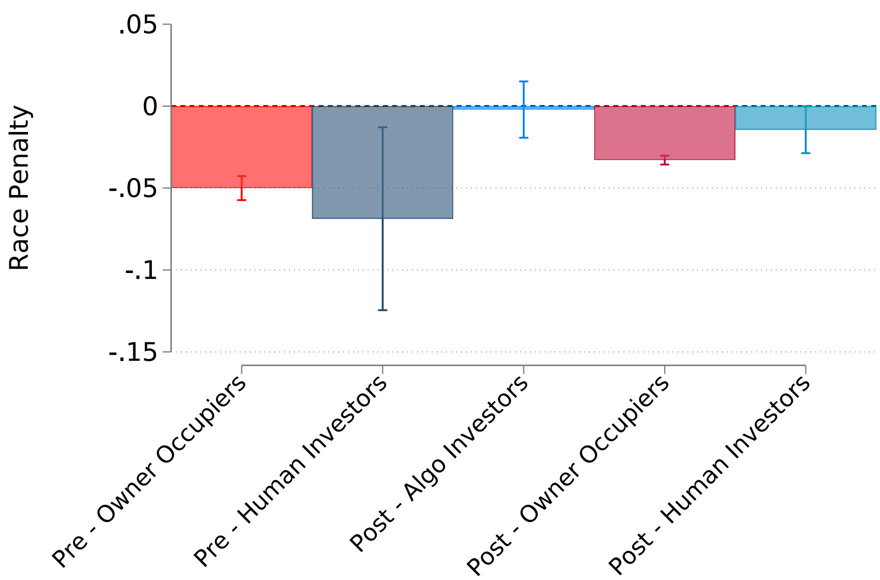
Notes: This table shows the race penalty or coefficient value that captures the residual difference in sales price between an observably similar house sold by Black or Hispanic homeowners and one sold by a White homeowner. The race penalty is calculated during the time before digitization. The regressions run include geography and year-fixed effects along with all available observable characteristics of the house. Standard errors are clustered at the relevant geography. All data comes from ATTOM Data.

Figure 8: Race Penalty, by Time to Digitization

A. Race Penalty, by Time to Digitization



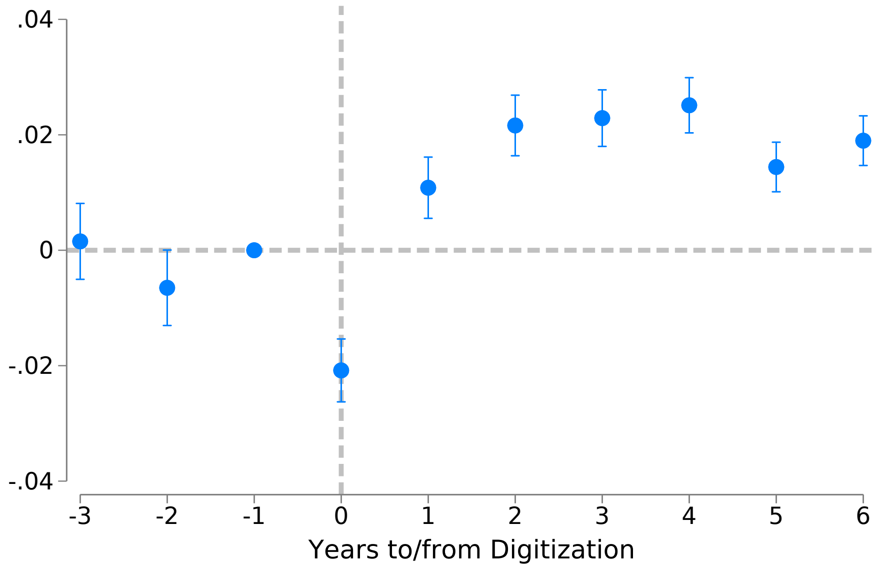
B. Race Penalty, by Digitization and Buyer



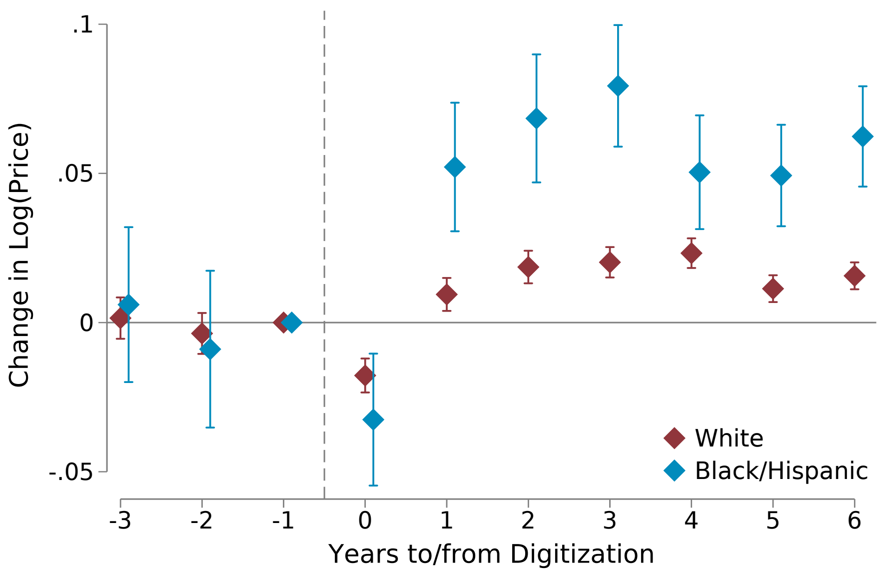
Notes: Panel A shows the race penalty, or residual difference in sale price between houses sold by White and minority homeowners by time to digitization. All specifications include census block group fixed effects and year-fixed effects, and standard errors are clustered at the block group level. Panel B shows the same coefficient plotted pre- and post-digitization for houses purchased by three different types of buyers: owner-occupiers, human investors, and algorithmic investors. All data comes from ATTOM Data and county digitization records.

Figure 9: Digitization on Price

A. Log(Price)

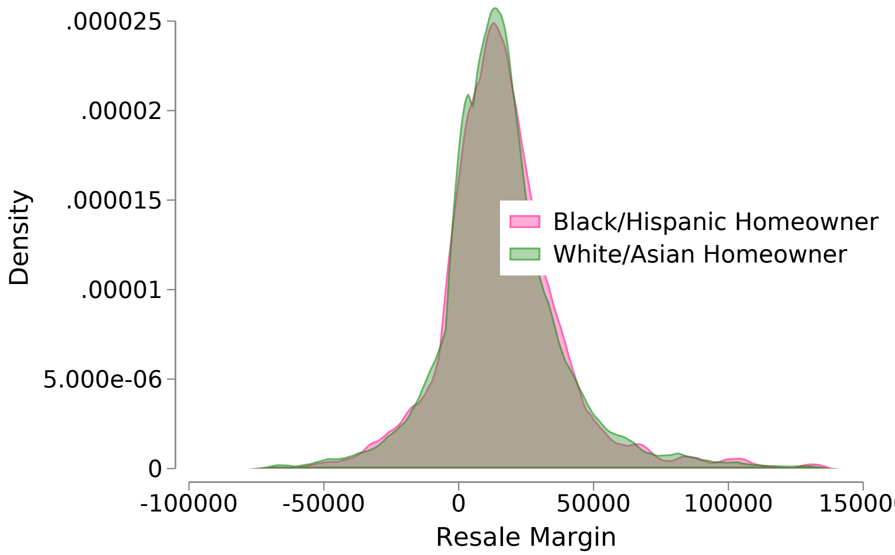


B. Log(Price), by Homeowner Race



Notes: This graph plots the impact of digitization on the natural log of housing transaction prices at the county level in aggregate and separately by White and minority homeowner. All specifications include census block and year-fixed effects; standard errors are clustered at the block level. All data comes from ATTOM Data, Zillow, and county digitization records.

Figure 10: Resale Margin, by Homeowner Race



Notes: This graph plots the gross margin or difference between the sale price and the purchase price for houses bought by algorithmic investors according to the race of the homeowner. All data comes from ATTOM Data and county digitization records.

Table 1: Summary Statistics on House Purchases, by Buyer Type

	(1) Owner Occupiers	(2) Human Investors	(3) Algo Investors
Sale Price	194,270.04	127,755.99	219,130.88
	(158,431.32)	(145,159.96)	(103,655.74)
Bedrooms	2.12	2.27	2.76
	(3.17)	(3.58)	(1.47)
Bathrooms	2.14	2.09	2.47
	(2.38)	(5.30)	(1.01)
Partial Baths	0.27	0.25	0.43
	(0.48)	(0.48)	(0.50)
Stories	1.25	1.18	1.57
	(0.75)	(0.86)	(0.64)
Additional Buildings	0.07	0.12	0.03
	(0.58)	(1.18)	(0.24)
Garage	0.56	0.48	0.82
	(0.50)	(0.50)	(0.38)
Fireplace	0.59	0.55	0.82
	(0.49)	(0.50)	(0.39)
Basement	0.17	0.13	0.17
	(0.37)	(0.34)	(0.38)
Parking Spaces	0.75	0.58	0.91
	(8.72)	(7.40)	(0.99)
House Age	30.94	36.30	21.31
	(25.89)	(29.26)	(15.53)
Age Since Remodel	24.27	28.82	18.85
	(21.18)	(25.10)	(13.96)
Observations	7223587	975776	111027

mean coefficients; sd in parentheses

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

Notes: This table shows the house characteristics of the transactions in our sample. The sample in column 1 includes all houses:those purchased by owner-occupiers (i.e., those buying houses to live in) and investors. Column 2 includes purchases made by human investors, and column 3 includes purchases by investors using algorithms. Houses with missing or zero transaction prices are removed from the sample. All data come from ATTOM Data and ZTRAX.

Table 2: Investors Purchases, by County Characteristics

Variable	(1) Human Investors	(2) Algorithmic Investors	(3) Difference
County 2010 Population	347,012.69	498,205.72	151,193.02***
	(317,242.69)	(327,782.25)	(0.00)
Total Housing Units	152,197.36	208,074.12	55,876.77***
	(138,687.72)	(141,824.16)	(0.00)
Share Black	27.83	28.58	0.76***
	(16.73)	(14.35)	(0.00)
Share Hispanic	7.77	10.63	2.86***
	(4.20)	(4.66)	(0.00)
Share White	58.71	53.05	-5.66***
	(19.49)	(17.41)	(0.00)
Share Asian	2.98	4.55	1.56***
	(2.29)	(2.93)	(0.00)
Share Persons under 18	24.47	26.39	1.93***
	(2.74)	(2.33)	(0.00)
Median Income	54,298.52	66,305.10	12,006.58***
	(12,705.78)	(12,281.54)	(0.00)
Median Rent	896.24	1,085.91	189.67***
	(191.01)	(182.29)	(0.00)
Share Families in Poverty	11.74	9.31	-2.43***
	(3.82)	(2.83)	(0.00)
Mean Family Size	3.18	3.29	0.10***
	(0.19)	(0.15)	(0.00)
Share Persons under 18	24.47	26.39	1.93***
	(2.74)	(2.33)	(0.00)
Observations	975,776	111,027	1,086,803

Notes: This table shows socioeconomic and demographic characteristics of counties where algorithmic and human investors purchase houses, weighted by the number of purchases. Data is at the house transaction level. All data comes from the US Decennial Census and the American Community Survey.

Table 3: Log(Houses Purchased) by Algorithmic Investors, Difference-in-Difference Estimators

	Point Estimate	Standard Error	Lower Bound 95% Confidence Interval	Upper Bound 95% Confidence Interval
TWFE-OLS	1.130	0.380	0.386	1.874
Borusyak-Jaravel-Spiess	2.451	0.446	1.578	3.325
Callaway-Sant'Anna	1.002	0.021	0.960	1.043
DeChaisemartin-D'Haultfoeuille	2.653	0.325	2.015	3.290
Sun-Abraham	1.988	0.286	1.428	2.549

Notes: This table shows the impact of county data digitization deployment on the log of houses purchased by algorithmic investors. I show results using the robust difference-in-differences estimators introduced in Borusyak, Jaravel and Spiess (2022), Callaway and Sant'Anna (2021), de Chaisemartin and D'Haultfœuille (2020) and Sun and Abraham (2021) along with a traditional two way fixed-effects. Callaway and Sant'Anna (2021) cannot be weighted, so I present the unweighted estimates. All regressions include county, year-fixed effects; standard errors are clustered at the county level. Regressions are weighted by the number of transactions.

Table 4: House Digitization on Algorithmic Investor Purchase

VARIABLES	(1) Algorithmic Investors	(2) Algorithmic Investors	(3) Algorithmic Investors	(4) Algorithmic Investors	(5) Human Investors
House Digitized	0.0023**	0.0021***	0.0009**		
	(0.0011)	(0.0008)	(0.0004)		
County Digitization x House **Not** Digitized				-0.0006	0.0017
				(0.0008)	(0.0023)
County Digitization x House Digitized				0.0098***	-0.0069***
				(0.0008)	(0.0023)
Observations	6,895,957	6,890,606	6,817,554	6,890,606	6,890,606
R-squared	0.0550	0.0598	0.1056	0.0600	0.0716
House Characteristics	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Location FE	Tract	Block Group	Block	Block Group	Block Group
Preperiod DV Mean	.00013	.00013	.00013	.00013	0.1247

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10

Notes: This table shows the results of cross-sectional difference-in-difference regressions estimating the impact of house record digitization on the purchase by an algorithmic investor. All specifications include house characteristics, year, and geography-fixed effects; standard errors are clustered at the geographic level. All data come from ATTOM Data, ZTRAX, and county governments.

Table 5: Race Penalty, with House Images

	(1) Log(Price)	(2) Log(Price)	(3) Log(Price)
Seller Black/Hispanic	-0.0557***	-0.0441***	-0.021***
	(0.0023)	(0.0042)	(0.0051)
Observations	30,130	30,130	29,037
R-squared	0.69	0.71	0.83
House + Lot	Yes	Yes	Yes
Year x Geo	Yes	Yes	Yes
Geographic FE	Tract	Block Group	Block
Adjusted R-squared	.571	.598	.688

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10

Notes: This table shows the race penalty, the residual difference in sale price between houses sold by White and minority homeowners. House exteriors are captured using a deep learning model to create vector representations of house images and included in the regressions as controls. All specifications include house characteristics, year, and geography-fixed effects, and standard errors are clustered at the geographic level. All data come from ATTOM Data, ZTRAX, Zillow and investor websites.

Table 6: House Digitization on Algorithmic Investor Purchase, by Homeowner Race

	(1) Algorithm Purchase	(2) Algorithm Purchase	(3) Algorithm Purchase
Seller Minority	-0.0037***	-0.0039***	-0.0049***
	(0.0005)	(0.0005)	(0.0004)
Digitization x Seller White	0.0022**	0.0020**	0.0007*
	(0.0011)	(0.0008)	(0.0004)
Digitization x Seller Minority	0.0044***	0.0042***	0.0043***
	(0.0007)	(0.0006)	(0.0005)
Geography FE	Tract	Block Group	Block
Year FE	Yes	Yes	Yes
Sample	All	All	All
DV Mean	.0018	.0018	.0018
Observations	6895957	6890606	6817554

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows the results of cross-sectional difference-in-difference regressions estimating the impact of house record digitization on the purchase by an algorithmic investor. I separately estimate effects by homeowner race. All specifications include house characteristics, year, and geography-fixed effects;standard errors are clustered at the relevant geographic level. All data come from ATTOM Data, ZTRAX, and county governments.

Table 7: Resale Margin

VARIABLES	(1) Log(Resale Margin)	(2) Log(Resale Margin)	(3) Log(Resale Margin)	(4) Log(Resale Margin)	(5) Log(Resale Margin)	(6) Log(Resale Margin)	(7) Log(Resale Margin)	(8) Log(Resale Margin)
Seller Minority = 1	0.010	0.010	-0.003	0.002	0.095***	0.101***	0.121***	0.066
	(0.009)	(0.010)	(0.018)	(0.017)	(0.019)	(0.021)	(0.033)	(0.041)
Seller Minority x Minority Neighborhood = 1				0.012				0.047
				(0.021)				(0.048)
Observations	6,775	5,630	2,212	5,630	56,459	45,527	23,449	45,527
R-squared	0.459	0.515	0.646	0.515	0.403	0.452	0.474	0.452
FE	Year x Tract	Year x Block Group	Year x Block	Year x Block Group	Year x Tract	Year x Block Group	Year x Block	Year x Block Group
Resale Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyers	Algorithms	Algorithms	Algorithms	Algorithms	Humans	Humans	Humans	Humans
DV Mean	0.0646	0.0576	0.0549	0.0576	0.369	0.362	0.307	0.362

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10

Notes: This table shows the difference in the natural log of the price the house sells for in the future, or the resale price, and the natural log of the price paid, or the *gross margin*. The *Seller Minority* variable indicates if the house was bought from Black or Hispanic homeowners or White homeowners. *Minority neighborhood* indicates if the house is in a census block group with an above average minority resident share. All specifications includes the resale year and sale year by geography-fixed effects; standard errors are clustered at the geographic level. All data comes from ATTOM Data.

Appendix Materials – For Online Publication

Figure A.1: Capitalization Rate Example

DELPHI PROPERTY GROUP

PRO FORMA

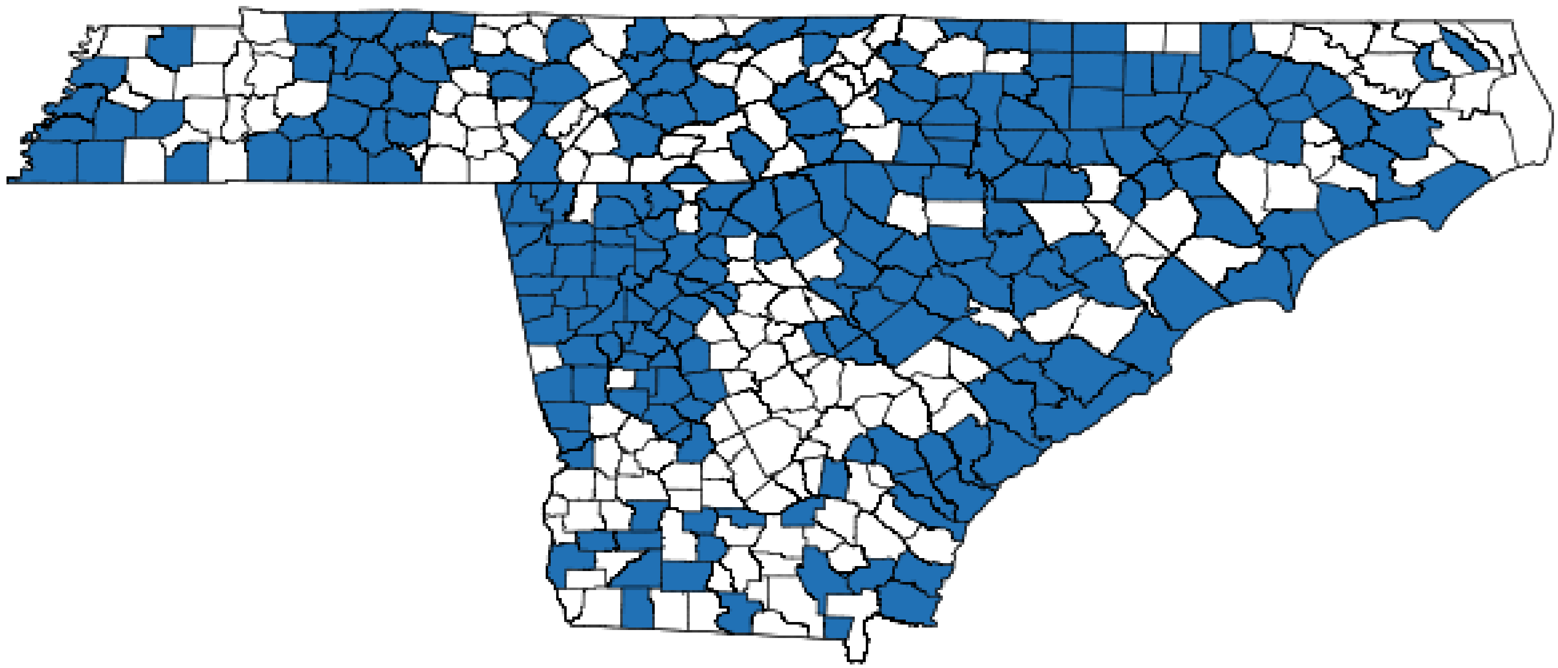
Income:			
Residential			
Gross Revenue		$	778,200.00
Vacancy; 5%		$	(38,910.00)
Effective Gross Residential Income:		**$**	**739,290.00**
Commercial			
Gross Revenue		$	150,000.00
Vacancy; 5%		$	(7,500.00)
Effective Gross Commercial Income:		**$**	**142,500.00**
Total Gross Revenue		**$**	**881,790.00**
Expenses:			
Taxes		$	74,176.13
Management Fee	*5.0%*	$	44,089.50
CAM - *Estimated*		$	45,000.00
Miscallaneous - *Estimated*		$	30,000.00
Insurance		$	11,511.00
Electric (Common)		$	10,000.00
Water		$	5,000.00
Trash		$	141.60
Advanced Disposal		$	3,468.36
Total Expenses		**$**	**223,386.59**
Net Operating Income		**$**	**658,403.41**

Pricing	
Sale Price	$11,000,000.00
Number of Units	47 Apartments & 2 Commercial
Price / Unit	$224,489.80
Gross Building Area	54,000 SF
Price PSF	$203.70

Investment Summary	
Cap Rate	6.0%
NOI	$658,403.41

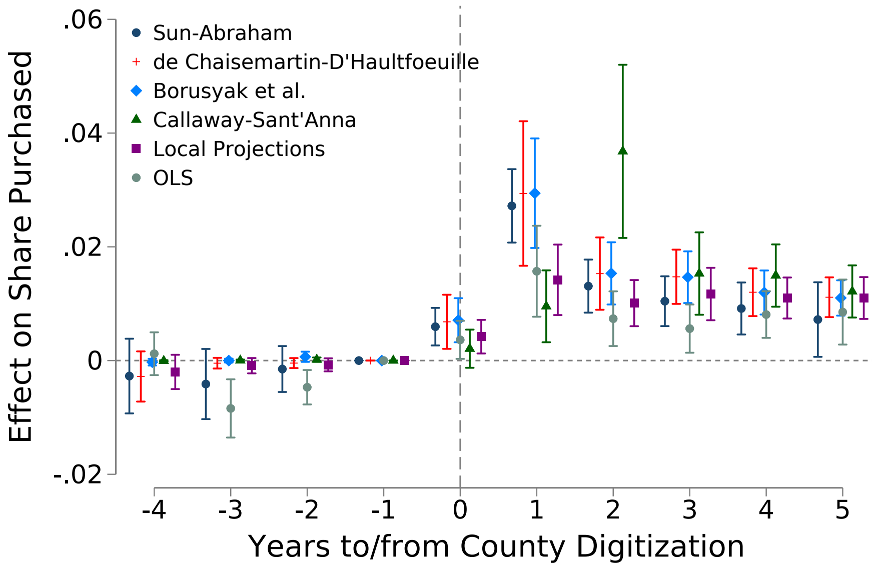
Notes: This shows a sample of the marketing material for 1 West Main St., Norristown, Penn., a mixed-use, multifamily apartment building. This page includes the building capitalization rate.

Figure A.2: Investor Activity



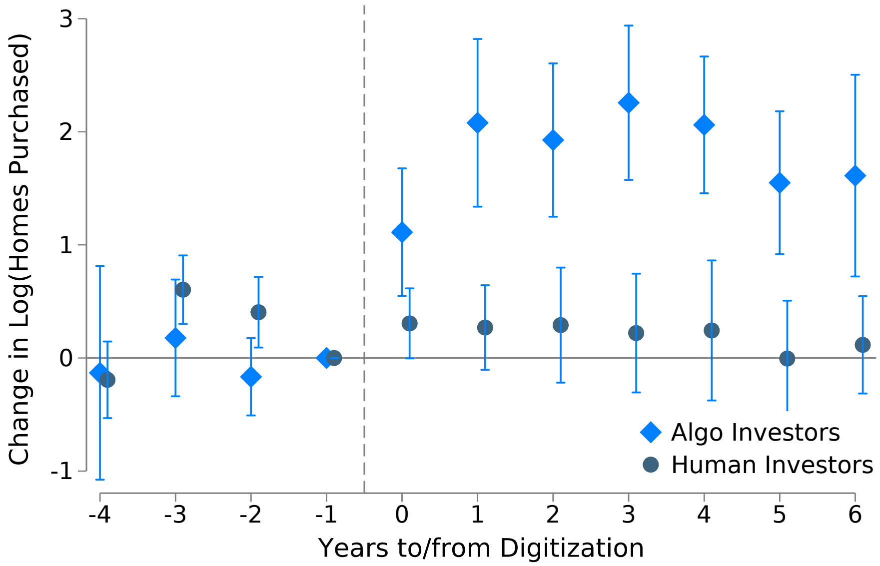
Notes: This graph shows the counties in Georgia, North Carolina, South Carolina, and Tennessee where human and algorithm investors are active in blue. Counties in white have only human investors.

Figure A.3: Alternative Event Studies, County Digitization on Algorithmic Investment



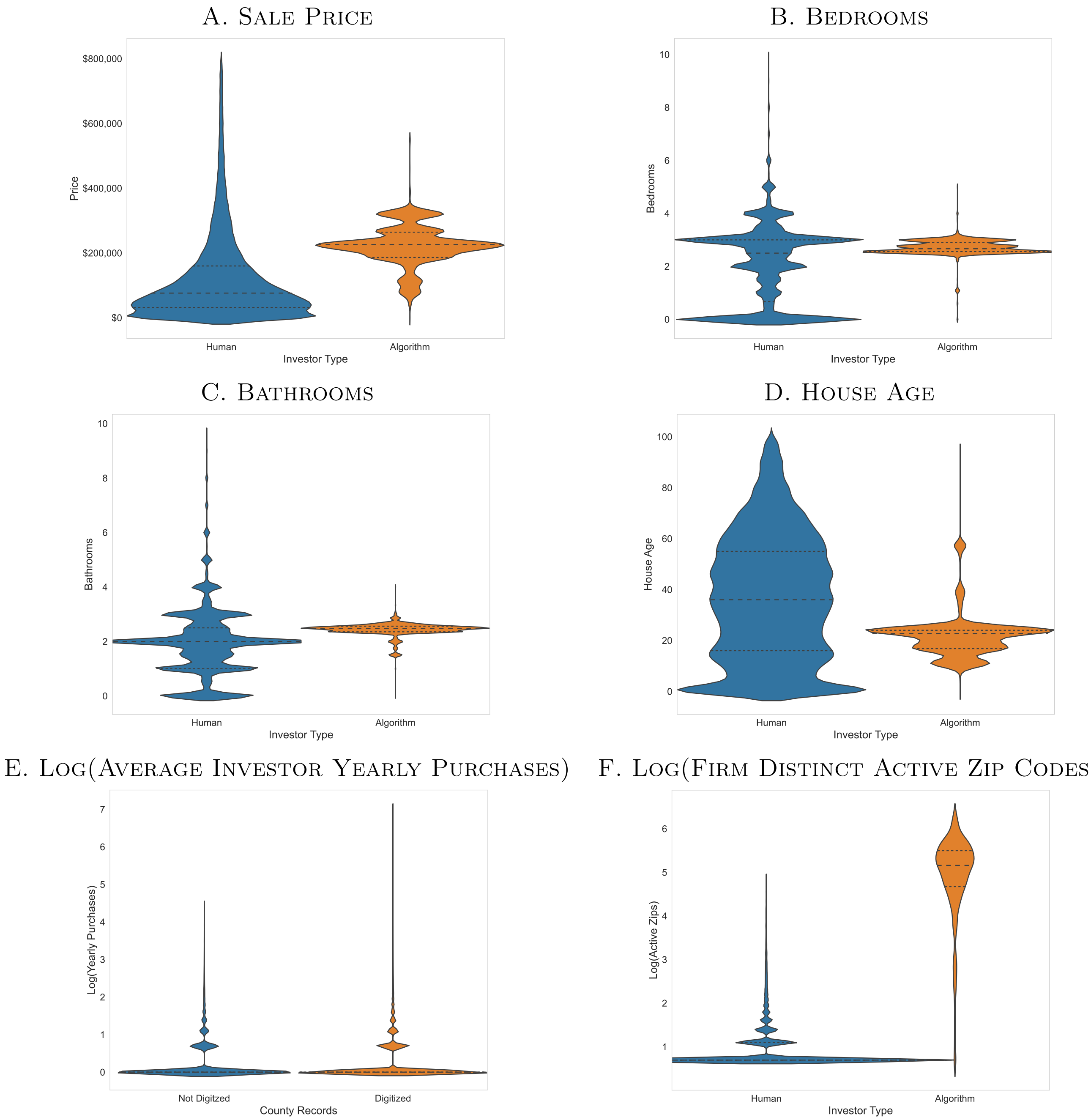
Notes: These figures plot the coefficients and 95 percent confidence intervals using a variety of robust dynamic difference-in-differences estimators introduced in Borusyak, Jaravel and Spiess (2022), de Chaisemartin and D'Haultfœuille (2020), Sun and Abraham (2021) and a standard two-way fixed effects regression model. All specifications include state and year fixed effects, controls for county population, share minority residents, and wealth; standard errors are clustered at the county level and are weighted by the number of transactions. All data comes from ATTOM Data, Zillow, and county digitization records.

FIGURE A.4: Log(House Purchases) by Investor Type



NOTES: This figure plots coefficients and 95 percent confidence interval from Sun and Abraham (2021) interaction-weighted event study regressions of county digitization on the natural log of the quantity of homes. I plot these results separately for the number of houses purchased by human or algorithmic investors in each county and year, weighted by the number of transactions. The regression includes state and year fixed effects, and standard errors are clustered at the county level. All data comes from ATTOM Data, Zillow, and county digitization records.

Figure A.5: House Characteristics, by Investor



Notes: This figures plots characteristics of houses purchased by human and algorithmic investors. Panel A plots the purchase prices of houses;Panel B plots the number of house bedrooms. Panel C shows the number of bathrooms, and panel D shows the age of the house. Panel E plots the natural log of average houses purchased by investors before and after digitization. Panel F plots the natural log of zip codes where investors are active each year. All data come from ATTOM Data and Zillow.

Figure A.6: House Exterior Images

Notes: This shows an example of the exterior images of the house used in the deep learning model.

Table A.1: Balance Table: Counties, by Year of Digitization

Variable	(1) Early Digitizers	(2) Late Digitizers	(3) Difference
Population	84,157.59	49,187.26	-34,970.33***
	(144,805.80)	(51,842.46)	(0.00)
Unemployment Rate	4.69	4.51	-0.19
	(1.81)	(1.69)	(0.36)
Share in Labor Force	56.67	54.77	-1.89**
	(7.01)	(6.24)	(0.01)
Share Units Occupied	82.58	81.33	-1.25
	(8.68)	(8.82)	(0.23)
Share Vacant	2.17	1.90	-0.27
	(1.44)	(2.17)	(0.27)
Median Rent	710.86	679.33	-31.53*
	(164.93)	(161.44)	(0.10)
Share Families in Poverty	14.66	14.62	-0.04
	(5.25)	(4.99)	(0.95)
Mean Family Size	3.14	3.07	-0.07***
	(0.29)	(0.20)	(0.01)
Median Income	44,399.43	42,521.30	-1,878.12
	(11,331.64)	(12,210.29)	(0.19)
Share Black	22.99	19.92	-3.07
	(18.08)	(19.44)	(0.17)
Share Hispanic	5.81	4.63	-1.18***
	(4.61)	(3.64)	(0.01)
Share White	67.19	71.74	4.56*
	(19.80)	(20.63)	(0.06)
Share Asian	1.25	0.85	-0.40***
	(1.31)	(0.95)	(0.00)
Observations	303	97	400

Notes: This table shows the covariate balance table for counties digitized before and after the median. All variables are calculated at the county level. All data come from ATTOM Data, ZTRAX, and the US Census.

Table A.2: Balance Table: Houses, by Year of Digitization

Variable	(1) Early Digitizers	(2) Late Digitizers	(3) Difference
Years since Sale	10.55	9.79	-0.76***
	(9.49)	(8.74)	(0.00)
Sale Price	210,262.55	202,658.52	-7,604.03***
	(959,260.94)	(804,228.75)	(0.00)
Bedrooms	2.19	2.07	-0.12***
	(1.68)	(3.29)	(0.00)
Bathrooms	2.03	2.14	0.11***
	(2.69)	(2.24)	(0.00)
Partial Baths	0.29	0.27	-0.02***
	(0.52)	(0.47)	(0.00)
Stories	1.17	1.26	0.09***
	(0.88)	(0.69)	(0.00)
Buildings	0.05	0.07	0.01***
	(0.42)	(0.53)	(0.00)
Garage	0.55	0.56	0.02***
	(0.50)	(0.50)	(0.00)
Fireplace	0.60	0.58	-0.02***
	(0.49)	(0.49)	(0.00)
Basement	0.18	0.17	-0.01***
	(0.38)	(0.37)	(0.00)
Parking Spaces	0.97	0.69	-0.28***
	(17.96)	(1.77)	(0.00)
House Age	33.12	30.18	-2.94***
	(24.85)	(26.03)	(0.00)
Age Since Remodel	27.87	23.68	-4.19***
	(21.54)	(21.13)	(0.00)
Minority Homeowner	0.04	0.04	-0.00***
	(0.20)	(0.19)	(0.00)
Homeowner Asian	0.02	0.03	0.01***
	(0.15)	(0.17)	(0.00)
Homeowner White	0.88	0.87	-0.01***
	(0.33)	(0.34)	(0.00)
Observations	1,096,423	3,684,075	4,780,498

Notes: This table shows the covariate balance table for houses that digitized before and after the median ("Early Digitizers") or later ("Late Digitizers"). The unit of observation is at the house level. All data come from ATTOM Data, ZTRAX, and the US Census.

Table A.3: Balance Table: Neighborhood Characteristics, by Year of Digitization

Variable	(1) Early Digitizers	(2) Late Digitizers	(3) Difference
Population	2,065.07	2,215.21	150.13***
	(1,284.76)	(1,448.16)	(0.00)
Housing Units	928.23	992.07	63.84***
	(540.46)	(600.80)	(0.00)
Share White	68.00	67.66	-0.34***
	(26.67)	(28.04)	(0.00)
Share Black	20.04	21.25	1.21***
	(21.69)	(24.18)	(0.00)
Share Asian	2.50	2.85	0.35***
	(3.12)	(4.33)	(0.00)
Share Under 18	23.66	24.01	0.35***
	(6.34)	(6.22)	(0.00)
Median Earnings	53,533.62	54,046.14	512.52***
	(13,100.19)	(13,607.33)	(0.00)
Rent	864.28	878.85	14.57***
	(200.30)	(199.76)	(0.00)
Age	38.23	38.09	-0.13***
	(4.39)	(4.43)	(0.00)
Mortgage Costs	1,310.06	1,332.71	22.65***
	(236.65)	(256.49)	(0.00)
Median List Price	216,445.45	205,233.47	-11,211.99***
	(75,300.37)	(70,603.58)	(0.00)
Days on the Market	109.73	107.64	-2.08***
	(30.66)	(27.31)	(0.00)
Observations	1,096,423	3,684,075	4,780,498

Notes: This table shows the covariate balance table for houses that digitized before and after the median (“Early Digitizers”) or later (“Late Digitizers”). When possible, all statistics are at the census block group level. Information from Zillow is at the zip code level, and the unit of observation is at the house level. All data come from ATTOM Data, ZTRAX, and the US Census.

Table A.4: County Digitization and Algorithmic Investors Buying

VARIABLES	(1) Ln(Q_Algo)	(2) Ln(Q_Algo)	(3) Ln(Q_Algo)
County Digitization	1.130**	0.780**	0.749**
	(0.380)	(0.221)	(0.229)
Observations	3,962	3,962	3,962
R-squared	0.798	0.812	0.816
Year FE	Yes	Yes	Yes
Location FE	Yes	Yes	Yes
SocioEconomics	-	Yes	Yes
Housing Stock	-	-	Yes
DV Mean	2.597	2.597	2.597

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10

Notes: This table shows the results of county-level difference-in-difference regressions estimating the effect of county record digitization on the natural log of houses purchased by algorithmic investors. All specifications include house characteristics, and year and geography fixed effects; standard errors are clustered at the county level. Column 2 includes county population, demographics, poverty, unemployment rate, and educational characteristics. Column 3 adds housing stock characteristics, such as the number of housing units and rent burden. All data comes from ATTOM Data, ZTRAX, the US Census, and county governments.

Table A.5: Algorithmic Investor Purchase, by Homeowner Race, Investor Sample

	(1) Algorithm Purchase	(2) Algorithm Purchase	(3) Algorithm Purchase
Seller Minority	-0.0133***	-0.0137***	-0.0194***
	(0.0036)	(0.0036)	(0.0045)
Digitization x Seller White	0.0079*	0.0075**	0.0042*
	(0.0045)	(0.0035)	(0.0022)
Digitization x Seller Minority	0.0415***	0.0396***	0.0389***
	(0.0043)	(0.0042)	(0.0050)
Geography FE	Tract	Block Group	Block
Year FE	Yes	Yes	Yes
Sample	Investors	Investors	Investors
DV Mean	.0018	.0018	.0018
Observations	898975	898061	802192

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows the results of cross-sectional difference-in-difference regressions estimating the impact of house record digitization on the purchase by an algorithmic investor, by homeowner race. The sample includes only investor purchases, so the coefficients are interpreted as the likelihood of being purchased by an algorithmic investor compared with a human investor. All specifications include house characteristics, year, and geography fixed effects; and standard errors are clustered at the geographic level. All data comes from ATTOM Data, ZTRAX, and county governments.

Table A.6: IV Analysis: Algorithmic Investors and Race Penalty

	(1) First Stage	(2) 2SLS	(3) First Stage	(4) 2SLS	(5) First Stage	(6) 2SLS
Digitization	0.043***		0.046***		0.067***	
	(0.004)		(0.004)		(0.006)	
Algo Buyer		0.289		0.279		0.291
		(0.434)		(0.320)		(0.209)
AlgoxSellerBlack/Hisp		0.526***		0.529***		0.527***
		(0.127)		(0.116)		(0.107)
Geo Level	Tract+Year	Tract+Year	BG+Year	BG+Year	Block+Year	Block+Year
DV Mean	.002	164167	.002	164167	.002	164167
Adj R-squared	.317	.345	.317	.345	.344	.345
Observations	222666	222772	221537	222686	151452	222686

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10

Notes: This table shows the results of cross-sectional 2SLS regressions that estimate the algorithmic investor purchase on the race penalty, instrumenting for the algorithmic purchase with house-level digitization. All specifications include house characteristics, year, and geography fixed effects; and standard errors are clustered at the geographic level and use log sale price as the outcome. All data comes from ATTOM Data, ZTRAX and county governments.

Table A.7: Assessment Margin

VARIABLES	(1) Log(Assess Margin)	(2) Log(Assess Margin)	(3) Log(Assess Margin)	(4) Log(Assess Margin)	(5) Log(Assess Margin)	(6) Log(Assess Margin)	(7) Log(Assess Margin)	(8) Log(Assess Margin)
Seller Minority = 1	0.004	0.004	0.002	0.003	-0.013*	-0.005	0.022**	-0.052***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.007)	(0.008)	(0.010)	(0.012)
Seller Minority x Minority Neighborhood = 1				0.003				0.090***
				(0.006)				(0.015)
Observations	81,387	76,312	46,294	76,312	467,588	440,142	238,501	440,142
R-squared	0.714	0.748	0.820	0.748	0.362	0.470	0.723	0.470
FE	Year x Tract	Year x Block Group	Year x Block	Year x Block Group	Year x Tract	Year x Block Group	Year x Block	Year x Block Group
Assessment Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyers	Algorithms	Algorithms	Algorithms	Algorithms	Humans	Humans	Humans	Humans
DV Mean	0.0911	0.0839	0.0577	0.0839	0.698	0.703	0.795	0.703

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10

Notes: This table shows the difference in the natural log of estimated house market value and the natural log transformation of the price paid, or *assessment margin*. *Minority neighborhood* indicates if the house is in a census block group with an above average minority resident share. All specifications include tax estimate and sale year by geography fixed effects. Standard errors are clustered at the geographic level. All data comes from ATTOM Data.

A Data and Model Appendix