Title

Essays in Labor and Transportation Economics

Author

Ladd, Daniel

Publication Date

2024

UNIVERSITY OF CALIFORNIA,
IRVINE

Essays in Labor and Transportation Economics

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Economics

by

Daniel Ladd

Dissertation Committee:
Distinguished Professor Jan Brueckner, Chair
Distinguished Professor David Neumark
Professor Matthew Freedman

2024

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the invaluable help, mentorship and continued support from a large number of people. I am incredible grateful to all those who have supported me throughout my journey.

I am grateful to my academic mentors, Caitline Myers, Jan Brueckner and David Neumark for this constant guidance and advice. All were able to interest and inspire me in different ways and I would not have gotten to where I am now without their help and support.

I would also like to thank my cohort members who have been a big part of my journey. They helped me survive comps and all aspects of Ph.D. life. Especially I would like to thank Derek Christopher and Daniel Firoozi who were both instrumental in both my research and constantly driving my passion and excitement for economics.

To my parents and sisters, I am eternally grateful for the constant support and unconditional love. To Laura, thank you for being an incredible partner, confidant, and friend. I never would have been able to do this without you and thank you for helping to make me a better version of myself each and every day.

Financial support was provided by the University of California, Irvine and by the Sloan Foundation

VITA

Daniel Ladd

2014	B.A. in Economics and Geography, Middlebury College
2014-2017	Charles River Associates
2017-2020	Teaching Assistant, Economics, University of California, Irvine
2000-02	Research Assistant, Program in Global Peace and Conflict Studies, University of California, Irvine
2019	M.A. in Economics, University of California Irvine
2019-2022	Research Assistant, Department of Economics, University of California, Irvine
2021	Instructor, Department of Economics, University of California, Irvine
2024	Ph.D. in Economics, University of California, Irvine

FIELD OF STUDY

Labor, Urban and Transportation Economics

PUBLICATIONS

2023 *Stereotypes of older workers and perceived ageism in job ads: evidence from an experiment.* Journal of Pension Economics & Finance. (with Ian Burn, Daniel Firoozi, and David Neumark)

2023 *Workplace Injuries and Receipt of Benefits From Workers' Compensation and Social Security Disability Insurance.* Journal of Occupational and Environmental Medicine. (with David Neumark)

2023 *Age Discrimination and Age Stereotypes in Job Ads.* Federal Reserve Bank of San Fransisco Economic Letter. (with Ian Burn, Daniel Firoozi, and David Neumark)

2020 *Did parental involvement laws grow teeth? The effects of state restrictions on minors' access to abortion.* Journal of Health Economics. (with Caitlin Myers)

ABSTRACT OF THE DISSERTATION

Essays on Labor and Transportation Economics

by

Daniel Ladd

Doctor of Philosophy in Social Ecology

University of California, Irvine, 2024

Distinguished Professor Jan Brueckner, Chair

This dissertation tries to answer some important questions in labor and transportation economics in an attempt to broaden understanding of different topics.

In the first chapter "*Correlated Labor Market Risk and Housing Investment*" I show that households have lower levels of housing investment when they live in areas with labor markets that are more correlated with their industry of employment. I find that if a household lives in an area where many other households work in the same or similar industries, then housing may be a riskier asset as it is more correlated with labor market income. In response household decrease their investment in housing, and this This decline is driven by concentrations and riskiness of other correlated industries, suggesting agglomeration in one industry can have negative spillovers to workers of other related industries.

The second chapter "*Loyalty rewards and redemption behavior: Stylized facts for the U.S. airline industry*" joint with Alexander Luttmann, provides a novel identification method for frequent flyer tickets in a comment industry database, DB1B. Using the FFAs we identify, we show how the characteristics of award tickets differ from paid tickets and how these

characteristics have changed over time. We then demonstrate how various market and product quality characteristics influence the share of passengers traveling on FFAs. Finally, we find that price dispersion increases on routes with larger shares of frequent flyer passengers, implying that airline loyalty programs enhance market power.

In the third chapter, "*Help Really Wanted? The Impact of Age Stereotypes in Job Ads on Applications from Older Workers*" joint with Ian Burn, Daniel Firoozi and David Neumark, We construct job ads for administrative assistant, retail, and security guard jobs, using language from real job ads collected in a prior large-scale correspondence study. We modify the job-ad language to randomly vary whether the job ad includes ageist language regarding age-related stereotypes. In contrast to a correspondence study in which job searchers are artificial and researchers study the responses of real employers, in our research the job ads are artificial and we study the responses of real job searchers. We find that job-ad language related to ageist stereotypes, even when the language is not blatantly or specifically age-related, deters older workers from applying for jobs. The change in the age distribution of applicants is large, with significant declines in the average and median age, the 75th percentile of the age distribution, and the share of applicants over 40. Based on these estimates and those from the correspondence study, and the fact that we use real-world ageist job-ad language, we conclude that job-ad language that deters older workers from applying for jobs can have roughly as large an impact on hiring of older workers as direct age discrimination in hiring.

Correlated Labor Market Risk and Housing Investment*

Daniel Ladd

University of California-Irvine

June 14, 2024

Abstract

This paper shows that households have lower levels of housing investment when they live in areas with labor markets that are more correlated with their industry of employment. In other words, if a household lives in an area where many other households work in the same or similar industries, then housing may be a riskier asset as it is more correlated with labor market income. Thus households decrease their investment in housing. Using US microdata from 2007-2017 a one-standard deviation increase in a household's correlated labor market risk is associated with a decline in housing investment by around $6,750. This decline is driven by concentrations and riskiness of other correlated industries, suggesting agglomeration in one industry can have negative spillovers to workers of other related industries.

*I received helpful feedback from Jan Brueckner, Derek Christopher, Edward Coulson, Tom Davidoff, Matt Freedman, Daniel Firoozi, David Neumark, and from members of the Applied Microeconomics Workshop at UC Irvine..

1 Introduction

For households in the United States, housing investment plays an crucial role in wealth accumulation. Data from the Federal Reserve's *Survey of Consumer Finances* highlights that more than half of the median household's net wealth is attributed to single-family residential housing. Additionally, housing investments can greatly influences local government funding through their reliance on property taxes. The Urban Institute's 2016 report estimates notes that property taxes, comprising nearly 30% of local governments' general revenue, generated $487 billion annually.[1] These factors underscore the need to understand key drivers behind the housing investment decisions of households.

Within the economics literature, housing is distinct, serving as both a consumptive good and an investment asset. Research has shown that this dual role that housing plays for households often leads to housing assets comprising a large fraction of a homeowner's overall net worth.[2] Consequently, the choice to rent or buy, and the amount spent on purchasing a home, critically impacts the risk profile of households' investment portfolios. While housing assets can make up the majority of a household's assets, wages derived from individuals' labor are the primary source of income for the vast majority of households. As such, households should adjust their investment decisions in response to both labor market risk[3] and income variability.[4] Rational households will therefore incorporate both overall labor market risk, and the degree of correlation between their labor market income and housing assets when making housing investment decisions.

In any given market, house prices are susceptible to decline when a large number of households decide to sell simultaneously. This scenario often coincides with widespread job losses, diminishing the effectiveness of housing investment as a hedge against labor market risk. Households are likely to be cautious about investing in housing if they perceive a high correlation between their own job security and that of others in their vicinity. Optimal portfolio management for households involves

[1] These taxes were split between residential and commercial properties, but Gravelle and Wallace 2007 suggest that more than 60% of property tax revenue is from residential properties.

[2] See Henderson and Ioannides 1983, Brueckner 1997 and Cocco 2005.

[3] See Betermier et al. 2011, Heaton and Lucas 2000.

[4] See Haurin 1991, Gathergood 2011, Robst et al. 1999, Diaz-Serrano 2005a, Diaz-Serrano 2005b.

evaluating the value of housing capital, and housing is a less valuable investment when its return is correlated with the return on a household's human capital (wages).

Several papers have examined aspects of this housing investment decision faced by households. Ortalo-Magne and Rady (2006) in 2006 built a theoretical life-cycle model of housing investment that, in a special case, predicts when house prices and incomes are positively correlated, households will be more likely to rent. In his 2006 study, Davidoff (2006) expanded on Ortalo-Mange and Rady's model. He utilized data from the 1990 U.S. Census to demonstrate that households living in cities where house prices and incomes in their employment sector are positively correlated tend to invest less in housing.

This paper proposes a mechanism to explain the phenomenon identified by Davidoff, focusing on the concept of correlated labor market risk. This paper puts forward a unqique measure of correlated labor market risk to assess whether households curb their housing investments when their risk of unemployment is correlated with that of others living in their same geographic area. This correlated risk measure is driven by both the concentration and riskiness of an individual's own industry and the concentration and risk associated with other closely related industries within the same geographic area. The measure created in this paper attempts to separate the effect on housing investment of risk arising from a household's own industry and the risk coming from other closely related industries.

While this paper focuses on the role of inter-household correlated labor market risk and its impact on housing investment, several previous papers have looked at the effects of intra-household correlated labor market risk on housing investment. Shore and Sinai (2010) find that in two-income households where individuals work in the same industry, there is greater investment in housing. Jansson (2017) replicates this finding using Swedish microdata, and proposes that the pattern is likely due to those two-income households facing a lower overall risk of any decline in income relative to two-income households with no overlap in industries.[5] Jansson also finds that despite these

[5]Households with individuals that work in the same industry are more likely to lose their jobs at the same time, but households with individuals working in different industries have a higher risk of seeing any decline in income as the risk of unemployment associated with their jobs is less correlated.

same-industry households investing more in housing, they have a lower probability of ownership.

This paper looks at the linkage between inter-household correlated labor market risk and housing investment decisions by combining microdata from multiple US sources to create a measure of correlated labor market risk. This risk measure is calculated at the industry-year-market level. The calculation of the correlated labor market risk measure is described in more detail in the next section of the paper, and the third section describes the data used. The fourth section details the empirical approach used to estimate the relationship between the correlated labor market risk measure and household housing investment decisions. The estimation uses individual-level American Community Survey data to asses the effect of the measure on both the intensive and extensive margins of housing investment. By creating a risk measure for each industry, market and year combination, estimates can be identified even in the presence of industry, market, and state-year fixed effects. The fifth section details the results, which suggest that a one standard deviation increase in the risk measure leads to a reduction in the value of housing owned of approximately $6,750. This effect is primarily driven by the risk coming from households in the same market who work in different, but related industries from the homeowner. The results also imply that a one standard deviation increase in the correlated risk measure leads to a 2% decline in the probability of becoming a homeowner. Additional specifications suggest the magnitude of the effect of correlated risk on housing investment is lower for households with other sources of savings and in states with higher levels of unemployment benefits. The sixth section concludes.

2 Correlated Labor Market Risk Measure

A simple labor market risk measure would incorporate the concentration of various industries in a market interacted with each industry's level of riskiness. This computation would lead to a weighted risk measure of the form:

$$UncorrelatedRiskMeasure_{m,t} = \sum_{j}[Risk_{j,t} * Concen_{j,m,t}] \tag{1}$$

In (1), $Risk_{j,t}$ is the chance that a worker in industry j loses his job in year t and $Concen_{j,m,t}$) is the concentration of the specific industries d in market (m) and year t.[6] Note that this weighted risk measure is specific to a market and time period, but does not vary across industries. Figure 1 displays what this uncorrelated risk measure looks like across PUMAs for the contiguous United States.[7]

By contrast, a *correlated* labor market risk measure aims to identify the risk that, when an individual loses his job, other people in the same market may lose their jobs as well. The risk of many people in the same market losing their jobs will capture the risk of a decline in the local housing market relevant to a particular industry. Thus, the correlated risk measure includes both the risk of unemployment an individual faces from their own employment, as well as the risk of individuals working in the same market also facing unemployment. The likelihood that individual i loses her job when a individual j loses his own job is a function of both how correlated the individuals' industries are and how risky individual j's job is. Finally, these industry risks are weighted by the size of each industry in each particular market. These considerations lead to the following correlated risk measure:

$$RiskMeasure_{d,m,t} = \sum_{j} [Risk_{j,t} * Concen_{j,m,t} * Corr_{d,j}] \tag{2}$$

$$RiskMeasure_{d,m,t} = Risk_{d,t} * Concen_{d,m,t} + \sum_{j \neq d} [Risk_{j,t} * Concen_{j,m,t} * Corr_{d,j}] \tag{3}$$

In (2), the correlated risk measure now includes a term ($Corr_{d,j}$) that gives the correlation[8] between industry j and the given worker's own industry, d. Thus if industry j and d are closely correlated, then a worker in industry j is more likely to lose his job at the same time as a worker in

[6]In this paper, the defined market is at the Public Use Microdata Area (PUMA) level. PUMAs are created from census geography and follow county or census-designated place boundaries. They generally contain between 100,000 and 200,000 residents. PUMAs created using 2000 Census boundaries are used for all analyses that follow. Industry is defined at the 3-digit NAICS code.

[7]The risk measure is expressed in percentages as it is a weighted measure of $Risk_{j,t}$ which is calculated as an annual percentage chance of an individual losing their job while working in industry j.

[8]The correlation between two industries is calculated as the correlation between the 1-year employment changes in industry j and d from 2000-2018. This is described in more detail in a following section.

industry d.[9] The risk measure Equations in (2) and (3) are equivalent as $Corr_{d,d}$ is equal to one.

This risk measure encompasses both the direct effect of labor market risk on a household's housing investment through their own risk of job loss ($Risk_{d,t}$), and the indirect effect of the correlated labor market risk that impacts a household's housing investment decision through potential housing price declines. The risk measure calculated in (3) shows how the effect of a worker's own industry ($Risk_{d,t} * Concen_{d,m,t}$ referred to as "Own Industry Risk" in the rest of the paper) can be separated from the effect of other industries in a worker's market ($\sum_{j \neq d}[Risk_{j,t} * Concen_{j,m,t} * Corr_{d,j}]$ referred to as "Other Industry Risk" in the remainder of the paper). "Own Industry Risk" can be further broken down by including $Risk_{d,t}$ and $Concen_{d,m,t}$ as separate variables in a regression. The correlated risk measure is equal to the sum of the Own Industry Risk and Other Industry Risk terms as seen in Equation (3).

The risk of job loss is calculated nationally for each industry and each year.[10] This risk is weighted by the concentration ($Concen_{j,m,t}$) of the specific industries in each market (m) as in Equation (1). It should be noted that industry concentration in Equations (1), (2) and (3) are computed not for the entire labor market, but instead by residential area (at the PUMA level). The local focus of PUMA-level employment concentrations and correlations directly affect both worker's income and his own home value, which reflects prices in the local(PUMA-level) housing market.

Consider the example of construction workers in a specific Boston Public Use Microdata Area (PUMA). The measure in Equation (3) takes labor market riskiness for construction workers nationally,[11] and weights it by the concentration of construction workers in that PUMA in Boston.[12] This portion of the measure will be higher for jobs with higher inherent levels of unemployment risk, and be higher for individuals who live in local areas with higher concentrations of individuals

[9] Implicitly $Risk_{d,t}$ is weighted by the correlation between industry d and itself, but that correlation is equal to one and is thus not shown in the equation.

[10] The job loss risk of other industries changes each year, but the correlation between industries is assumed constant for the sample period.

[11] Labor market riskiness is calculated as the chance that a worker in that industry loses his job in a given year.

[12] The concentration is calculated as the share of employed individuals in that PUMA working in construction.

working in their same industry.[13] This first part of the measure is then combined with the riskiness of each other industry adjusted by that industry's correlation with the construction industry and weighted by that industry's concentration in that particular PUMA in Boston. Thus, this correlated risk measure will vary across industries, markets, and time, and differ for retail workers in the same Boston PUMA[14] or for construction workers who live in a PUMA in the Hartford, CT metro area.[15]

Figure 2 displays the average risk measure from 2007-2017 across PUMAs weighted by the number of observed households in each industry for each PUMA.[16] Though there are some similarities with the geographic distribution of uncorrelated risk (Figure 1) there is different spatial variation to the average risk measure. Figure 3 shows the average correlated risk measure across the US for a specific industry code (NAICS code 336) corresponding to Transportation Equipment Manufacturing. The geographic patters here show higher risk in areas with higher levels of car manufacturing employment (the Midwest and Tennessee/Alabama) but also show higher levels of risk in areas not traditionally associated with transportation manufacturing (Central Florida and Las Vegas).

Figure 4 shows the geographic distribution of the Other Industry Risk term for the Transportation Equipment Manufacturing industry code. This map shows there are higher risks in areas with closely related industries (northern Indiana/Ohio and Tennessee) as well as areas with high concentrations of other cyclical industries (Nevada and Florida). Figure 5 displays the distribution of Own Industry Risk across the US for the Transportation Equipment Manufacturing industry. This map displays the average own industry risk value across years so it primarily serves as a map of industry concentration.[17] The concentration of manufacturing in the Midwest and south of the country is

[13]Implicitly, this is designed to capture the likelihood that if you have been laid off, it is likely that other individuals working in that same industry are more likely to also be laid off at the same time.

[14]Although they live in the same PUMA, retail workers have a different industry risk level and have different correlations with other industries than construction workers do.

[15]Each of those PUMAs will have different concentrations of both construction workers as well as different shares of other industry employment.

[16]The risk measure is expressed in percentages as it is a weighted measure of $Risk_{j,t}$ which is calculated as a percentage chance of an individual losing a job in industry j. The correlated risk measure is lower than the uncorrelated risk measure as the maximum value $Corr_{d,j}$ can take is one.

[17]$Risk_{d,t}$ is calculated nationally and thus does not vary at the PUMA level.

readily apparent.[18]

The riskiest industries in a given market tend to be Construction or a closely related field.[19] Examples of other risky industries include Furniture and Related Product Manufacturing, Food Services and Drinking Places,and Rental and Leasing Services. The safest industries in a given PUMA are often Nursing and Residential Care Facilities and National Security and International Affairs.[20] Examples of other safe industries include Administration of Human Resource Programs, Administration of Economic Programs, and Utilities.

3 Data

The correlated risk measure from above (Equations 2 and 3) requires measures of industry risk, local industry concentration and cross-industry correlations. These measures are combined into a risk measure calculated at the industry-year-PUMA level and then appended to individual-level data on housing investment. All industries are defined using three-digit NAICS codes. Small industry categories are combined to insure each that classification has a large number of observations,[21] leading to 77 different industry codes being used in the main analysis.

3.1 Industry Risks

The Current Population Survey (CPS) is used to identify an individuals' industry risk. The CPS is a monthly survey of approximately 60,000 US households that follows individuals for four months, after which they are removed from the survey for eight months, and then returned to the survey for four more months. Data from 2000 to 2017 provided by the University of Minnesota's

[18]The concentrations in the Seattle area are mostly due to airplane manufacturing for Boeing, and concentrations in Connecticut and South-Eastern Virginia are due to shipbuilding industries.

[19]Nonmetallic Mineral Product Manufacturing has a correlation of .966 with the construction industry.

[20]National Security and International Affairs are primarily civilian and military employees of the Department of Defense.

[21]Industry subcodes that average fewer than 200,000 employees a month in the Quarterly Census of Employment and Wages (QCEW) from 2000-2018 are merged into related larger subcodes. Examples of small industry codes eliminated include '521-Monetary authorities - central bank', '712-Museums, historical sites, zoos, and parks' and '927-Space research and technology'.

Integrated Public Use Microdata (IPUMS) is used to create a file covering each of the eight months an individual was interviewed. The industry risk is calculated for each industry, where the risk is defined as the probability of having an unemployment spell after working in a given industry.[22] In calculating the unemployment risks associated with working in a given industry the sample is limited to respondents between the ages of 26 and 60.[23] Industry risk measures are calculated for each year from 2000 to 2017 at the national level.

3.2 Industry Concentrations

Industry concentrations are calculated using American Community Survey (ACS) estimates provided by IPUMS at the Public Use Microdata Area (PUMA) level. The American Community Survey is an annual survey that replaced the long form census to provide more timely updates than the decennial census. The industry concentrations are calculated primarily using three-year ACS files, which are available from 2007 to 2013. Since the three-year files were discontinued after 2013, five-year estimates are used from 2014 to 2017.[24] Industry shares are calculated for each PUMA in each year from 2007 to 2017.

3.3 Industry Correlations

Correlations between industries are calculated using the Quarterly Census of Employment and Wages (QCEW) from the Bureau of Labor Statistics. The QCEW is created through the tabulation of employment and wages of establishments that report to the unemployment insurance programs, and it represents around 97% of all wage and salary civilian employment. Using national data from

[22] An individual is categorized as unemployed if he is unemployed, unable to work, or unpaid and working less than 15 hours a week.

[23] This limits the risk of identifying unemployment spells that are due to retirement or schooling instead of labor market risk.

[24] The ACS is given to roughly 1% of the US population each year. However, the Census Bureau recommends using three- or five-year ACS files for aggregation at geographic units when precision is more important than the currency of the data. The Census Bureau weights data collected in the previous three (or five) years to estimate current populations. Since all individuals are making housing investment decisions before they appear in the sample, there is no benefit to using less precise one-year samples. In an alternate specification, 2000 Census data are used as a baseline, and current year industry concentrations are calculated by assuming each industry in each PUMA grew at its respective national industry growth rate (this approach is often referred to as a Bartik shift share).

2000-2018, correlation coefficients for one-year employment growth rates are calculated for each pair of the 77 different industries.[25] The industry correlations are unique for each industry pair, but do not vary by year or across PUMAs.

3.4 Housing Investment Decisions

The primary data source for the level of housing investment comes from the one-year ACS files from IPUMS. The ACS includes measures of an individual's industry, occupation, self-reported income and home value along with individual controls such as marital status, age, education levels, race, ethnicity and PUMA of residence. The data are from the 2007 to 2017 period and are restricted to households that report an individual who is employed, has positive income and is between the ages of 26-60.

3.5 Sample Restrictions

PUMAs are defined by population, and thus can comprise large geographic areas that may not accurately represent local housing or labor markets. PUMAs in the sample are restricted to those within a metropolitan area that contains at least 5 PUMAs.[26] To get a better sense of how individuals make housing decisions in response to labor market decisions,[27] the sample is limited to households that have moved into their current home in the last 9 years.[28] As home values are self reported, more recent homebuyers are likely those with more accurate views of their home value. Since many homeowners do not update housing decisions frequently, limiting to this sample likely focuses on individuals whose current labor market status influenced the housing investment decision.

[25] The percent change in employment accurately captures correlations between job risk across industries. Additionally the one-year growth rate smooths out seasonal variation in industries such as education by comparing employment in a given month to employment in the same month one year prior.

[26] This restriction limits the sample to the 44 largest metropolitan areas, which contain 48.3% of all homeowners and 55.6% of all renters. These large metropolitan areas contain 62.6% of all homeowners and 66.8% of all renters who live in metropolitan areas.

[27] Additionally, this restriction reduces the odds that individuals may have changed jobs or industries after purchasing a home.

[28] 39.1% of all homeowners have moved in the last 9 years and, of those households in larger metropolitan areas, 40.1% have moved in the last 9 years. Only 69.6% of all homeowners report when they moved into their current home.

To protect an individual's privacy, the Census Bureau recodes households that report home values in the top 1% in each state and year. These high-home-value households list a home value equal to the average of those topcoded. Thus the top 1% of home values in each state-year are the same and do not accurately represent housing investment. Additionally, there is likely some geographic correlation within a given state as to where these topcoded households may live. As a result, top-coded households are dropped from the analyses that focus on the value of housing investment but are included in estimations of tenure choice. Finally, each household is assigned the demographic and industry information of the highest wage earner in the household.[29]

The final sample includes 907,695 homeowners and 790,166 renters. Table 1 presents summary statistics for these households. The summary statistics show that homeowners tend to be older, whiter, wealthier and more educated than renters, with higher rates of marriage and larger household sizes. Homeowners appear to work in slightly less risky and less concentrated industries, but there does not appear to be a meaningful gap between the two.

4 Empirical Strategy

Correlated labor market risk can influence an individual's home investment decisions either by reducing the likelihood of purchasing a home (extensive margin), or by reducing the amount invested in a home once one is purchased (intensive margin). Focusing on the intensive margin leads to the following main specification:

$$HouseValue_{i,d,m,s,t} = \alpha + \theta RiskMeasure_{d,m,t} + \beta X_i + \gamma D_d + \psi T_{t,s} + \omega M_m + \nu_i \quad (4)$$

where $HouseValue$ is the reported value of house owned by household i. All regressions account for industry (D_d) and PUMA fixed effects (M_m). Additionally, to control for state specific housing or labor market shocks, state-year fixed effects are added ($T_{t,s}$). X is a vector of individual level

[29]The household is likely to care most about the highest wage earner's correlated labor market risk.

characteristics.[30] Included in the individual specific demographic controls are indicators for race, ethnicity, educational attainment, veteran status,[31] self-employment,[32] marital status, presence of a mortgage as well as controls for age, age squared, number of employed workers in the household, and number of individuals in the household. Additionally, the regression includes controls for indicators of how long a household has lived in their current house. Consistent with prior literature, the analysis should show that when correlated labor market risks faced by a household are higher, that household will invest less in housing as the asset is less of a good hedge against their own labor market income risk.

As detailed above in Equation (3), the correlated risk measure is a linear combination of a household's own industry risk and the level of risk driven by other industries. The coefficient on other industry risk is expected to be negative as higher concentrations of more risky, more closely related industries should increase the risk associated with a household investing in housing. The predicted effect of the own industry risk measure is theoretically ambiguous. Higher levels of own industry risk could reduce a household's level of housing investment through the same channel as that driven by other industry risk, but the constituent components of own industry risk could lead to higher housing investment for a household. For a household employed in an industry with a higher level of local concentration there may exist more chances for career advancement which would increase a household's expectations of future earnings and thus increase housing investment. Additionally, the presence of agglomeration effects within a particular concentrated local industry may also increase a household's expectations of future productivity (and thus earnings) growth.

For notational convenience several additional controls are included in the regression but not shown in Equation (4). To control for within and across-industry variation in housing investment decisions that could be due to the particular job held, indicator variables for two-digit occupation categories are included.[33] Controls for household income cannot be included since the variable of

[30] As stated in the previous section, the characteristics of the individual come from the highest wage earner in the household.

[31] Veterans can qualify for reduced cost loans which may influence their housing investment decision.

[32] It is likely that these individuals may face additional labor market risks not directly tied to their industry.

[33] These variables control for the fact that, although doctors and nurses work in the same industry, they likely have

interest, correlated labor market risk, influences both an individual's housing investment decision and their wage. As income likely influences a household's housing investment decision, the regressions include controls for the median income by industry and the median income by occupation for each metropolitan area.[34]

In addition to looking at the impact of local labor market risk on housing investment conditional on purchasing a house, a separate regression explores how this risk can influence the decision to become a homeowner. For this extensive margin the regression equation is:

$$Homeowner_{i,d,m,s,t} = \alpha + \theta RiskMeasure_{d,m,t} + \beta X_i + \gamma D_d + \psi T_{t,s} + \omega M_m + \nu_i \qquad (5)$$

For this regression the sample includes all homeowners and renters. This regression does not require accurate estimates of an individual's home value, so households with top-coded home values are included in the sample. This regression includes the same controls as those in the home value regression, with the exception of a control for presence of a mortgage.[35] Equation (5) is estimated both as a linear probability model and a probit, with similar results. The linear probability model estimates are shown for ease of convenience. Prior literature would predict that higher correlated labor market risk should decrease the probability of being a homeowner as the benefit of housing as an investment asset is decreased.

5 Results

Of primary interest is the effect of the correlated labor market risk measure on households' housing investments. Table 2 presents a series of three regressions that include the various components of the correlated labor market risk measure. Each coefficient on the risk or concentration measure gives the effect of a one-standard deviation change in that measure. The first column reports the

different levels of housing investment.

[34] Implicitly, one individual's wage should not directly influence the median income in an area, but that individuals wage and the median income of other workers in the same industry and occupation in the same geographic region should be strongly correlated.

[35] The presence of a mortgage is limited to only those households who own a home.

coefficient on the correlated labor market risk measure and shows that for a one-standard deviation increase in a household's correlated labor market risk, households reduce investment in housing by \$6,750. This result is of similar direction and magnitude to Davidoff's (2006) estimate, which showed that a one standard-deviation in his income-price covariance measure led to a decline in housing investment of approximately \$7,500.

The analysis differentiates how households adjust housing investments in response to risks from their own industry versus those driven by the risks and concentrations of other industries. In column 2, the risk measure is divided into 'own' and 'other' industry risks. The findings reveal a substantial, negative, and statistically significant correlation with 'other industry risk,' indicating that households reduce their housing investment in response to higher correlations and riskiness of other industries in the same local market. Conversely, a positive coefficient for 'own industry risk' suggests that households increase their housing investment in response to higher levels of concentration and risk from their own industry.

The third column breaks apart the own industry risk term into its constituent components ($Risk_{d,t}$ and $Concen_{d,m,t}$) and shows there is a positive relationship between the concentration of a given industry in a PUMA and the household's housing investment. The positive result could come from agglomeration effects, where more concentrated industries in a particular PUMA may be more productive due to spillovers. Additionally, households may perceive less risk when there is a large concentration of same-industry jobs as they may feel they have more employment opportunities available if there is a firm specific shock as opposed to an industry specific one. Finally, a large industry concentration may lead to more opportunities for upward advancement on the career ladder, and thus households may have expectations of higher future wages, raising their investment in housing.

All of the regressions in Table 2 include state-year fixed effects and fixed-effects for PUMAs, industry codes and occupation codes as well as indicator variables for number of individuals and number of employed individuals in the household.

The coefficients on the control variables in the Table 2 models match economic intuition.

Households where a female is the highest wage earner are likely to have lower wages (and thus lower investment) due to the gender wage gap in the US. Married individuals invest more in housing as do white households. Increasing levels of education lead to more housing investment, and self-employed individuals also invest more in housing. Mortgages allow households to invest more in housing as they are less credit constrained. Households that have moved more recently invest more in housing than those who have remained in place.[36] Additionally, households with higher wages in their industry or occupation also invest more in housing.

Table 3 shows the results of the linear probability model described in Equation (4). The outcome variable is tenure choice. The homeowners in this sample now include the top-coded individuals who were excluded from the previous regressions on housing investment as the outcome is now tenure choice. Similarly to Table 2, the models in Table 3 start with the effect of correlated labor market risk and suggest that a one-standard deviation increase in the correlated labor market risk measure is associated with a two percentage point decline in the probability that a household is a homeowner. In column 2 results show that both other industry and own industry risk reduce the probability of being a homeowner, and column 3 show the negative effect of own industry risk is primarily drive by changes in $Concen_{d,m,t}$. With the tenure choice model we see there is a negative effect on home ownership probabilities for both own and other industry risk, matching theoretical predictions. Using a probit or logit model for this estimation gives similar results.[37]

The coefficients on the control variables in Table 3 also match economic intuition described above, with a lower probability of home ownership for households with the highest wage earner being a female. Married households are more likely to own a home as are white households. Increasing levels of education lead to higher home ownership and individuals who who have recently moved are less likely to own a home.

[36] It is possible households that have moved less recently are reporting the value of their house when they bought it, as opposed to its current market value which could also explain this decline in home value.

[37] Probit and logit coefficients are available from the author upon request.

5.1 Sources of Bias

Homeowners are a selected sample of households, as a sample of those that purchase or own a home may have different characteristics to those of all households. This potential selection bias can to be controlled for using a two-step procedure. First, a probit on likelihood of owning a home is run, and the results are then used to calculate an Inverse Mills ratio which is included as an additional variables in the baseline housing investment equation. Table 4 displays the coefficients of interest from a Heckman Two-Step Selection model for housing investment. The results in Table 4 show that controlling for this sample selection issue does not appear to have a large impact on the coefficients of interest. The coefficients in Table 4 are nearly identical to those reported in Table 2, and the coefficient on the Inverse Mills Ratio is not statistically significant. As selection does not appear to be biasing the effects of risk measures on housing investment levels, the results presented below do not incorporate this selection procedure.

The regression coefficients shown thus far are reflective of a particular sample of individuals. The sample used is limited to households living in large metropolitan areas,[38] who moved into their homes in the last 9 years.[39] Figure 6 shows the coefficient on the risk measure for a variety of choices of restrictions on how recently households moved in, and how many PUMAs a metropolitan area must have to be included. Moving from left to right across the figure restricts the sample to larger and larger metropolitan areas and leads to a slight increase in the magnitude of the coefficient.[40]. Restricting to more recent movers[41] appears to increase the magnitude of the coefficient as well. The shapes of the coefficient points correspond to their level of statistical significance. The largest triangle corresponds to the regression presented in column 1 of Table 2. The figure suggests that although there is variation in estimated coefficients for different samples, the impact of the correlated

[38]Large metropolitan areas here mean metropolitan areas with at least 5 PUMAS. PUMAs are defined by population, not geography, and thus can comprise large geographic areas that may not accurately represent local housing or labor markets. Limiting to PUMAs in larger cities restricts to PUMAs that encompass smaller geographic areas.

[39]These households are less likely to have changed industry since purchasing their home, and likely provide more accurate information about their own home's current value to the Census.

[40]The coefficients are all negative so a decline in the chart is an increase in the size of the coefficient.

[41]The different color lines each refer to a different restriction on when the household must have moved into their current home.

labor market risk measure is consistently negative and mostly statistically significant.

Another source of potential bias given the data used in these models is that the time period where the American Community Survey is available (2007 to 2017), overlaps with both the great recession and the recovery afterwards. It is plausible that areas with higher levels of labor market risk were more effected by the housing crash and thus those areas would have lower house prices and lower levels of housing investment. This would lead to a larger negative estimate of the effect of the risk measure on housing investment. Although PUMA and state-year fixed effects should absorb most of this potential, the influence of labor market risk on housing investment could have changed in response to the great recession. To identify the effect of the risk measure across time, the first column of Table 5 reports the coefficients of the risk measure interacted with the years in the sample. The coefficients suggest that the effect of risk measure on housing investment actually decreased in magnitude in the immediate aftermath of the great recession. The second column interacts own industry and other industry risk with year as well. The results suggest that the effects are relatively stable across time.

There is some concern about potential endogeneity in estimating the effect of the correlated risk measure on housing investment. Such endogeneity might arise if households choose their employment location or industry based on available housing investment options. Tables 6 and 7 present ways to address two different potential sources of endogeneity. One source of bias is that individuals in certain areas may change jobs in response to this labor market risk, or the effect of risk could change the composition of people moving to the area. Both effects could result in places where people want to invest more in housing to adjust industry concentrations and increase the proportion of safer industries. This would potentially bias the estimates of the effect of the risk measure to be more negative as places with higher housing investment would have lower levels of risk. Table 6 addresses this concern by calculating the risk measure using industry concentrations from the 2000 Census and national industry growth rates. This method assumes that various industries grow at the same level in each local market as they do nationwide. As with the other models, the households span from 2007-2017 and have all moved into their current homes in the last 9 years. Thus nearly all households in this sample purchased their home after the 2000 Census. Table 6 suggests that the effect of the correlated risk measure on housing investment is similar to or perhaps slightly larger in magnitude than that estimated in the baseline specification.

Another potential source of endogeneity is more closely related to the individual homeowner. Individuals with unobserved preferences for greater housing consumption could self-select into industries (or local markets) with lower levels of correlated labor market risk. This self-selection would increase the estimated coefficient of the labor market risk variable as the unobserved housing demand would be negatively correlated with the risk measure. To address this concern, Table 7 presents results limited to households where the highest wage earner is 35 or younger and has a college degree. Chen and Rosenthal (2008) provide evidence that these young, educated individuals primarily move to locations with more favorable business environments to access more and better jobs over moving to areas with greater amenities or lower house prices. This results strongly suggests that these individuals are more likely to choose a job first and then optimize housing investment thus mitigating the endogeneity discussed above. The results in Table 7 show that these individuals are, if anything, more responsive to labor market risks and reduce their housing investment by over $9,400 with a one standard-deviation increase the correlated risk measure.

Investments in housing serve both as a hedge against labor market risk, but also as an investment generally. To the extent that locations with higher average risk lead to lower levels of housing investment the correlated labor market risk measure could be capturing the effect of households investing less in housing due to its lower return. Although the inclusion of PUMA fixed effects should capture most of the across-industry risk of housing investment in an area, Table 8 includes a measure of uncorrelated labor market risk (defined in Equation 1). The smaller coefficient on the risk measure indicates it may have been capturing some of this effect. The coefficient is still negative and of a similar magnitude to previous specifications. The coefficient on the uncorrelated risk measure is large and negative suggesting that households do invest less in housing in riskier markets. The results from column 2 suggest that most of the decline in the coefficient on the risk measure is coming from the other industry risk coefficient which declines relative to the baseline model (Table 2) while the own industry risk coefficient stays stable.

Similarly, a household that has lived in a house for a while should be less concerned about the impact of the risk measure on their housing investment.[42] These households should still worry about the effect of the uncorrelated risk measure as housing is still an investment for them. Table 9 Panel A shows the coefficients on the risk measure for households that have lived in their homes for different periods of time. Households that have lived in a house for longer have a smaller effect of the risk measure. Panel B of Table 9 includes the uncorrelated risk measure into the same regressions from Panel A. Including the uncorrelated risk measure reduces the magnitude of the coefficient on the risk measure, but it stays negative and statistically significant for more recent movers (column 1). While the effect of the risk measure decreases over the length of time living in a home, the coefficient on the uncorrelated risk measure stays more consistent across households.

5.2 Alternate Specifications

The specifications presented thus far allow for risk measures calculated to vary at the PUMA level. This is the smallest geographic level of aggregation publicly available and assumes households care about the labor market risk to housing caused by the people living in close vicinity to themselves.[43] As PUMAs are relatively arbitrary geographic designations it may be better to think about this relationship occurring at the metropolitan level. Table 10 reports the effects of risk measures calculated at the MSA level on housing investment. The results are noisier, but consistent with prior results presented suggesting a one-standard deviation increase in the MSA-calculated risk measure leads to a decline in housing investment of $7,770.

Instead of using individual level data, Table 11 reports coefficients on the effect of the risk

[42] These households have built up higher levels of home equity, giving them larger a cushion to support a job loss. Additionally with time households could change jobs or adjust housing investment levels (through remodeling or an addition) to reduce the impact of the risk measure. Finally these households may reflect housing investment decisions that predate the more currently calculated correlated risk measure and thus would not be sensitive to them.

[43] That is, households care more about the risk to local housing markets posed by the chance that people who live in their more immediate area lose their jobs and/or sell their houses. The regressions implicitly assume labor markets are at the city level by including metropolitan area average industry and occupation wages. The assumption is the value of your house depends more on the value of other houses in a more local area , but wages are competitive across the entire city.

measure for collapsed PUMA-Industry-Year cells. The previous regressions implicitly give weight to areas with larger populations[44] and industries with higher levels of employment. The results from Table 11 suggest that these factors serve to reduce the magnitudes of the estimated coefficients as a one-standard deviation increase in the risk measure is now associated with a decrease in housing investment of $11,586. Table 12 combines the approaches of Tables 10 and 11 by calculating the effect of the risk measure for collapsed MSA-Industry-Year cells. The sample size is greatly reduced and the results are not statistically significant, but the magnitudes of the estimated coefficients are consistent with the previous estimates (-7,558 vs. -6,749 in the baseline specification). Finally Table 13 reports results from a regression on collapsed PUMA-Industry-Year cells but for all metropolitan areas.[45] As in the baseline specifications including smaller metropolitan areas reduces the magnitude of the coefficient on the risk measure,[46] but the results tell a similar story.

5.3 Heterogeneous Effects

The results above suggest that households adjust their housing investment in response to correlated labor market risks. However not all households should react in the same way to higher levels of risk. Households that derive a large fraction of their earnings from the labor market will be more responsive to correlated labor market risk. Conversely, households will be less concerned with correlated labor market risk when they have other forms of savings to rely on if they lose their job. Columns 1 and 2 of Table 14 restrict the sample to households that only report wage income. This restriction reduces the sample size,[47] and the coefficient on the overall correlated labor market risk measure increases slightly in comparison to the baseline sample. Similarly, when restricting to households with no investment income (columns 3 and 4) the coefficient on the overall correlated labor market risk measure is still similar to the baseline model. However, when limiting

[44] Given that PUMAs are created using population this weights results more for PUMAs that have seen greater levels of population growth.

[45] Table 11 was limited to the same set of large metropolitan areas as all other regressions.

[46] See Figure 6.

[47] The number of observations fall from 879,398 households to 461,980.

to households that have positive investment income[48] the coefficient on the risk measure declines greatly and is no longer statistically significant. Crucially, this decline seems primarily driven by the decline in the coefficient on other industry risk, as the effect of own industry risk does not appear to drastically change across samples.

The results from Table 14 suggest that households that have alternate sources of income are less likely to adjust their housing investment in response to correlated labor market risk. When these households lose a job, they are less concerned about the effects on local house prices as they do not need to capitalize their housing investment to provide for living expenses. Thus households with savings or investments are insured to a degree against the threat of job loss. Unemployment insurance provides a similar guarantee. In the United States, unemployment insurance is provided by the states, with each state choosing its own level of income provided in the event of a job loss and the duration for which benefits can be received. Thus each state has a different maximum benefit amount that can be obtained through unemployment insurance.[49] Table 15 shows the effect of the risk measures on housing investment in states with higher and lower levels of maximum unemployment insurance benefits. States with higher levels of benefits (columns 1, 3, 5, and 7) have smaller effects of correlated labor market risk on housing investment, while states with lower levels of benefits see higher effects. The differences in coefficients match with the theory that households change their housing investment less when they have alternative sources of living expenses in the event of job loss.

6 Conclusion

For most households, housing comprises their single largest asset and wages provide the vast majority of their income. Previous literature has shown that when incomes and house prices

[48]Positive investment income indicates the household likely has positive savings or investments that generate this income.

[49]The maximum unemployment benefit is calculated as the maximum weekly benefit multiplied by the maximum number of weeks covered. The maximum weekly benefit is calculated as a fraction of income up to an income cap. As the sample here is limited to homeowners in large metropolitan areas, the vast majority of households are above the income cap (average wage income in the sample of owners is over $88,000).

are correlated, households invest less in housing. This paper builds off those results to suggest that households incorporate correlated labor market risk into their housing investment decisions. The results presented reveal a significant decrease in housing investment of around $6,750 for households facing a one-standard deviation increase in their estimated correlated labor market risk. Additionally the model estimates that households adjust on the extensive margin and that a one-standard deviation increase in correlated labor market risk is associated with a 2% decline in the probability of a household owning a home. Intriguingly, this decrease appears largely driven by risks stemming from sectors outside the household's own industry.

The study also highlights the complexities municipalities face in attracting businesses. Municipalities are increasingly using targeted incentives to attract particular businesses to their area. Though the results presented here do not contradict benefits from agglomeration, they do suggest an unintended side effect. As concentration in one industry grows, the correlated labor market risk for households in other closely related industries increases. Municipalities derive large fractions of their revenue from property taxes, and the increase in tax revenue from bringing in outside companies in one particular industry or sector may be offset to a degree by decreases in housing investment by households in other closely related industries. The results of this paper suggest there may be benefits to these municipalities instead focusing on diversifying their industry concentrations to increase home values (and thus tax revenue) for a larger share of their residents.

Finally, these results also suggest there may be underprovision of unemployment insurance available for households. Households may use housing investment as a hedge against loss in labor market income, but this hedge is less effective if labor market risks are highly correlated. In states with higher levels of unemployment insurance, there is suggestive evidence that households' housing investment decisions are less driven by correlated labor market risk and may lead to a more optimal household investment portfolios.

References

[1] Abowd, J.M. and O.C. Ashenfelter. 1981, "Anticipated unemployment, temporary layoffs, and compensating wage differentials", *Sherwin Rosen Studies in Labor Markets*, National Bureau of Economic Research, 141-186.

[2] Betermier, Sebastien and Jansson, Thomas and Parlour, Christine A. and Walden, Johan, 2011, "Hedging Labor Income Risk", *Riksbank Research Paper Series*, No. 86.

[3] Brueckner, Jan K. 1997, "Consumption and Investment Motives and the Portfolio Choices of Homeowners", *Journal of Real Estate Finance and Economics*, 15(2), 159-180.

[4] Chen, Yong and Rosenthal, Stuart 2008, "Local amenities and life-cycle migration: Do people move for jobs or fun?", *Journal of Urban Economics*, 64(3), 519-537.

[5] Cocco, Jao F. 2005, "Portfolio Choice in the Presence of Housing", *The Review of Financial Studies*, 18(2), 535-567.

[6] Cubas, German and Pedro Silos, 2017, "Career Choice and the Risk Premium in the Labor Market", *Review of Economic Dynamics*, 18(2), 535-567.

[7] Davidoff, Thomas, 2006, "Labor income, housing prices, and homeownership", *Journal of Urban Economics*, 59, 209-235.

[8] Diaz-Serrano, L., 2005, "On the negative relationship between labor income uncertainty and homeownership: risk-aversion vs. credit constraints", *Journal of Housing Economics*, 14, 109-126.

[9] Diaz-Serrano, L., 2005, "Labour income uncertainty, skewness and homeownership: a panel data study for Germany and Spain", *Journal of Urban Economics*, 58, 156-176.

[10] Feinberg, R.M. 1981, "Earnings-risk as a compensating differential", *Southern Economic Journal*, 48(1), 156-163.

[11] Flavin, Marjorie and Takashi Yamashita, 2002, "Owner-Occupied Housing and the Composition of the Household Portfolio", *American Economic Review*, 92(1), 345-362.

[12] Gathergood, John, 2011, "Unemployment risk, house price risk and the transition into home ownership in the United Kingdom", *Journal of Housing Economics*, 20(3), 200-209.

[13] Gravelle, Jennifer and Sally Wallace. 2009. "Introduction and Overview: Erosion of the Property Tax Base: Trends, Causes, Consequences." Washington, DC: Lincoln Institute of Land Policy and George Washington University Property Tax Roundtable Proceedings.

[14] Haurin, Donald R. 1991, "Income variability, homeownership, and housing demand", *Journal of Housing Economics*, 1(1), 60-74.

[15] Heaton, John and Deborach Lucas, 2000, "Portfolio Choice and Asset Prices: The Importance of Entrepreneurial Risk", *Journal of Finance*, 55(3), 1163-1198.

[16] Henderson, J.V., Ioannides, Y.M. 1983, "A model of housing tenure choice", *American Economic Review*, 73(1), 98-113.

[17] Jansson, Thomas, 2017, "Housing choices and labor income risk", *Journal of Urban Economics*, 99, 107-119.

[18] Leigh, J.P. 1983, "Job choice across industries when earnings are uncertain", *Quarterly Review of Economics and Business*, 23(3), 54-69.

[19] Ortalo-Magne, Francois and Sven Rady, 2006, "Housing Market Dynamics: On the Contribution of Income Shocks and Credit Constraints", *The Review of Economic Studies*, 73(2), 459-485.

[20] Robst, J and R. Deitz and K. McGoldrick, 1999 "Income variability, uncertainty and housing tenure choice", *Regional Science and Urban Economics*, 29, 219-229.

[21] Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek, 2019, IPUMS USA: Version 9.0 [dataset]. Minneapolis, MN: IPUMS, https://doi.org/10.18128/D010.V9.0

[22] Shore, Stephen H. and Todd Sinai, 2006, "Commitment, Risk, And Consumption: Do Birds of a Feather Have Bigger Nests?", *The Review of Economics and Statistics*, 92(2), 408-424.

Table 1
Summary Statistics

	Owners		Renters	
	mean	s.d.	mean	s.d.
Economic Data				
Average Home Value/Monthly Rent	$368,830	$293,887	$1,179.71	$648.39
Household Income	$133,589	$109,277	$69,433	$67,664
Wage Income	$91,638	$85,373	$52,305	$54,977
Labor Market Conditions in Percents				
Risk Measure	2.77	2.00	2.83	2.10
Other Industry Risk	2.48	1.94	2.50	2.04
Own Industry Risk	0.29	0.31	0.33	0.33
$Risk_{d,t}$	7.69	3.23	8.34	3.34
$Concen_{d,m,t}$	3.71	3.24	3.95	3.50
Demographic Controls				
Age	42.37	9.07	39.51	9.58
White	0.77	0.37	0.61	0.47
Black	0.09	0.28	0.20	0.40
College	0.56	0.49	0.39	0.45
Male	0.63	0.48	0.53	0.50
Self Employed	0.08	0.28	0.07	0.26
Veteran	0.07	0.25	0.05	0.21
Household Composition				
Fraction Married	0.69	0.46	0.32	0.47
Household Size	3.07	1.50	2.41	1.51
Employed Workers	1.71	0.70	1.40	0.66
Fraction with Children	0.59	0.49	0.43	0.50

Notes: Household-weighted summary statistics calculated for the highest wage earner in the household from 2007-2017. All dollars are adjusted to 2017 dollars. Data comes from 2007-2017 one-year ACS samples from the US Census Bureau.

Table 2
OLS Regression of Risk on Housing Investment

	(1)	(2)	(3)
Risk Measure	-5,846**		
	(2,746)		
Other Industry Risk		-11,774***	-12,521***
		(2,524)	(2,543)
Own Industry Risk		4,897***	
		(1,233)	
$Risk_{d,t}$			2,016*
			(1,151)
$Concen_{d,m,t}$			5,498***
			(1,152)
Female	-15,964***	-15,955***	-15,945***
	(739)	(738)	(738)
Age	12,999***	12,994***	13,001***
	(393)	(393)	(394)
Age Squared	-104***	-104***	-104***
	(4)	(4)	(4)
Married	42,857***	42,829***	42,824***
	(1,187)	(1,186)	(1,187)
Veteran	-21,148***	-21,121***	-21,099***
	(1,257)	(1,254)	(1,251)
Black	-45,593***	-45,609***	-45,611***
	(3,147)	(3,146)	(3,143)
Native American	-40,776***	-40,579***	-40,489***
	(4,062)	(4,052)	(4,047)
Asian	-16,208***	-16,332***	-16,336***
	(3,400)	(3,397)	(3,394)
Other Race	-15,063***	-15,045***	-15,049***
	(2,912)	(2,922)	(2,922)
Two or More Races	-23,072***	-22,991***	-22,944***
	(2,158)	(2,155)	(2,151)
Not Hispanic	49,354***	49,455***	49,430***
	(3,386)	(3,391)	(3,389)
GED	30,639***	30,770***	30,728***
	(2,911)	(2,910)	(2,886)
HS Grad	23,134***	23,525***	23,501***
	(1,740)	(1,731)	(1,731)
Some College	38,147***	38,556***	38,545***
	(1,780)	(1,770)	(1,765)
Associate's Degree	44,782***	45,196***	45,204***
	(2,080)	(2,072)	(2,069)
Bachelor's Degree	85,511***	85,915***	85,954***
	(2,863)	(2,863)	(2,863)
Advanced Degree	122,810***	123,067***	123,043***
	(3,994)	(3,999)	(3,998)
Mortgage	23,690***	23,680***	23,671***
	(2,038)	(2,037)	(2,035)
Moved in last 2 years	-1,666*	-1,667*	-1,655*
	(864)	(863)	(863)
Moved in last 4 years	-12,142***	-12,169***	-12,166***
	(879)	(877)	(877)
Moved in last 9 years	-28,701***	-28,737***	-28,727***
	(1,022)	(1,021)	(1,021)
Median Metro Ind Wage (000s)	1,064***	1,012***	983***
	(103)	(103)	(103)
Median Metro Occ Wage (000s)	992***	991***	989***
	(37)	(37)	(37)
Self Employed	89,343***	89,502***	89,509***
	(3,993)	(3,994)	(3,989)
State-Year F.E.	Yes	Yes	Yes
PUMA F.E.	Yes	Yes	Yes
Industry F.E.	Yes	Yes	Yes
Occupation F.E.	Yes	Yes	Yes
Observations	1,225,902	1,225,902	1,225,902
R-squared	.54	.54	.54

Robust standard errors in parentheses clustered at the PUMA level. Data is limited to employed homeowners that are the highest wage earner in their household, who are located in a major metropolitan area and moved into their current home in the last 9 years. *$p < 0.10$ **$p < 0.05$ ***$p < 0.01$.

Table 3
Linear Probability Model of Risk on Tenure Choice

	(1)	(2)	(3)
Risk Measure	-0.01866***		
	(0.00336)		
Other Industry Risk		-0.01371***	-0.01350***
		(0.00320)	(0.00322)
Own Industry Risk		-0.00701***	
		(0.00147)	
$Risk_{d,t}$			0.00134
			(0.00189)
$Concen_{d,m,t}$			-0.00808***
			(0.00130)
Female	0.00044	0.00042	0.00040
	(0.00112)	(0.00112)	(0.00112)
Age	0.01796***	0.01794***	0.01792***
	(0.00069)	(0.00068)	(0.00068)
Age Squared	-0.00015***	-0.00015***	-0.00015***
	(0.00001)	(0.00001)	(0.00001)
Married	0.14192***	0.14194***	0.14195***
	(0.00202)	(0.00202)	(0.00202)
Veteran	-0.01169***	-0.01171***	-0.01175***
	(0.00208)	(0.00208)	(0.00208)
Black	-0.15357***	-0.15357***	-0.15356***
	(0.00357)	(0.00357)	(0.00357)
Native American	-0.05351***	-0.05366***	-0.05375***
	(0.00650)	(0.00650)	(0.00650)
Asian	-0.03467***	-0.03458***	-0.03459***
	(0.00543)	(0.00543)	(0.00542)
Other Race	-0.03496***	-0.03497***	-0.03495***
	(0.00327)	(0.00327)	(0.00327)
Two or More Races	-0.04361***	-0.04368***	-0.04384***
	(0.00348)	(0.00348)	(0.00348)
Not Hispanic	0.06311***	0.06306***	0.06309***
	(0.00312)	(0.00314)	(0.00314)
GED	0.04805***	0.04793***	0.04816***
	(0.00358)	(0.00359)	(0.00360)
HS Grad	0.04102***	0.04076***	0.04073***
	(0.00231)	(0.00231)	(0.00232)
Some College	0.07694***	0.07666***	0.07664***
	(0.00259)	(0.00259)	(0.00260)
Associate's Degree	0.10964***	0.10936***	0.10929***
	(0.00308)	(0.00308)	(0.00309)
Bachelor's Degree	0.16397***	0.16370***	0.16359***
	(0.00326)	(0.00325)	(0.00326)
Advanced Degree	0.19006***	0.18991***	0.18990***
	(0.00366)	(0.00366)	(0.00366)
Moved in last 2 years	0.04015***	0.04014***	0.04013***
	(0.00235)	(0.00235)	(0.00235)
Moved in last 4 years	0.13076***	0.13076***	0.13075***
	(0.00270)	(0.00270)	(0.00270)
Moved in last 9 years	0.29469***	0.29469***	0.29464***
	(0.00501)	(0.00501)	(0.00501)
Median Metro Ind Wage (000s)	0.00103***	0.00107***	0.00112***
	(0.00009)	(0.00008)	(0.00008)
Median Metro Occ Wage (000s)	0.00127***	0.00127***	0.00128***
	(0.00003)	(0.00003)	(0.00003)
Self Employed	0.08717***	0.08703***	0.08701***
	(0.00217)	(0.00217)	(0.00217)
State-Year F.E.	Yes	Yes	Yes
PUMA F.E.	Yes	Yes	Yes
Industry F.E.	Yes	Yes	Yes
Occupation F.E.	Yes	Yes	Yes
Observations	2,210,166	2,210,166	2,210,166
R-squared	.34	.34	.34

Robust standard errors in parentheses clustered at the PUMA level. Data is limited to employed homeowners that are the highest wage earner in their household, who are located in a major metropolitan area and moved into their current home in the last 9 years. $*p < 0.10$ $**p < 0.05$ $***p < 0.01$.

Table 4
OLS Regression of Risk on Housing Investment: Heckman Sample-Selection Results

	(1)	(2)	(3)
Risk Measure	-5,458*		
	(2,901)		
Other Industry Risk		-11,518***	-12,281***
		(2,586)	(2,606)
Own Industry Risk		5,087***	
		(1,326)	
$Risk_{d,t}$			2,013*
			(1,152)
$Concen_{d,m,t}$			5,683***
			(1,242)
Inverse Mills Ratio	-10,700	-11,021	-10,903
	(10,840)	(10,833)	(10,817)
State-Year F.E.	Yes	Yes	Yes
PUMA F.E.	Yes	Yes	Yes
Industry F.E.	Yes	Yes	Yes
Occupation F.E.	Yes	Yes	Yes
Observations	1,225,901	1,225,901	1,225,901
R-squared	.54	.54	.54

Robust standard errors in parentheses clustered at the PUMA level. Regressions include controls for age, age squared, marital status, race, education, ethnicity, self employment status and veteran status of the household's highest wage earner. Regressions also include controls for presence of a mortgage, years since moved into home, metropolitan median wage by industry and occupation. Data is limited to employed homeowners that are the highest wage earner in their household, who are located in a major metropolitan area and moved into their current home in the last 9 years. $^{*}p < 0.10$ $^{**}p < 0.05$ $^{***}p < 0.01$.

Table 5
Housing Investment with Risk Interacted with Year

	(1)		(2)	
Risk Measure				
2007	-10,521***	3,684		
2008	-6,455	3,946		
2009	-5,282	3,697		
2010	-7,066**	3,440		
2011	-8,108**	3,313		
2012	-6,960**	3,188		
2013	-6,969**	3,261		
2014	-6,756*	3,629		
2015	-8,017**	3,873		
2016	-9,083**	4,279		
2017	-9,718**	4,386		
Other Industry Risk				
2007			-17,835***	3,419
2008			-14,622***	3,600
2009			-13,018***	3,411
2010			-14,335***	3,166
2011			-14,782***	3,055
2012			-13,270***	2,949
2013			-13,392***	2,986
2014			-13,713***	3,281
2015			-15,211***	3,502
2016			-16,254***	3,849
2017			-17,403***	3,949
Own Industry Risk				
2007			3,266**	1,400
2008			6,079***	1,565
2009			6,087***	1,476
2010			5,634***	1,393
2011			4,164***	1,255
2012			3,491***	1,167
2013			4,371***	1,347
2014			5,770***	1,728
2015			4,863***	1,653
2016			3,434*	1,867
2017			4,991**	2,276
State-Year F.E.	Yes		Yes	
PUMA F.E.	Yes		Yes	
Industry F.E.	Yes		Yes	
Occupation F.E.	Yes		Yes	
Observations	879,398		879,398	
R-squared	.54		.54	

Robust standard errors in parentheses clustered at the PUMA level. Data is limited to employed homeowners that are the highest wage earner in their household, who are located in a major metropolitan area and moved into their current home in the last 9 years. $*p < 0.10$ $**p < 0.05$ $***p < 0.01$.

Table 6
OLS Regression of Risk on Housing Investment: Industry Concentration from 2000 Census

	(1)	(2)	(3)
Risk Measure	-8,772***		
	(2,837)		
Other Industry Risk		-12,983***	-13,667***
		(2,745)	(2,767)
Own Industry Risk		2,887***	
		(1,106)	
$Risk_{d,t}$			2,276*
			(1,235)
$Concen_{d,m,t}$			3,616***
			(1,063)
State-Year F.E.	Yes	Yes	Yes
PUMA F.E.	Yes	Yes	Yes
Industry F.E.	Yes	Yes	Yes
Occupation F.E.	Yes	Yes	Yes
Observations	879,398	879,398	879,398
R-squared	.54	.54	.54

Robust standard errors in parentheses clustered at the PUMA level. Regressions include controls for age, age squared, marital status, race, education, ethnicity, self employment status and veteran status of the household's highest wage earner. Data is limited to employed homeowners that are the highest wage earner in their household, who are located in a major metropolitan area, moved into their current home in the last 9 years are 35 or younger and have a college degree. *$p < 0.10$ **$p < 0.05$ ***$p < 0.01$.

Table 7
OLS Regression of Risk on Housing Investment: Restricted to Age<= 35 and College Degrees

	(1)	(2)	(3)
Risk Measure	-9,402**		
	(3,965)		
Other Industry Risk		-13,402***	-13,086***
		(3,938)	(3,970)
Own Industry Risk		3,145**	
		(1,586)	
$Risk_{d,t}$			-1,978
			(2,671)
$Concen_{d,m,t}$			2,477**
			(1,206)
State-Year F.E.	Yes	Yes	Yes
MSA F.E.	Yes	Yes	Yes
Industry F.E.	Yes	Yes	Yes
Occupation F.E.	No	No	No
Observations	146,012	146,012	146,012
R-squared	.54	.54	.54

Robust standard errors in parentheses clustered at the PUMA level. Regressions include controls for age, age squared, marital status, race, education, ethnicity, self employment status and veteran status of the household's highest wage earner. Data is limited to employed homeowners that are the highest wage earner in their household, who are located in a major metropolitan area and moved into their current home in the last 9 years. $*p < 0.10$ $**p < 0.05$ $***p < 0.01$.

Table 8
OLS Regression of Risk on Housing Investment: Including Uncorrelated Risk

	(1)	(2)	(3)
Risk Measure	-5,916*		
	(3,124)		
Other Industry Risk		-11,606***	-12,588***
		(2,799)	(2,841)
Own Industry Risk		4,661***	
		(1,267)	
$Risk_{d,t}$			2,804**
			(1,294)
$Concen_{d,m,t}$			5,197***
			(1,171)
Uncorrelated Risk	-11,702**	-11,159*	-10,535*
	(5,732)	(5,697)	(5,636)
State-Year F.E.	Yes	Yes	Yes
PUMA F.E.	Yes	Yes	Yes
Industry F.E.	Yes	Yes	Yes
Occupation F.E.	Yes	Yes	Yes
Observations	879,398	879,398	879,398
R-squared	.54	.54	.54

Robust standard errors in parentheses clustered at the PUMA level. Regressions include controls for age, age squared, marital status, race, education, ethnicity, self employment status and veteran status of the household's highest wage earner. Data is limited to employed homeowners that are the highest wage earner in their household, who are located in a major metropolitan area and moved into their current home in the last 9 years. $*p < 0.10$ $**p < 0.05$ $***p < 0.01$.

Table 9
OLS Regression of Risk on Housing Investment: When Households Moved into Current Home

	(1)	(2)	(3)	(4)
Panel A:				
Risk Measure	-7,857**	-6,749**	-3,209	-3,141
	(3,242)	(3,064)	(2,606)	(2,125)
Panel B:				
Risk Measure	-6,762**	-5,916*	-2,475	-2,189
	(3,293)	(3,124)	(2,665)	(2,187)
Uncorrelated Risk	-15,447**	-11,702**	-10,326*	-13,712**
	(6,268)	(5,732)	(5,883)	(5,658)
Moved In	0-4 Years Ago	0-9 Years Ago	10-19 Years Ago	10+ Years Ago
State-Year F.E.	Yes	Yes	Yes	Yes
PUMA F.E.	Yes	Yes	Yes	Yes
Industry F.E.	Yes	Yes	Yes	Yes
Occupation F.E.	Yes	Yes	Yes	Yes
Observations	487,483	879,398	475,187	715,865
R-squared	.54	.54	.55	.55

Robust standard errors in parentheses clustered at the PUMA level. Regressions include controls for age, age squared, marital status, race, education, ethnicity, self employment status and veteran status of the household's highest wage earner. Data is limited to employed homeowners that are the highest wage earner in their household, who are located in a major metropolitan area and moved into their current home in the last 9 years. $*p < 0.10$ $**p < 0.05$ $***p < 0.01$.

Table 10
OLS Regression of Risk on Housing Investment: Industry Concentration at MSA Level

	(1)	(2)	(3)
Risk Measure	-7,770**		
	(3,893)		
Other Industry Risk		-10,953***	-11,919***
		(3,462)	(3,648)
Own Industry Risk		2,599	
		(1,645)	
$Risk_{d,t}$			2,354*
			(1,361)
$Concen_{d,m,t}$			4,310***
			(1,574)
State-Year F.E.	Yes	Yes	Yes
PUMA F.E.	Yes	Yes	Yes
Industry F.E.	Yes	Yes	Yes
Occupation F.E.	Yes	Yes	Yes
Observations	879,398	879,398	879,398
R-squared	.54	.54	.54

Robust standard errors in parentheses clustered at the PUMA level. Regressions include controls for age, age squared, marital status, race, education, ethnicity, self employment status and veteran status of the household's highest wage earner. Data is limited to employed homeowners that are the highest wage earner in their household, who are located in a major metropolitan area and moved into their current home in the last 9 years. *$p < 0.10$ **$p < 0.05$ ***$p < 0.01$.

Table 11
OLS Regression of Risk on Housing Investment: Collapsed to PUMA-Industry-Year Level

	(1)	(2)	(3)
Risk Measure	-11,586***		
	(2,748)		
Other Industry Risk		-16,370***	-16,963***
		(2,809)	(2,836)
Own Industry Risk		6,648***	
		(1,390)	
$Risk_{d,t}$			72
			(1,716)
$Concen_{d,m,t}$			8,018***
			(1,283)
State-Year F.E.	Yes	Yes	Yes
PUMA F.E.	Yes	Yes	Yes
Industry F.E.	Yes	Yes	Yes
Occupation F.E.	No	No	No
Observations	226,467	226,467	226,467
R-squared	.62	.62	.62

Robust standard errors in parentheses clustered at the PUMA level. Data is limited to employed homeowners that are the highest wage earner in their household, who are located in a major metropolitan area and moved into their current home in the last 9 years. Household data is collapsed to Industry-PUMA-Year cells. Regressions include controls for average age, age squared, marital status, race, education, ethnicity, self employment status and veteran status of the Industry-PUMA-Year cell. *$p < 0.10$ **$p < 0.05$ ***$p < 0.01$.

Table 12

OLS Regression of Risk on Housing Investment: Collapsed to MSA-Industry-Year Level

	(1)	(2)	(3)
Risk Measure	-7,558		
	(5,429)		
Other Industry Risk		-10,145*	-10,590*
		(5,672)	(5,871)
Own Industry Risk		6,510*	
		(3,461)	
$Risk_{d,t}$			674
			(2,513)
$Concen_{d,m,t}$			5,843*
			(3,133)
State-Year F.E.	Yes	Yes	Yes
MSA F.E.	Yes	Yes	Yes
Industry F.E.	Yes	Yes	Yes
Occupation F.E.	No	No	No
Observations	44,942	44,942	44,942
R-squared	.71	.71	.71

Robust standard errors in parentheses clustered at the MSA level. Data is limited to employed homeowners that are the highest wage earner in their household, who are located in a metropolitan area and moved into their current home in the last 9 years. Household data is collapsed to Industry-MSA-Year cells. Risk measures calculated at MSA level. Regressions include controls for average age, age squared, marital status, race, education, ethnicity, self employment status and veteran status of the Industry-MSA-Year cell. $*p < 0.10$ $**p < 0.05$ $***p < 0.01$.

Table 13
OLS Regression of Risk on Housing Investment: Collapsed to PUMA-Industry-Year Level for all Metros

	(1)	(2)	(3)
Risk Measure	-7,681***		
	(2,014)		
Other Industry Risk		-11,276***	-11,916***
		(2,104)	(2,132)
Own Industry Risk		4,646***	
		(995)	
$Risk_{d,t}$			1,235
			(1,197)
$Concen_{d,m,t}$			6,330***
			(940)
State-Year F.E.	Yes	Yes	Yes
PUMA F.E.	Yes	Yes	Yes
Industry F.E.	Yes	Yes	Yes
Occupation F.E.	No	No	No
Observations	354,202	354,202	354,202
R-squared	.62	.62	.62

Robust standard errors in parentheses clustered at the PUMA level. Data is limited to employed homeowners that are the highest wage earner in their household, who are located in a metropolitan area and moved into their current home in the last 9 years. Household data is collapsed to Industry-PUMA-Year cells. Regressions include controls for average age, age squared, marital status, race, education, ethnicity, self employment status and veteran status of the Industry-PUMA-Year cell. $*p < 0.10$ $**p < 0.05$ $***p < 0.01$.

Table 14
OLS Regression of Risk on Housing Investment: Income Restrictions

	(1)	(2)	(3)	(4)	(5)	(6)
Risk Measure	-7,476**		-6,604**		-2,769	
	(3,161)		(2,644)		(5,911)	
Other Industry Risk		-12,793***		-11,848***		-7,396
		(2,850)		(2,467)		(5,687)
Own Industry Risk		4,105***		4,194***		4,060**
		(1,359)		(1,217)		(1,866)
Income Restriction	100% Wage	100% Wage	No Investment Inc	No Investment Inc	Investment Inc	Investment Inc
State-Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes
PUMA F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Occupation F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	461,980	461,980	671,036	671,036	201,422	201,422
R-squared	.53	.53	.52	.52	.53	.53

Robust standard errors in parentheses clustered at the PUMA level. Regressions include controls for age, age squared, marital status, race, education, ethnicity, self employment status and veteran status of the household's highest wage earner. Data is limited to employed homeowners that are the highest wage earner in their household, who are located in a major metropolitan area and moved into their current home in the last 9 years. $^{*}p < 0.10$ $^{**}p < 0.05$ $^{***}p < 0.01$.

Table 15
Housing Investment in High and Low Maximum Unemployment Insurance Benefit States

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Risk Measure	1,824	-15,608***			-5,689	-8,412**		
	(4,846)	(3,383)			(4,288)	(4,008)		
Other Industry Risk			-5,552	-19,141***			-8,015**	-15,063***
			(4,331)	(3,464)			(3,801)	(3,788)
Own Industry Risk			6,556***	1,877			1,389	5,564***
			(1,873)	(1,236)			(1,764)	(1,578)
UI Benefits	Above Median	Below Median	Above Median	Below Median	Above Mean	Below Mean	Above Mean	Below Mean
State-Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
PUMA F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	506,856	372,542	506,856	372,542	384,079	495,319	384,079	495,319
R-squared	.54	.52	.54	.52	.48	.55	.48	.55

Robust standard errors in parentheses clustered at the PUMA level. Regressions include controls for age, age squared, marital status, race, education, ethnicity, self employment status and veteran status of the household's highest wage earner. Data is limited to employed homeowners that are the highest wage earner in their household, who are located in a major metropolitan area and moved into their current home in the last 9 years. $*p < 0.10$ $**p < 0.05$ $***p < 0.01$.

Figure 1
Average Uncorrelated Risk by PUMA

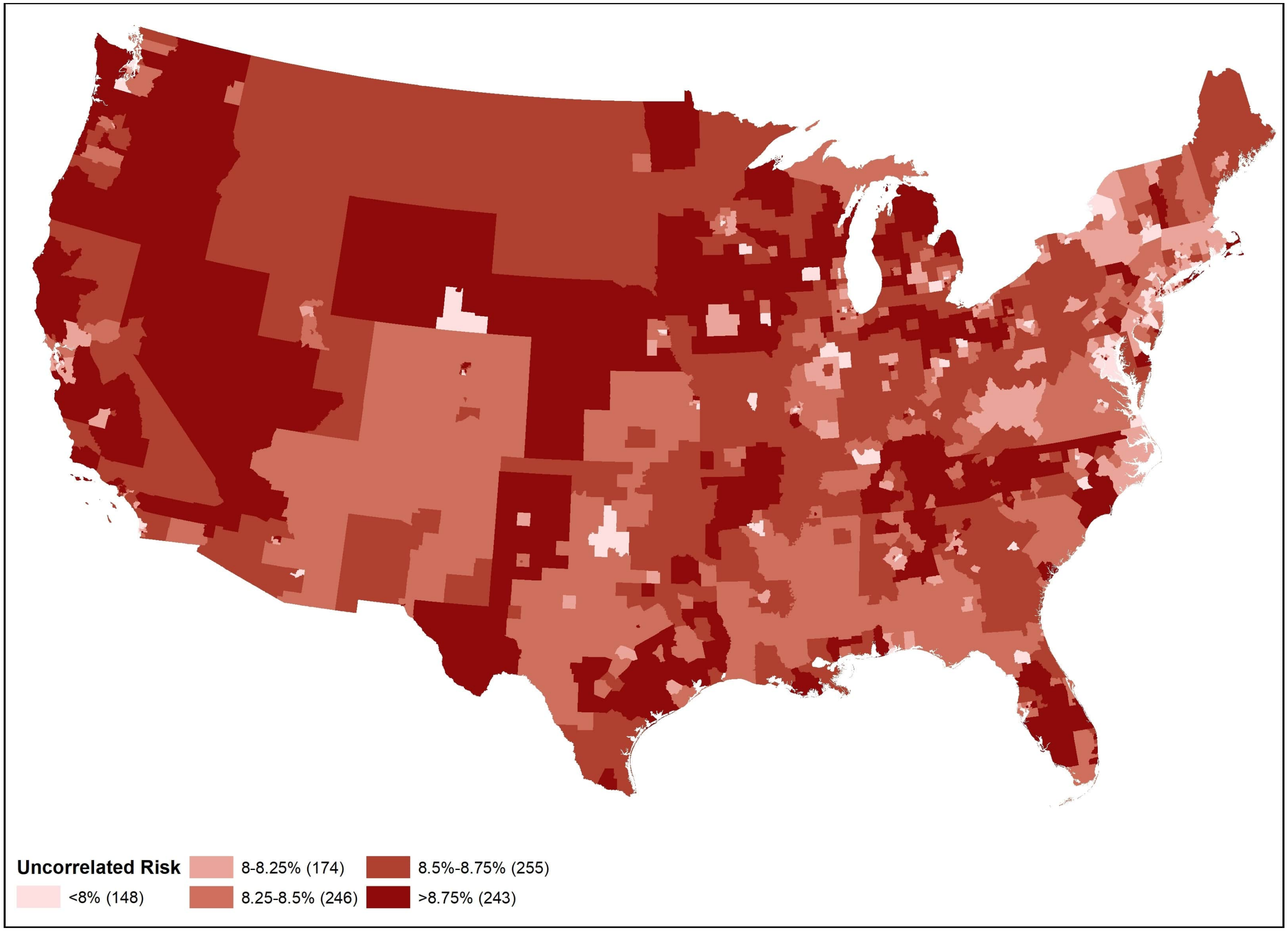


Figure 2
Average Risk Measure by PUMA

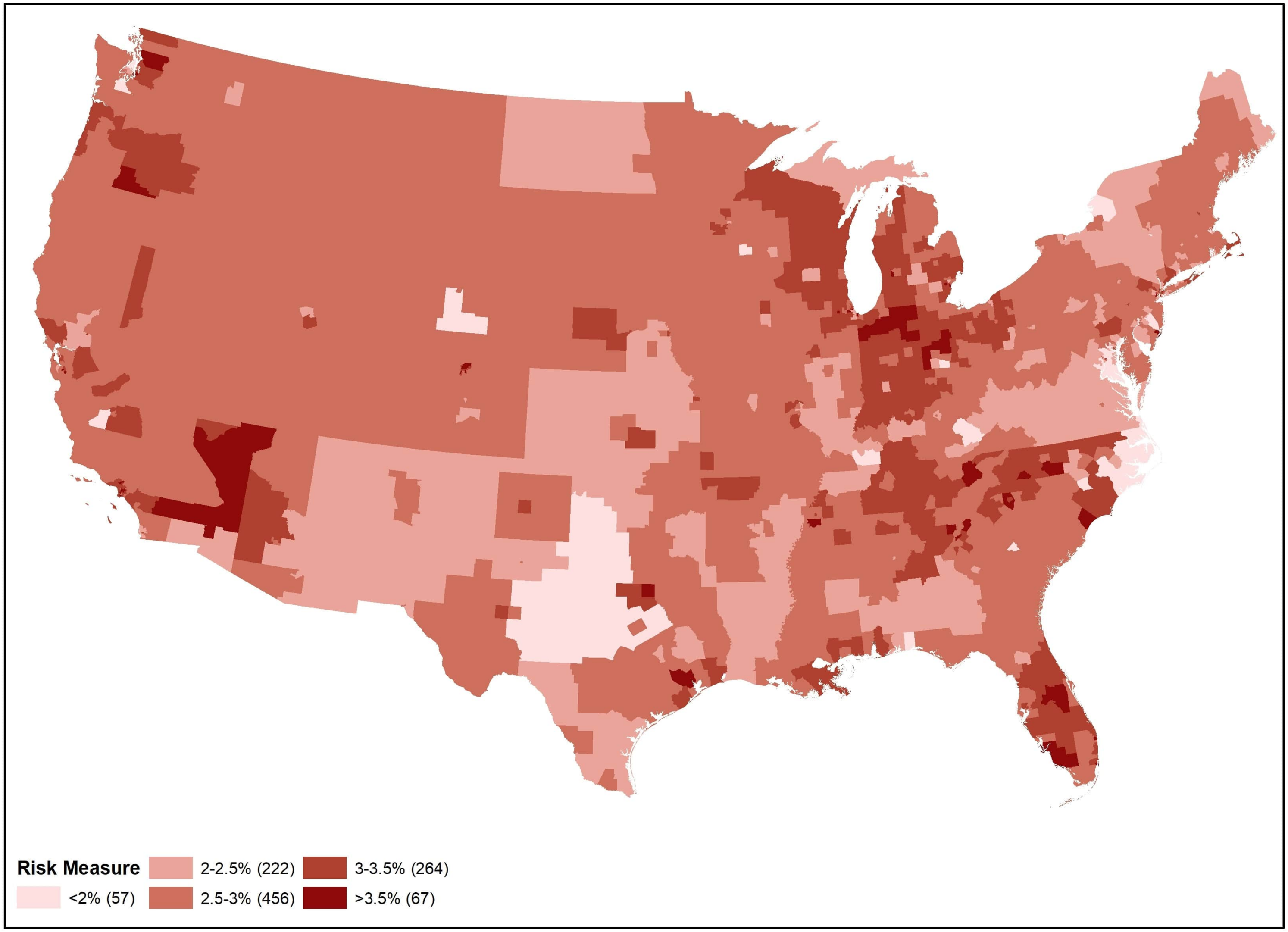


Figure 3
Average Risk Measure by PUMA for Industry Code 336 (Transportation Equipment Manufacturing)

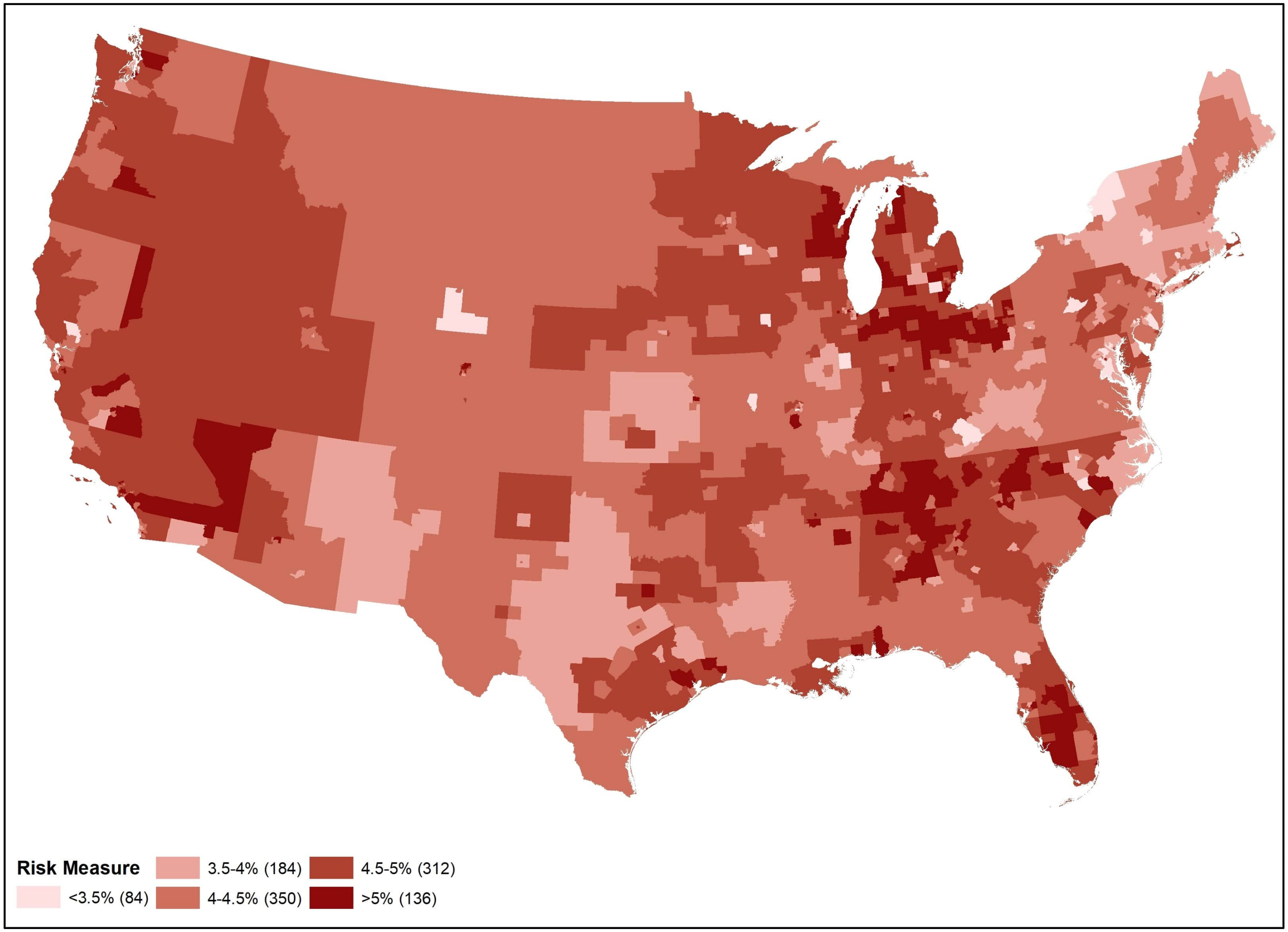


Figure 4
Average Other Industry Risk by PUMA for Industry Code 336 (Transportation Equipment Manufacturing)

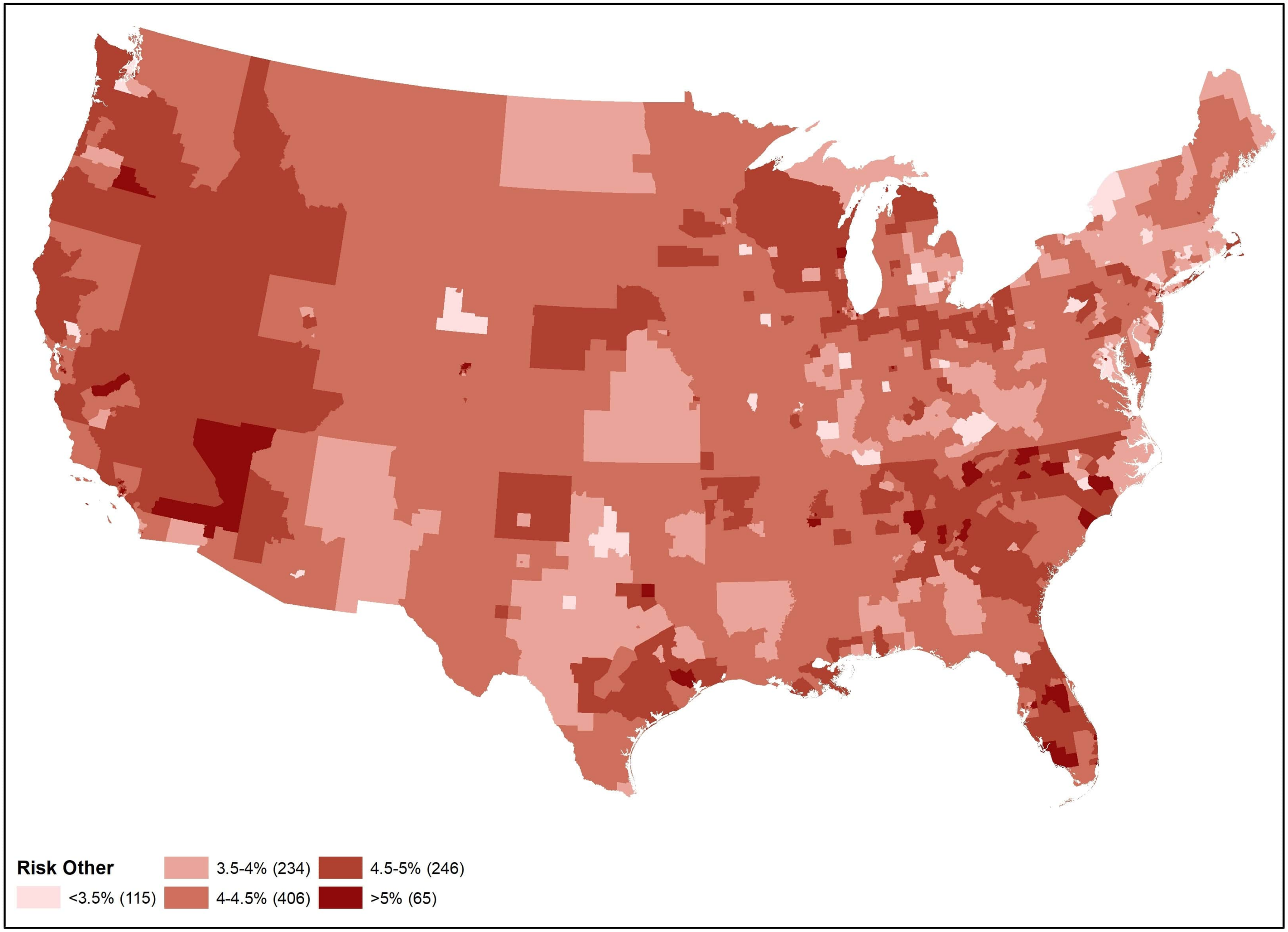


Figure 5
Average Own Industry Risk by PUMA for Industry Code 336 (Transportation Equipment Manufacturing)

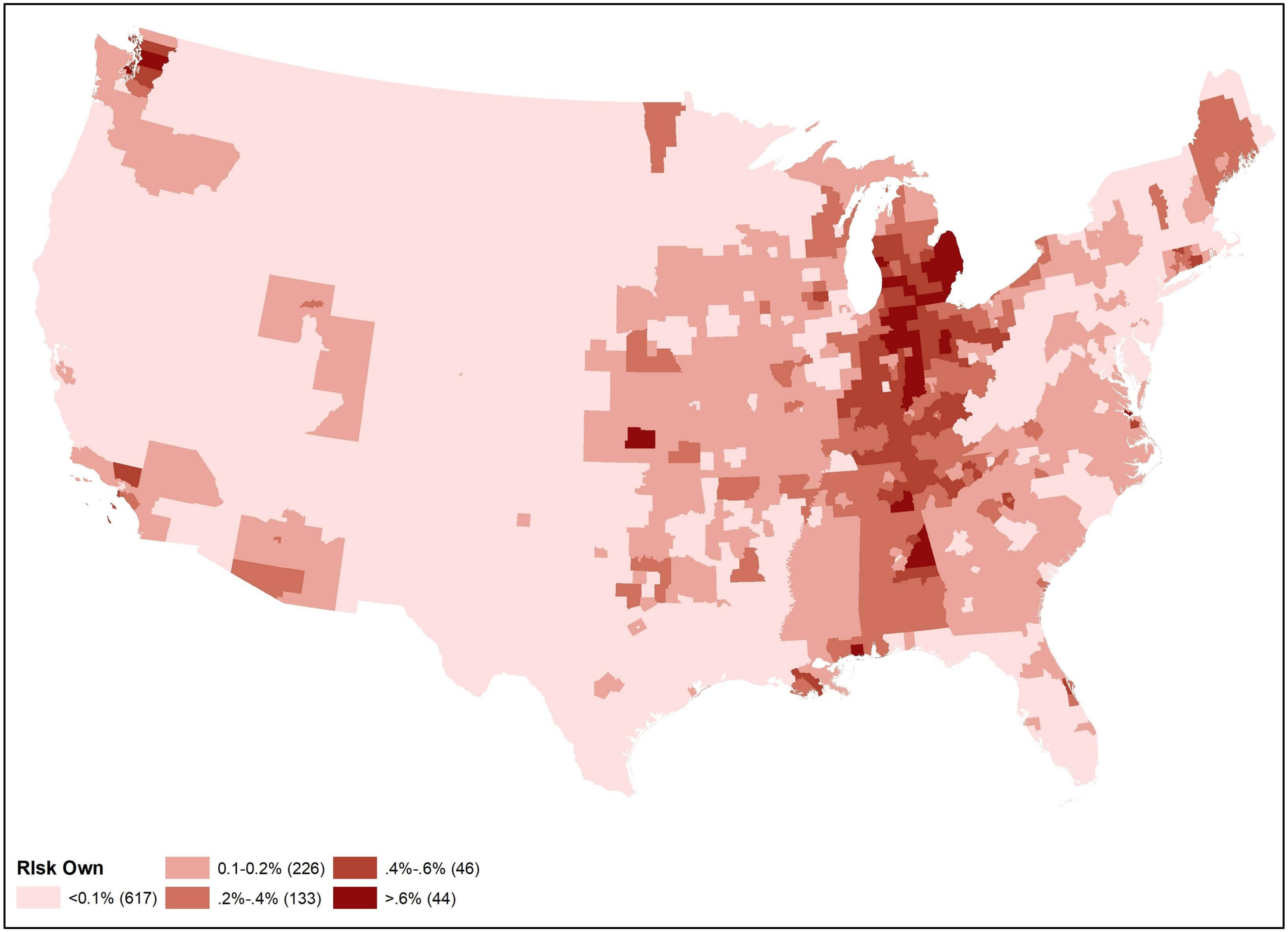
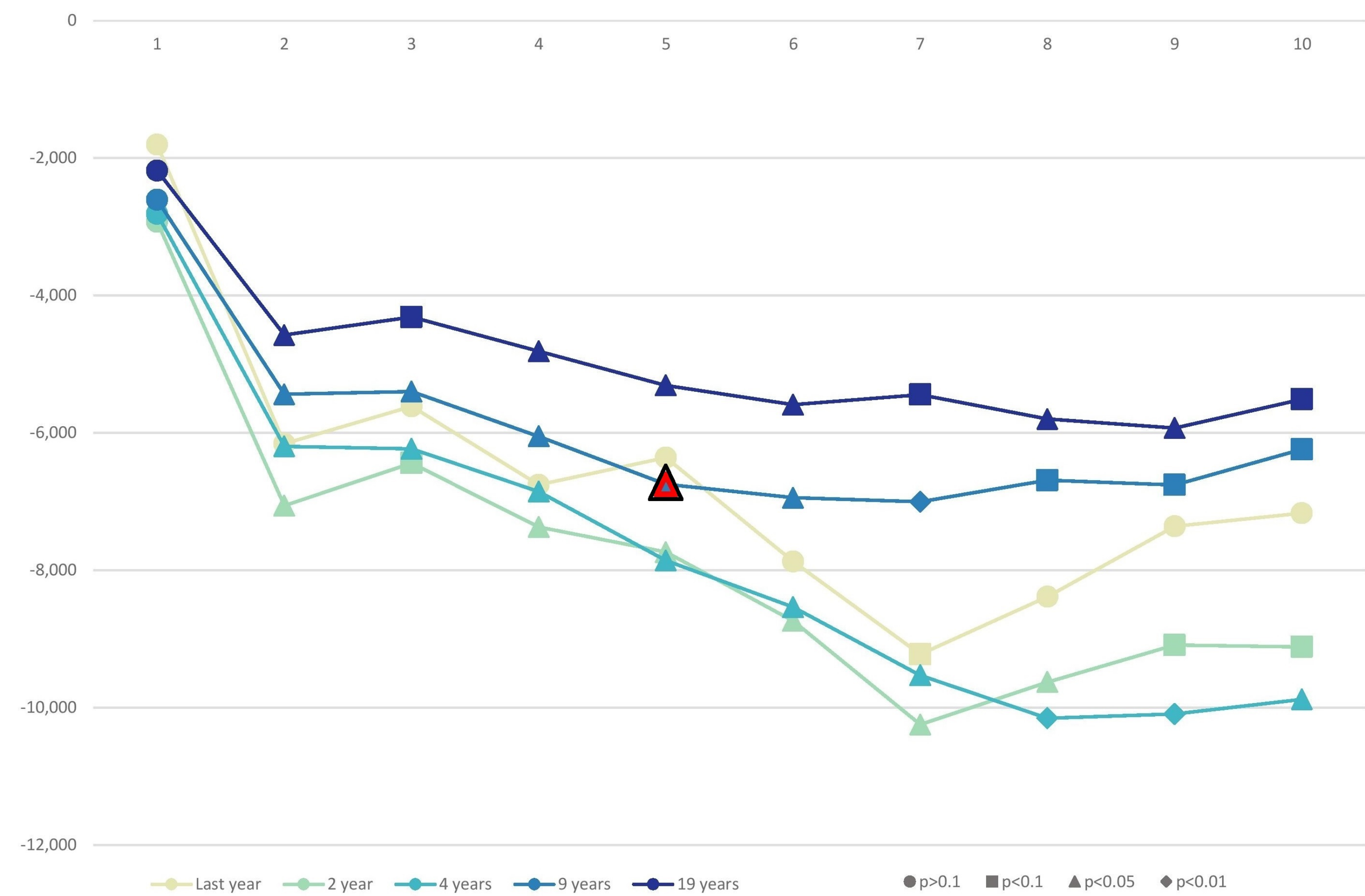


Figure 6

Coefficient on Correlated Labor Market Risk by Metro Size and Years Since Moved into Current House



Notes: Coefficients on the Correlated Labor Market Risk variable for regressions with varying restrictions on sample. Moving across the Y axis, these regressions limit the sample to metropolitan areas with at least X PUMAs (1 to 10) and the different lines correspond to samples limited to households who moved into their current homes in the last 1 to 19 years. Regressions include all controls listed in Table 2. The highlighted coefficient corresponds to the coefficient in column 1 of Table 2. The shape of the marker for each point on the Figure above corresponds to the p-value associated with that regression's estimated of the coefficient of the Correlated Labor Market Risk variable.

Loyalty rewards and redemption behavior: Stylized facts for the U.S. airline industry*

Daniel Ladd
University of California-Irvine

Alexander Luttmann†
The MITRE Corporation

June 14, 2024

Abstract

Over the past forty years, one of the most important datasets in industrial organization has been the Airline Origin and Destination Survey (DB1B). Most studies relying on these data remove tickets with fares less than \$20, assuming that these are heavily discounted frequent flyer awards (FFAs). We investigate the validity of this approach by first defining the size of the frequent flyer market using annual Form 10-K filings. Exploiting a federal regulation, we then outline a novel approach to identify FFAs in the DB1B. Our method indicates that the \$20 cutoff used by researchers is too high and may be lowered to \$12 for tickets appearing in the DB1B after February 1, 2002. Using the FFAs we identify, we show how the characteristics of award tickets differ from paid tickets and how these characteristics have changed over time. We then demonstrate how various market and product quality characteristics influence the share of passengers traveling on FFAs. Finally, we find that price dispersion increases on routes with larger shares of frequent flyer passengers, implying that airline loyalty programs enhance market power.

JEL Codes: L11, L13, L14, L93, M31, R40, R49
Keywords: Airlines, competition, loyalty rewards, frequent flyer tickets, product quality

*We thank Jan Brueckner, Kerry Tan, and participants at the 93rd Annual Meeting of the Southern Economic Association for providing insightful comments. All errors are our own.

†*Corresponding author.* Transportation Performance and Economic Analysis Department, The MITRE Corporation, 7515 Colshire Drive, McLean, VA 22102; email: `aluttmann@mitre.org`. The author's affiliation with the MITRE Corporation is for identification purposes only, and is not intended to convey or imply MITRE's concurrence with, or support for, the positions, opinions, or viewpoints expressed by the author.

1 Introduction

Loyalty programs that reward consumers for repeat purchases are common in a variety of retail markets including airlines, car rentals, clothing, credit cards, hotels, restaurants, and supermarkets. By implementing loyalty programs, firms exercise a degree of market power over repeat customers by introducing a cost of switching to a competitor's product. These switching costs often include foregoing future rewards in addition to transaction costs associated with switching suppliers, learning to use new brands, and uncertainty about the quality of the competing product (Klemperer, 1995).

Although the typical consumer is a member of several different rewards programs, it is unclear whether these programs enhance or reduce competition in markets where they are present. For example, many have argued that loyalty programs are anticompetitive because switching costs generally raise prices, create barriers to entry, generate deadweight losses, facilitate tacit collusion, and reduce product variety (Banerjee and Summers, 1987; Cairns and Galbraith, 1990; Fong and Liu, 2011; Kim et al., 2001; Klemperer, 1995). Others argue that loyalty programs are "business stealing devices" that enhance competition and increase total surplus (Caminal, 2012; Caminal and Claici, 2007; Caminal and Matutes, 1990).[1]

Despite the importance of determining whether loyalty programs increase consumer welfare, empirical evidence on the effects of these programs are limited. Data on program membership and redemption behavior are often proprietary, and thus researchers have struggled to obtain appropriate data. However, there are some exceptions for the airline industry. Due to regulatory reporting requirements, airline data is often available to researchers. Using geocoded data from a major European frequent flyer program (FFP), De Jong et al. (2019) find that national airlines enjoy a substantial loyalty advantage in their home country.[2] Lederman (2007) used airline mergers to

[1] In addition, Basso et al. (2009) argue that within the airline industry, firms create frequent flyer programs to take advantage of the principal-agent problem between employers who pay for airline tickets and the employees who book them. In their theoretical model, they find that while these programs likely raise prices, airlines may end up worse off than if they had not created the programs due to intensified competition in the form of frequent flyer benefits.

[2] In particular, De Jong et al. (2019) find that foreign consumers earn about 60% less miles and are 70% less likely to be FFP members than domestic consumers.

instrument for enhancements to an airline's FFP and found that these enhancements are associated with increases in demand on routes that depart from an airline's hub airports (i.e., airports where the airline is dominant). In a follow-up paper, Lederman (2008) finds that FFPs enable airlines to charge higher fares on routes that depart from its hubs.

Nevertheless, many questions surrounding the economics of loyalty programs remain unexplored. For example, are consumers encouraged to redeem rewards on high or low quality products when the firm operating the loyalty program vertically differentiates its products? Do consumers disproportionately redeem rewards on high or low price products when multiple redemption options are available? How do loyalty programs affect price dispersion in differentiated product markets? In this paper, we shed light on these questions by presenting empirical evidence on frequent flyer redemption behavior for the U.S. airline industry.

Empirical questions concerning the economics of loyalty programs have been difficult to answer due to difficulties in identifying award redemptions in public and proprietary datasets. Our first contribution outlines an approach to credibly identify frequent flyer awards in one of the most widely used datasets in empirical industrial organization and transportation economics, the Department of Transportation's Airline Origin and Destination Survey (referred to as database DB1B). Released quarterly, the DB1B data are a 10% random sample of all airline tickets that originate in the United States on domestic carriers.[3] Over the past forty years, researchers relying on these data have provided empirical evidence on important questions surrounding competition policy and the functioning of oligopolistic markets. For example, the DB1B have been used to examine topics such as how incumbents respond to the threat of entry[4], the relationship between competition and price dispersion[5], how competition affects prices and profitability[6], the price effects of merg-

[3]The Department of Transportation relies on these data to determine air traffic patterns, air carrier market shares, and passenger flows.

[4]E.g., see Goolsbee and Syverson (2008), Gayle and Wu (2013), Gayle and Xie (2018), Morrison (2001), and Tan (2016).

[5]E.g., see Borenstein and Rose (1994), Gerardi and Shapiro (2009), Dai et al. (2014), Luttmann (2019), and Kim et al. (2023).

[6]E.g., see Berry and Jia (2010), Brueckner et al. (2013), and Kwoka et al. (2016).

ers[7], the price effects of granting antitrust immunity in international markets[8], the price effects of domestic alliances[9], how multimarket contact may facilitate tacit collusion[10], how capacity constraints affect prices[11], the revenue effects of product unbundling[12], and the competitive effects of common ownership[13], among others.[14]

A common theme among papers relying on DB1B data are their approaches towards removing low-fare tickets. Almost all papers remove observations with fares below a $20 or $25 cutoff, assuming that these fares represent heavily discounted frequent flyer award (FFA) tickets.[15] We focus on this portion of the dataset and find that researchers are removing approximately 7%-10% of observations (tickets) when applying a $20 cutoff. We then investigate the validity of this approach by comparing these dropped ticket counts to the number of FFAs reported by airlines in their annual Form 10-K filings, finding that FFAs account for roughly 7%-8% of revenue passenger miles in a given year.[16]

To identify frequent flyer tickets in the DB1B data, we exploit a February 1, 2002 federal regulation that established the Passenger Fee, also known as the September 11 Security Fee. This fee is collected by all commercial air carriers at the time airfare is purchased, *including when passengers redeem frequent flyer awards*. Although not the focus of this paper, the method of identifying frequent flyer tickets that we describe in Section 3 can serve as a starting point for empirical researchers interested in exploring questions such as the effects of mergers and codesharing on FFAs,

[7] E.g., see Luo (2014), Carlton et al. (2017), Shen (2017), and Li et al. (2022).

[8] E.g., see Brueckner and Whalen (2000), Whalen (2007), Brueckner et al. (2011), Gayle and Thomas (2016), Calzaretta Jr et al. (2017), Brueckner and Singer (2019), and Gayle and Xie (2019).

[9] E.g., see Gayle (2013).

[10] E.g., see Ciliberto and Williams (2014), Ciliberto et al. (2019), and Kim et al. (2021).

[11] E.g., see Fukui (2019).

[12] E.g., see Brueckner et al. (2015) and He et al. (2022).

[13] E.g., see Azar et al. (2018).

[14] The DB1B data have also been used to investigate how the internet influences price dispersion (Orlov, 2011) and how government legislation affects fares (Luttmann and Nehiba, 2020; Snider and Williams, 2015), among others.

[15] For example, Severin Borenstein has graciously posted raw and summary versions of the 1979Q1-2016Q3 DB1A/DB1B data on NBER's website. The summary files at the airline-route-quarter level are generated after removing tickets with fares below $20 and fares above $9,998. These datasets are available at `http://data.nber.org/data/dot-db1a/`.

[16] A revenue passenger mile is a standard industry metric that summarizes the number of miles flown by paying (revenue) passengers. For example, a revenue passenger mile is flown when a revenue passenger is transported one mile.

how the timing and introduction of airline branded credit cards impacts reward redemptions, and how frequent flyer program devaluations over time have affected consumer welfare, among others. Overall, our approach indicates that FFAs can be credibly identified in the DB1B and that the $20 cutoff used by most researchers is too high. In particular, the cutoff may be lowered to $12 for any tickets appearing in the DB1B after the establishment of the Passenger Fee in 2002.

Using the FFAs we identify from 2005-2019, our second contribution establishes how the characteristics of FFA tickets differ from paid tickets and how these differences have changed over time. Foremost, we find that FFAs are disproportionately redeemed on less concentrated (i.e., more competitive) routes to leisure destinations, suggesting that passengers either have a preference for redeeming award tickets for travel to vacation destinations or airlines are successful in restricting capacity on more concentrated routes where price markups are expected to be higher. Second, FFAs are also found to be disproportionately redeemed on seasonal routes (i.e., routes that are not served continuously throughout the year).[17] In comparison to paid tickets, we also find that FFAs have more flight segments, are longer on average, and more likely to be roundtrip. However, these differences have declined over time. By 2019, FFAs are generally not statistically distinguishable from paid tickets in terms of these observable characteristics. This finding suggests that devaluations of frequent flyer programs and airline consolidation during our sample period have resulted in consumers treating reward redemptions more like cash.[18]

Additionally, we tie our results to the extensive literature on the hub premium.[19] In particular, we find that FFAs are disproportionately redeemed on routes from origins with higher carrier specific concentrations of passengers. This finding supports the classic hub premium story that fares for routes leaving from hub airports are higher due to the value of frequent flyer miles.

---

[17]Examples of seasonal routes are airport-pairs involving Aspen, Colorado that are only served during the winter ski season (e.g., Delta's seasonal Los Angeles (LAX)-Aspen (ASE) or Atlanta (ATL)-Aspen (ASE) services).

[18]It is also possible that it is the airlines themselves who start treating frequent flyer miles more like cash. Historically, airline rewards programs awarded miles based on actual distance flown and set thresholds for award redemptions (i.e., 25,000 miles for a domestic round trip). More recently, airlines have moved to awarding points based on the amount paid for a consumer's flight and redemption thresholds now vary by route, season, and schedule. Additionally, with the growth in rewards credit cards, airlines now derive comparable revenue from selling their miles to credit card companies than to passengers flying their routes.

[19]For more on the hub premium, see Borenstein (1989), Lederman (2007), Lederman (2008), Lee and Luengo-Prado (2005), Ciliberto and Williams (2010), Escobari (2011), and Bilotkach and Pai (2016).

Our third contribution establishes how various market structure and product quality characteristics affect the share of an airline's passengers traveling on FFAs. We find that high fare routes do not have larger shares of frequent flyer awards. In contrast, we find that measures of competition and routing quality have larger effects on frequent flyer passenger shares. Airlines appear to limit FFAs on more competitive routes and more direct routes. Moreover, we also find that routes with low load factors have higher shares of FFAs, suggesting that airlines may restrict a customer's ability to redeem awards on densely traveled routes.

Finally, our results show that price dispersion increases on routes with higher shares of frequent flyer passengers. The increase in price dispersion is driven by both higher prices at the 10th percentile of fares and larger price increases at the 90th percentile of fares. These increases suggest that airlines reduce the availability of lower fare tickets for paying customers when more frequent flyer passengers are present on a route.

The rest of this paper is organized as follows. Section 2 provides details on the size of the frequent flyer market. Section 3 describes the Department of Transportation (DOT) data used in the analysis and describes the method used to identify frequent flyer tickets. After identifying these award tickets, Section 4 outlines the descriptive analysis used to identify how FFAs differ from paid tickets and presents results from this ticket level analysis. Section 5 outlines the empirical strategy used to determine how various market structure and product quality characteristics affect the share of frequent flyer passengers and presents results from this market level analysis. Section 6 provides evidence on how FFAs affect price dispersion. Finally, Section 7 concludes.

2 Size of the frequent flyer market

To provide statistics on the size of the frequent flyer market, we compiled data from annual Form 10-K filings for each of the major U.S. airlines from 2005-2019. Since 1934, publicly-traded U.S. companies are required to submit an annual report to the Securities and Exchange Commission providing a comprehensive overview of the company's business and financial condition.[20] Typical information reported on the Form 10-K include a company's organizational structure, risk factors, subsidiaries, and audited financial statements. Because frequent flyer programs are an important aspect of an airline's business strategy, many of the major airlines also report details on the size of their loyalty programs in these annual filings.

By airline and year, Table 16 reports the percentage of revenue passenger miles that are due to passengers traveling on frequent flyer awards. A revenue passenger mile is a standard industry metric summarizing the number of miles flown by paying (revenue) passengers.[21] The numbers in Table 16 indicate that passengers traveling on frequent flyer awards account for a sizeable fraction of total passenger traffic. Depending on carrier, award passengers accounted for between 6.0% and 14.1% of revenue passenger miles in 2019. For the two low-cost carriers in the table, there is a clear trend of increasing award traffic over time. For Southwest, award passengers accounted for 6.6% of revenue passenger miles in 2005 compared to 14.1% in 2019. For JetBlue, award passengers accounted for just 2.0% of revenue passenger miles in 2005 compared to 6.0% in 2019. In contrast, the trend for the three major legacy carriers remained relatively constant from 2005-2019. On American, Delta, and United, passengers traveling on frequent flyer awards accounted for 7%-9% of revenue passenger miles in most years.

[20]The Form 10-K reporting requirement was established as a result of the Securities and Exchange Act of 1934. For more information on the Form 10-K, see `https://www.sec.gov/fast-answers/answers-form10khtm.html`.

[21]A revenue passenger mile is flown when a paying (revenue) passenger is transported one mile.

Table 16: Frequent Flyer Award Traffic
(% of revenue passenger miles)

Airline	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Alaska (AS)	7.9%	8.6%	9.7%	12.4%	15.0%	9.0%	*	*	*	*	*	*	*	*	*
American (AA)	7.2%	7.5%	7.5%	9.7%	8.9%	8.8%	8.8%	8.6%	8.2%	5.5%	6.5%	6.3%	6.1%	7.6%	8.0%
Continental (CO)	7.0%	6.8%	7.2%	8.5%	6.0%	5.7%	5.6%	Acquired by United (UA)							
Delta (DL)	9.0%	*	*	*	*	8.3%	8.2%	8.0%	7.3%	7.4%	7.2%	7.9%	7.9%	8.2%	8.9%
Hawaiian (HA)	*	*	*	6.0%	5.0%	6.0%	5.7%	5.2%	4.8%	5.3%	5.0%	5.0%	5.0%	6.0%	6.0%
JetBlue (B6)	2.0%	2.0%	3.0%	4.0%	3.7%	3.0%	2.0%	3.0%	3.0%	3.0%	4.0%	4.0%	5.0%	5.0%	6.0%
Northwest (NW)	7.3%	7.3%	*	*	Acquired by Delta (DL)										
Southwest (WN)	6.6%	6.4%	6.2%	6.4%	7.7%	7.9%	8.6%	9.0%	9.5%	11.0%	12.0%	12.7%	13.8%	13.8%	14.1%
United (UA)	7.4%	8.1%	8.0%	9.1%	8.3%	7.5%	8.2%	7.1%	7.7%	7.1%	7.5%	7.7%	7.5%	7.1%	7.2%
U.S. Airways (US)	9.1%	4.0%	4.0%	4.0%	4.0%	4.0%	4.0%	4.0%	3.5%	Acquired by American (AA)					

Source: 2005-2019 Form 10-K filings for Alaska, American, Continental, Delta, Hawaiian, JetBlue, Northwest, Southwest, United, and U.S. Airways.

Notes: *Number not reported in Form 10-K filing. A revenue passenger mile is flown when a paying (revenue) passenger is transported one mile. AirTran (FL), Frontier (F9), Spirit (NK), and Virgin America (VX) are not included because they do not report the number of frequent flyer awards in their annual Form 10-K filings. Allegiant (G4) introduced their frequent flyer program in August 2023.

By airline and year, Table A1 reports the number of frequent flyer award tickets redeemed. Across all airlines, there is a clear increasing trend in the number of award tickets redeemed over time with number of award flights across reporting carriers, increasing from just over 18 million in 2005 to more than 50 million in 2019. This increasing trend is likely a result of a combination of factors including the introduction of airline branded credit cards that allowed reward program members to accrue frequent flyer miles, the completion of several mergers, and increasing travel demand over time.

Using data from 2019 Form 10-K filings, Table 17 compares airline revenue from loyalty programs to net income and liabilities accrued through their loyalty programs to the airline's long term debt. In 2019, each major domestic carrier reported frequent flyer program revenue that exceeded their net income. Additionally, carrier's liabilities from their frequent flyer programs are comparable to or exceed that of their long term debt.[22] Overall, the statistics presented in Tables 16 and 17 indicate that passengers traveling on frequent flyer awards represent a large and non-trivial fraction of total passenger traffic and play an important role in airline profitability.

Investors also immensely value loyalty programs given that several airlines collateralized the future cash flows of their frequent flyer programs to raise billions of dollars in loans during the Covid-19 pandemic. For example, United raised $6.8 billion in July 2020, Spirit $850 million in September 2020, Delta $9 billion in September 2020, and American $10 billion in March 2021.[23]

Table 17: Airline Loyalty Revenues and Liabilities 2019
(millions of dollars)

Airline	Loyalty Revenue	Net Income	Loyalty Liabilities	Long Term Debt
Alaska (AS)	$1,169	$769	$1,990	$1,264
American (AA)	$5,540	$1,972	$8,615	$22,372
Delta (DL)	$4,862	$4,767	$6,728	$8,052
Hawaiian (HA)	$244	$224	$350	$547
JetBlue (B6)	$592	$569	$661	$1,990
Southwest (WN)	$3,787	$2,300	$3,385	$1,846
United (UA)	$4,350	$3,009	$5,276	$13,145

Source: 2019 Form 10-K filings for Alaska, American, Delta, Hawaiian, JetBlue, Southwest, and United.

[22] Loyalty liabilities are not included in Form 10-K reports of Long Term Debt but are reported separately under other liabilities.

[23] For additional information on how loyalty programs helped save airlines during the Covid-19 pandemic, see `https://hbr.org/2021/04/how-loyalty-programs-are-saving-airlines`.

3 Method for identifying frequent flyer tickets in the DB1B database

In order to identify airline frequent flyer awards, itinerary and price data (inclusive of all ticket taxes and fees) are taken from the U.S. Department of Transportation's Airline Origin and Destination Survey (referred to as database DB1B). Data from this survey are released quarterly and generated from a 10% random sample of all airline tickets that originate in the United States on U.S. based carriers. Information from this survey include ticket characteristics such as the total fare, fareclass, origin and destination airports, operating and ticketing carriers, distance flown, number of route segments, connecting airports (if any), and an indicator specifying if the ticket is roundtrip.

To identify frequent flyer tickets in the DB1B data, we exploit a February 1, 2002 federal regulation that established the Passenger Fee, also known as the September 11 Security Fee.[24] This fee is collected by all commercial air carriers at the time airfare is purchased, *including when passengers redeem frequent flyer awards*.[25] Airlines then remit these fees to the Transportation Security Administration (TSA). Between February 1, 2002 and July 20, 2014, the TSA imposed a security fee of $2.50 per flight segment for a maximum of $5.00 per one-way trip or $10.00 per roundtrip. On July 21, 2014, the Passenger Fee was changed to $5.60 per one-way trip and $11.20 per roundtrip (i.e., fees no longer applied on a flight segment basis). Table 18 summarizes this regulation and the amendment made on July 21, 2014.

[24] The Passenger Fee was initially authorized under the Aviation and Transportation Security Act. For more information on this fee, see `https://www.tsa.gov/for-industry/security-fees`.

[25] Specifically, the original legislation states that "Direct air carriers and foreign air carriers must collect the security service fees imposed on air transportation sold on or after February 1, 2002. *The security service fee imposed by this interim final rule applies to passengers using frequent flyer awards for air transportation, but is not applicable to other nonrevenue passengers.*" See Federal Register Vol. 66, No. 250 available at `https://www.govinfo.gov/content/pkg/FR-2001-12-31/pdf/01-32254.pdf`.

Table 18: Passenger Fee Summary

Legislation	Effective Fee Date	Fee for one-way trips: One segment	Fee for one-way trips: Two segments	Fee for one-way trips: Three or more	Fee for Roundtrips: Two segments	Fee for Roundtrips: Three segments	Fee for Roundtrips: Four or more
Public Law 107–71	Feb. 1, 2002	\$2.50	\$5.00	\$5.00	\$5.00	\$7.50	\$10.00
Public Law 113–67/113–294	Jul. 21, 2014	\$5.60	\$5.60	\$5.60	\$11.20	\$11.20	\$11.20

Source: Federal Register Vol. 66, No. 250, Federal Register Vol. 79, No. 119, and Federal Register Vol. 80, No. 107. More information on the Passenger Fee is located at `https://www.tsa.gov/for-industry/security-fees`.
Notes: "The security service fee must be imposed on passengers who obtained the ticket for air transportation with a frequent flyer award, but may not be imposed on other nonrevenue passengers." Federal Register Vol. 79, No. 119.

Because passengers redeeming frequent flyer awards (FFAs) are required to pay the Passenger Fee, we identify FFAs in the DB1B by classifying tickets according to the fee structure in Table 18. Prior to July 21, 2014, one segment trips with \$2.50 fares are identified as FFAs.[26] One-way trips with two or more segments and roundtrips with two segments are identified as FFAs if the fare charged was \$5. For roundtrips with three segments, tickets with \$7.50 fares are identified as FFAs.[27] For roundtrips with four or more segments, tickets with \$10 fares are identified as FFAs. After July 20, 2014, one-way trips with \$5.60 fares and roundtrips with \$11.20 fares are identified as FFAs.[28]

Figures 7, 8, and 9 illustrate our strategy for identifying FFAs. In Figure 7, the distribution of DB1B fares under \$20 in 2005, 2010, 2015, and 2018 are presented. In line with the Passenger

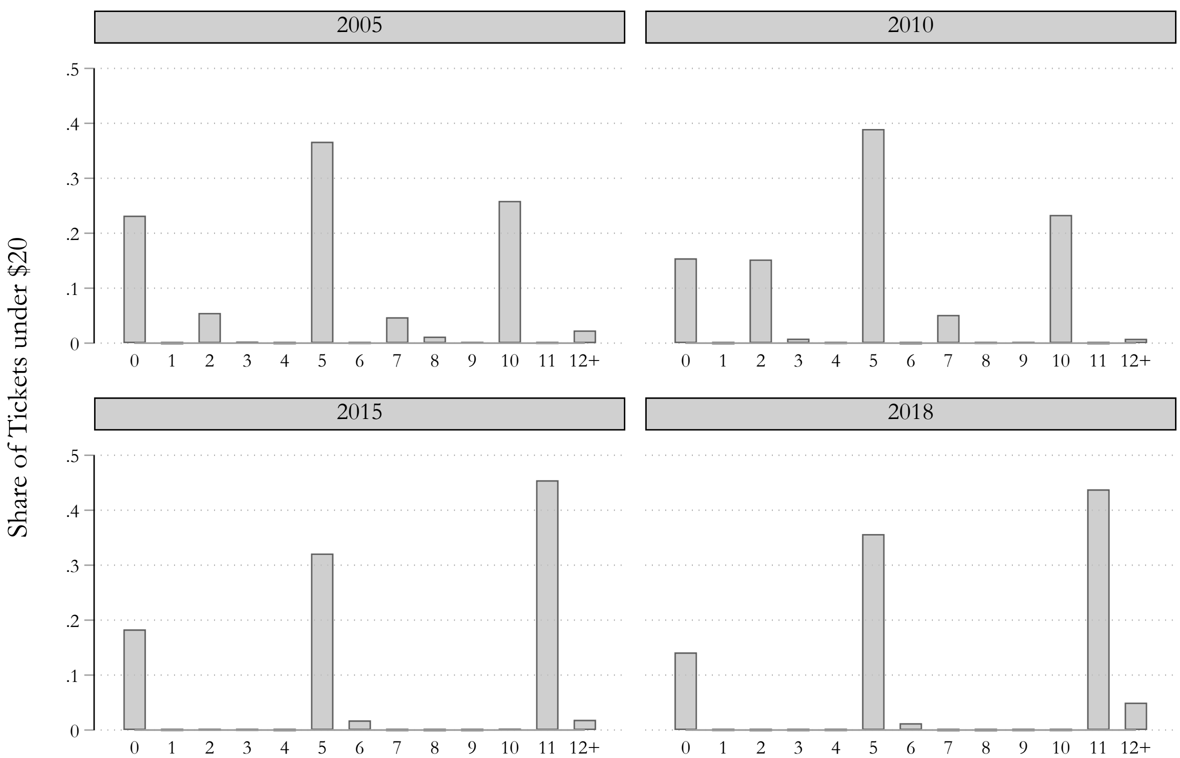
[26] Since fares in the DB1B are expressed as whole numbers, one segment trips with fares of \$2 or \$3 are classified as FFAs prior to July 21, 2014.

[27] Since fares in the DB1B are expressed as whole numbers, three segment roundtrips with fares of \$7 or \$8 are identified as FFAs prior to July 21, 2014.

[28] Since fares in the DB1B are expressed as whole numbers, one-way trips with fares of \$5 or \$6 are identified as FFAs after July 20, 2014. Similarly, roundtrips with \$11 or \$12 fares are identified as FFAs after July 20, 2014.

Fees in effect prior to July 21, 2014, spikes in the distribution occur at \$2, \$5, \$7, and \$10 in 2005 and 2010.[29] Reflecting the Passenger Fee change in 2014, spikes occur at \$5 and \$11 in 2015 and 2018. However, not all tickets with fares less than \$20 are FFAs. For example, the spikes observed at \$0 in Figure 7 reflect nonrevenue passengers such as airline employees and friends and family of airline employees flying standby. In addition, some fares under \$20 are actual paid fares (e.g., Allegiant's \$9 flight sales and Frontier's \$15 and \$19 flash sales).

Figure 7: Distribution of DB1B Fares Under \$20 in 2005, 2010, 2015, and 2018



Notes: Data from DB1B files for 2005, 2010, 2015 and 2018. Bars represent the share of passengers with tickets under \$20 that reported that exact itinerary fare. Data are limited to one-way itineraries with three or fewer segments and round trip itineraries with six or fewer segments.

Figure 8 displays the distribution of DB1B fares under \$20 in 2013 (red bars) and 2015 (blue bars) for American (Panel A) and Delta (Panel B). The four charts in each panel correspond to the Passenger Fees charged to that itinerary under the 2013 fee regime. The chart titled "\$2.50"

[29] Itinerary prices in DB1B are reported in whole dollar increments.

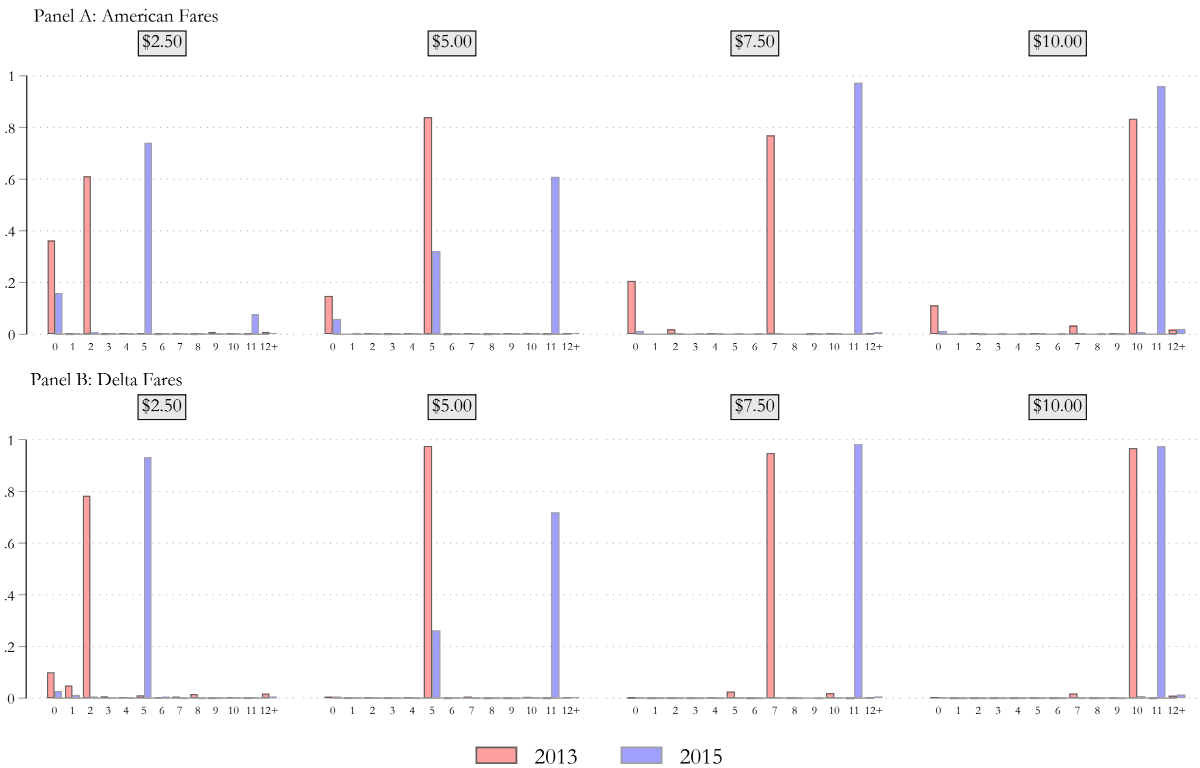
in each panel displays the distribution of fares under $20 for one segment trips as these trips were subject to a Passenger Fee of $2.50 in 2013. With the change in the Passenger Fee in 2014, these itineraries were then subject to a Passenger Fee of $5.60 in 2015. Consistent with Table 18, spikes in the distribution of American and Delta's fares occur at $2 in 2013 and $5 in 2015 for these trips. Our approach classifies all one segment trips with $2 fares before July 2014 and $5 fares after July 2014 as FFAs. All other one segment trips are not identified as FFAs (e.g., one segment trips with $0 fares are not identified as FFAs).

The "$5.00" charts in Figure 8 display the distribution of American (Panel A) and Delta's (Panel B) fares under $20 for multi-segment one-way trips and two segment roundtrips in 2013 and 2015. The Passenger Fee for one-way trips with multiple segments was $5 in 2013 and $5.60 in 2015 while the Passenger Fee for roundtrips with two segments was $5 in 2013 and $11.20 in 2015. Consistent with Table 18, fare spikes for these trips occur at $5 in 2013 (red bars) and at $5 and $11 in 2015 (blue bars). Accordingly, two segment trips with $5 fares before July 2014 and two segment trips with $5 (one-way trips) or $11 (roundtrips) fares after July 2014 are identified as FFAs. All other trips with two segments are not identified as FFAs.

The "$7.50" charts in Figure 8 display the distribution of American (Panel A) and Delta's (Panel B) fares under $20 for three segment roundtrips in 2013 and 2015. The Passenger Fee for roundtrips with three segments was $7.50 in 2013 and $11.20 in 2015. As expected, large spikes in the distribution of American and Delta's fares for three segment trips occur at $7 in 2013 (red bars) and $11 in 2015 (blue bars). In line with Table 18, we classify three segment one-way trips with $5 fares in addition to three segment roundtrips with $7 fares before July 2014 and $11 fares after July 2014 as FFAs. All other trips with three segments are not identified as FFAs.

The "$10.00" charts in Figure 8 display the distribution of American (Panel A) and Delta's (Panel B) fares under $20 for roundtrips with four or more segments in 2013 and 2015. The Passenger Fee for roundtrips with four or more segments was $10 in 2013 and $11.20 in 2015. As expected, large spikes in the distribution of American and Delta's fares for these trips occur at $10 in 2013 and $11 in 2015. In line with Table 18, roundtrips with four or more segments and $10 fares before July 2014 and $11 fares after July 2014 are classified as FFAs. All other trips with four or more segments are not identified as FFAs.

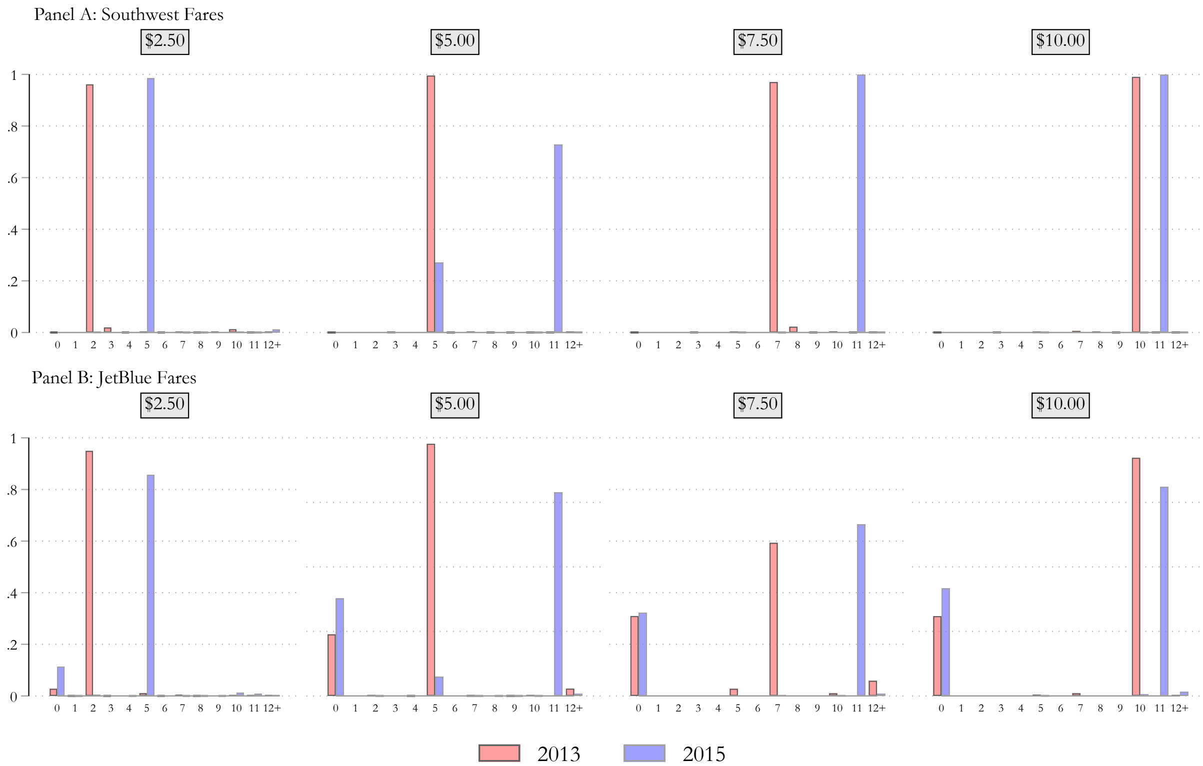
Figure 8: American and Delta Fares Under $20 by Expected Passenger Fee in 2013



Notes: Data from DB1B files for 2013 and 2015 are limited to observations on American or Delta Flights. Bars represent the share of passengers with tickets under $20 and the expected Passenger Fee in 2013 that reported that exact itinerary fare. Passenger Fees of $2.50 expected for one segment one-way flights. Passenger Fees of $5.00 expected for two or three segment one-way flights or two segment roundtrip flights. Passenger Fees of $7.50 expected for roundtrip flights with one leg having two or three segments while the other leg has only one. Passenger Fees of $10.00 expected for roundtrip flights with both legs having two or three segments.

Figure 9 is analogous to Figure 8, except that the distribution of fares under $20 in 2013 (red bars) and 2015 (blue bars) are displayed for Southwest (Panel A) and JetBlue (Panel B). Consistent with Figure 8, fare spikes for one segment trips ("$2.50" charts) occur at $2 in 2013 and $5 in 2015, at $5 in 2013 and at $5 and $11 in 2015 for two segment trips ("$5.00" charts), at $7 in 2013 and $11 in 2015 for three segment trips ("$7.50" charts), and at $10 in 2013 and $11 in 2015 for trips with four or more segments ("$10.00" charts).

Figure 9: Southwest and JetBlue Fares Under $20 by Expected Passenger Fee in 2013



Notes: Data from DB1B files for 2013 and 2015 are limited to observations on Southwest or JetBlue Flights. Bars represent the share of passengers with tickets under $20 and the expected Passenger fee in 2013 that reported that exact itinerary fare. Passenger fees of $2.50 expected for one segment one-way flights. Passenger fees of $5.00 expected for two or three segment one-way flights or two segment roundtrip flights. Passenger fees of $7.50 expected for roundtrip flights with one leg having two or three segments while the other leg has only one. Passenger fees of $10.00 expected for roundtrip flights with both legs having two or three segments.

Table 19 compares the results from using this method to identify FFAs with those reported in each airline's annual Form 10-K filings. Panel A of Table 19 is analogous to Table 16 except with revenue passenger mile calculations that are derived from our method of identifying FFAs in the DB1B data. Panel B gives the percentage point difference between Panel A and Table 16. Our method appears to identify similar proportions of FFA tickets, though there are some airline specific over and under estimate of awards. The observed differences between FFA revenue passenger miles in the DB1B data and those reported in the Form 10-Ks could be due to differences both in the number of FFAs identified and the total number of revenue passenger miles to compare to.[30] Our method of identifying frequent flyer awards is conservative as we likely underestimate the number of FFAs present in the DB1B data. For example, American and United used to charge their award program members a "close-in booking fee" to redeem frequent flyer miles for travel very close to the date of departure. Those passengers would have fares exceeding the passenger fee and would not be identified with our method. These close-in booking fees were recently eliminated by United on November 15, 2019 and American on January 15, 2020.[31] Additionally, we are not able to observe FFAs redeemed for international travel or total international revenue passenger miles as our analysis relies on the publicly available domestic version of the DB1B database.[32] If customers redeem awards disproportionately on domestic routes, then not including international data may lead to an overestimate of the percentage of revenue passenger miles from FFAs.[33] Despite these limitations, our method still appears to be identifying a decent share of FFAs.[34]

[30] It is important to note that the numbers reported in 10-K filings appear to be rounded for some airlines (e.g., Alaska, Hawaiian, JetBlue, and U.S. Airways).

[31] For more information on the elimination of these "close-in" booking fees, see `https://thepointsguy.com/news/united-pulling-award-chart/` and `https://thepointsguy.com/news/aa-eliminates-close-in-booking-fee/`.

[32] Access to the international version of the DB1B database is restricted to U.S. citizens. To access these data, U.S. citizens must submit an application to the Office of Airline Information within the Bureau of Transportation Statistics. The restricted data is unlikely to alleviate the problem of identification of international FFAs as many international itineraries include additional fees for customs, fuel, and landing.

[33] Similarly, any airline that has customers disproportionately redeeming FFAs for international trips would have a lower FFA share as a percentage of revenue passenger miles in the DB1B data.

[34] Our large overestimates of FFA share on Hawaiian Airlines may be due to not including their long distance international travel (26% of revenue in 2019). If Hawaiian Airlines has relatively few passengers that redeem FFAs for international trips then excluding them would increase the denominator for the percentage revenue passenger miles calculation while not changing the numerator. For reference, Hawaiian Airlines reported 720,000 total frequent flyer awards in 2019 (see Table A1) while we identify approximately 512,000 frequent flyer awards in the DB1B.

Table 19: Frequent Flyer Award Traffic From DB1B
(% of revenue passenger miles)

Panel A: Estimated Frequent Flyer Award Traffic From DB1B

Airline	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Alaska (AS)	9.3%	10.7%	11.2%	12.5%	11.3%	10.6%	9.7%	10.0%	10.0%	9.5%	9.1%	9.2%	8.3%	7.7%	7.2%
American (AA)	7.1%	7.1%	7.4%	8.3%	7.7%	7.9%	8.0%	7.9%	7.5%	7.0%	6.8%	5.7%	5.9%	6.8%	7.0%
Continental (CO)†	—	—	—	—	—	—	—	Acquired by United (UA)							
Delta (DL)	4.8%	4.2%	8.8%	7.0%	8.2%	7.7%	7.9%	7.8%	7.6%	7.6%	7.4%	8.2%	8.2%	8.2%	8.2%
Hawaiian (HA)	6.7%	7.7%	7.3%	9.7%	9.1%	10.0%	9.9%	9.7%	8.5%	8.1%	7.9%	8.3%	8.4%	8.6%	8.8%
JetBlue (B6)	1.5%	2.2%	2.5%	3.5%	3.6%	2.5%	2.7%	3.0%	3.4%	3.9%	4.2%	4.8%	5.6%	6.4%	7.2%
Northwest (NW)	6.2%	6.3%	6.4%	6.4%	Acquired by Delta (DL)										
Southwest (WN)	9.2%	8.6%	8.9%	9.2%	8.8%	8.9%	9.8%	10.8%	11.4%	12.4%	12.7%	13.6%	14.9%	15.5%	16.4%
United (UA)‡	8.5%	8.3%	8.7%	10.1%	9.2%	8.9%	8.9%	—	—	—	—	—	—	—	—
U.S. Airways (US)	5.3%	2.5%	3.5%	3.7%	3.2%	4.4%	4.3%	4.0%	4.1%	Acquired by American (AA)					

Panel B: Difference in Frequent Flyer Award Traffic From DB1B Compared to Reported Numbers from 10-K Filings (Table 1)

Airline	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Alaska (AS)	1.4%	2.1%	1.5%	0.1%	−3.7%	1.6%	*	*	*	*	*	*	*	*	*
American (AA)	−0.1%	−0.4%	−0.1%	−1.4%	−1.2%	−0.9%	−0.8%	−0.7%	−0.7%	1.5%	0.3%	−0.6%	−0.2%	−0.8%	−1.0%
Continental (CO)†	—	—	—	—	—	—	—	Acquired by United (UA)							
Delta (DL)	−4.2%	*	*	*	*	−0.6%	−0.3%	−0.2%	0.3%	0.2%	0.2%	0.3%	0.3%	0.0%	−0.7%
Hawaiian (HA)	*	*	*	3.7%	4.1%	4.0%	4.2%	4.5%	3.7%	2.8%	2.9%	3.3%	3.4%	2.6%	2.8%
JetBlue (B6)	−0.5%	0.2%	−0.5%	−0.5%	−0.1%	−0.5%	0.7%	0.0%	0.4%	0.9%	0.2%	0.8%	0.6%	1.4%	1.2%
Northwest (NW)	−1.4%	−1.3%	*	*	Acquired by Delta (DL)										
Southwest (WN)	2.6%	2.2%	2.7%	2.8%	1.1%	1.0%	1.2%	1.8%	1.9%	1.4%	0.7%	0.9%	1.1%	1.7%	2.3%
United (UA)‡	1.1%	0.2%	0.7%	1.0%	0.9%	1.4%	0.7%	—	—	—	—	—	—	—	—
U.S. Airways (US)	−3.8%	−1.5%	−0.5%	−0.3%	−0.8%	0.4%	0.3%	0.0%	0.6%	Acquired by American (AA)					

Source: 2005-2019 DB1B and 2005-2019 Form 10-K filings for Alaska, American, Continental, Delta, Hawaiian, JetBlue, Northwest, Southwest, United, and U.S. Airways.
Notes: *Number not reported in Form 10-K filing. —Number not calculated in DB1B. A revenue passenger mile is flown when a paying (revenue) passenger is transported one mile. † Continental appears to report FFA tickets in DB1B with an itinerary fare of zero. ‡ United also reports FFA tickets with an itinerary fare of zero after their merger with Continental (2012-2019). See Appendix B for a more detailed discussion.

4 Characteristics of frequent flyer tickets

Using the frequent flyer awards we identify, the goal of our descriptive analysis is to determine how the characteristics of frequent flyer awards differ from paid tickets. In Section 4.1, we outline the fixed effects model used to examine how FFAs differ from paid tickets. In Section 4.2, we present results from our ticket level analysis.

4.1 Descriptive analysis at the ticket level

To determine how the characteristics of frequent flyer tickets differ from paid tickets, we estimate equation (1) below,

$$y_{ijktn} = \beta_0 + \beta_1 \cdot \text{FrequentFlyer}_n + \beta_2 \cdot \text{Roundtrip}_n + \gamma_{ki} + \delta_{kj} + \theta_t + \varepsilon_{ijktn} \quad (1),$$

where y_{ijktn} is the dependent variable measured at the origin i, destination j, ticketing carrier k, quarter t, and ticket n, level. γ is an airline-origin fixed effect and δ an airline-destination fixed effect. These fixed effects control for the airline's level of dominance at the origin and destination airports. θ_t is a quarter-of-year fixed effect that controls for seasonality in our dependent variables. Dependent variables are (i) one-way distance traveled (including stopovers if any) (ii) the number of flight segments on the itinerary, (iii) indicator specifying if the ticket is nonstop, (iv) a measure of the number of competing airlines operating in the same market in that quarter, (v) the Herfindahl-Hirschman Index (HHI) for travel between origin airport i and destination airport j in quarter t, (vi) the average paid fare for travel between origin airport i and destination airport j on ticketing carrier k in quarter t, (vii) the maximum Seasonal Variation in Demand Index (SVID)[35] of the origin airport i and destination airport j, and (viii) a measure of the share of passengers on carrier k for the origin airport i and destination airport j. The coefficient of interest in our analysis is β_1, as this coefficient measures how frequent flyer award tickets differ from paid tickets with respect to the dependent variables.

[35]Following Appendix A of Li et al. (2022), SVID is calculated using monthly T100 data on passenger traffic as $\frac{\sum_{m=1,\ldots,M=12}(\frac{100 * Traffic_{a,m}}{Traffic_a} - 100)^2}{1000}$ where a refers to the airport. Large values of SVID indicate airports with considerable seasonality in passenger traffic.

During our fifteen year sample period, several airlines merged with other carriers, changed the structure of their frequent flyer programs, and introduced airline branded credit cards that enabled frequent flyer program members to accrue reward miles outside of flying. To allow the estimated effects to differ over time, equation (1) is estimated separately for each year across our sample period. All regressions are weighted by the number of passengers and standard errors are two-way clustered at the airport-pair and airline level.[36]

4.2 Results of ticket level analysis

Figure 10 displays the yearly coefficients on $FrequentFlyer$ when distance flown, number of flight segments, and the nonstop trip indicator are the dependent variables in equation (1). The blue lines in the figure display $FrequentFlyer$ coefficients when airline-origin and airline-destination fixed effects are included (our preferred specification) while the red lines display $FrequentFlyer$ coefficients when airline-origin and airline-destination fixed effects are excluded. The bars stemming from the yearly coefficients indicates the 95% confidence interval.

As demonstrated in panel (a), passengers redeeming frequent flyer awards (FFAs) traveled approximately 250 more miles than passengers traveling on paid tickets in 2005. However, this difference has steadily declined over time. By 2019, the average distance flown on FFAs was only 50 miles more than the distance flown on non-FFAs.

Panel (b) of Figure 10 displays the coefficients on $FrequentFlyer$ when the number of flight segments is the dependent variable. In 2005, FFAs had an average of 0.12 more flight segments than non-FFAs. This difference has also steadily declined over time. By 2019, FFAs involved an average of only 0.04 more segments.

Panel (c) of Figure 10 displays the coefficients on $FrequentFlyer$ when the nonstop trip indicator is the dependent variable. In 2005, FFAs were about 5% more likely to involve connecting flights. Since nonstop flights are of higher quality than connecting flights, this finding—along with the increase in flight segments found in Panel (b)–suggests that FFAs are redeemed on lower qual-

[36]Each ticket in the DB1B data contains a passenger count.

ity flights. However, this difference has also steadily declined over time. By 2019, FFAs were only 1% more likely to involve connecting flights than paid fares.

Figure 11 displays the coefficients on *FrequentFlyer* when the number of competitors (top panel) and HHI (bottom panel) are the dependent variables in equation (1). As both panels illustrate, FFAs were redeemed on less concentrated routes with approximately 0.20 more competitors in 2005. However, these differences have steadily declined over time. Relative to non-FFAs, FFAs were reedemed on routes with only 0.05 more nonstop competitors in 2019. Furthermore, FFAs and non-FFAs are not statistically different from each other in terms of the route's market concentration or the number of nonstop competitors in the specification without airline-origin and airline-destination fixed effects during a large portion of our sample period.

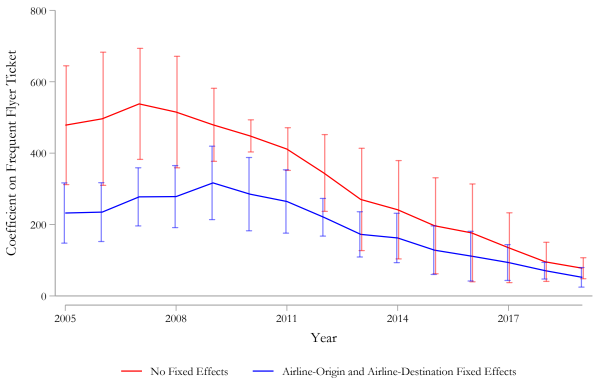
Panel (a) of Figure 12 displays *FrequentFlyer* coefficients when the average paid fare is the dependent variable. In the specification with airline-origin and airline-destination fixed effects, FFAs and non-FFAs are not statistically different from each other in terms of the average fare on routes where FFAs are redeemed. In the specification without airline-origin and airline-destination fixed effects, FFAs are redeemed on higher fare routes from 2005-2015 (routes that are $7.50-$13.50 more expensive). However, consistent with the specification with airline-origin and airline-destination fixed effects, FFAs and non-FFAs are not statistically different from each other in terms of the average paid fare by the end of our sample period (2016-2019). The disagreement between the models with and without airline-origin and airline-destination fixed effects in the pre-2016 period suggest that FFAs are disproportionately on routes to or from particular cities.

Panel (b) of Figure 12 displays *FrequentFlyer* coefficients when the maximum value of the Seasonal Variation In Demand (SVID) measure between the origin and destination airports is the dependent variable. High values of SVID indicate airports with substantial seasonal variation in demand (e.g., ski destinations such as Aspen (ASE), Eagle County (EGE), Jackson Hole (JAC), and Telluride (MTJ) where demand spikes during the winter). As demonstrated in panel (b), FFAs are disproportionately redeemed on more seasonal routes.

Panel (c) of Figure 12 displays *FrequentFlyer* coefficients when the carrier share at the origin

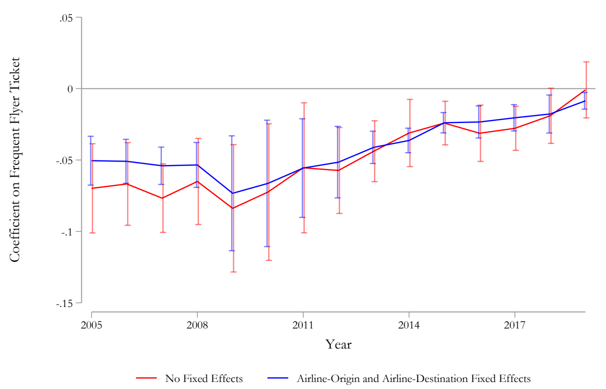
Figure 10: Difference in Distance, Flight Segments, and Fraction of Nonstop Tickets

(a) Difference in Number of Miles Flown



(b) Difference in the Number of Flight Segments

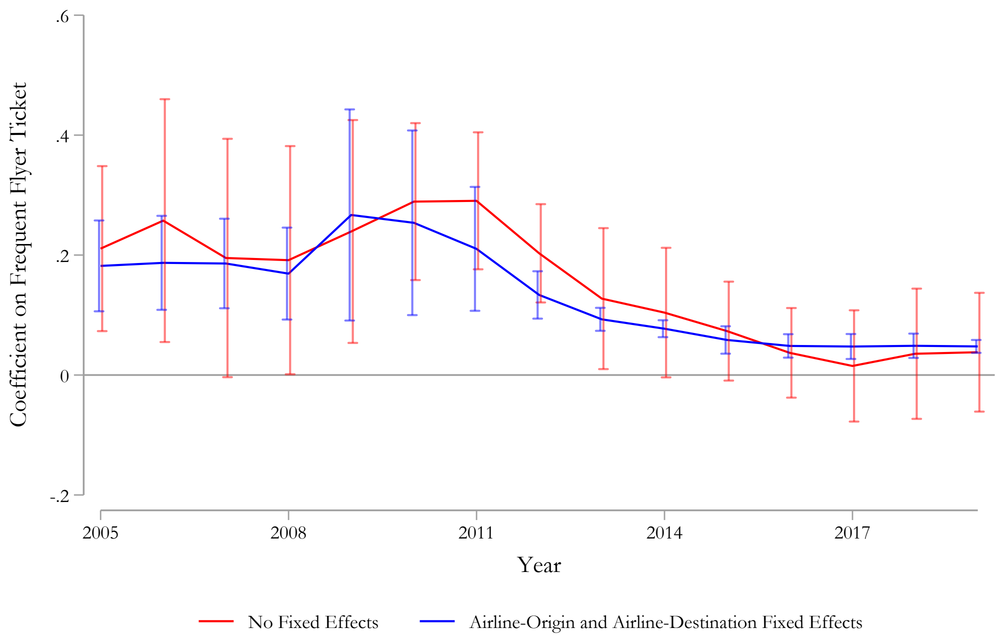
(c) Difference in the Fraction of Nonstop Tickets



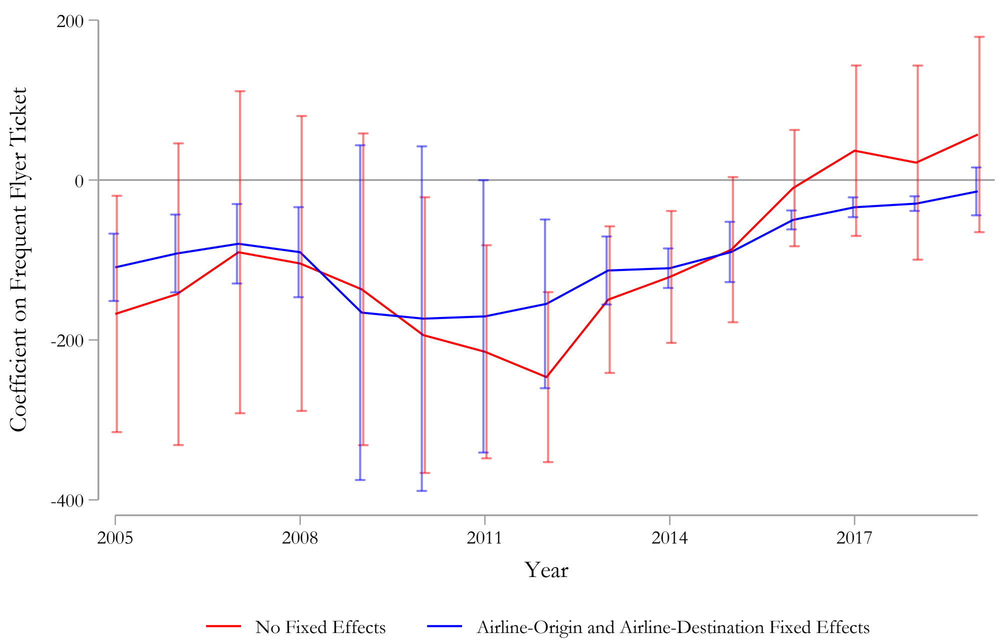
Notes: Charts display the yearly coefficients on $FrequentFlyer$ for regressions with the respective dependent variable for each panel. All regressions include controls for roundtrip status and airline. The blue lines in the figure display $FrequentFlyer$ coefficients when airline-origin and airline-destination fixed effects are included (our preferred specification) while the red lines display $FrequentFlyer$ coefficients when airline-origin and airline-destination fixed effects are excluded. The bars stemming from the yearly coefficients indicates the 95% confidence interval. Data are from the DB1B (2005-2019) and limited to one-way tickets with three or fewer segments and round-trip tickets with six or fewer segments.

Figure 11: Difference in Market Structure

(a) Difference in the Number of Competitors



(b) Difference in the Herfindahl-Hirschman Index (HHI)



Notes: Charts display the yearly coefficients on $FrequentFlyer$ for regressions with the respective dependent variable for each panel. All regressions include controls for round-trip status and airline. The blue lines in the figure display $FrequentFlyer$ coefficients when airline-origin and airline-destination fixed effects are included (our preferred specification) while the red lines display $FrequentFlyer$ coefficients when airline-origin and airline-destination fixed effects are excluded. The bars stemming from the yearly coefficients indicates the 95% confidence interval. Data are from the DB1B (2005-2019) and limited to one-way tickets with three or fewer segments and round-trip tickets with six or fewer segments.

and destination airports are the dependent variables. In this panel, the red line displays the results when the dependent variable is the origin airport own-carrier share while the blue line displays the results from the destination airport own-carrier share. This specification suggests that FFAs are more likely to originate at airports with higher own-carrier shares (e.g., hub airports) where it may be easier for consumers to accrue frequent flyer miles. These FFAs are disproportionately used to travel to destinations with lower own-carrier shares which may reflect more leisure or seasonal destinations.

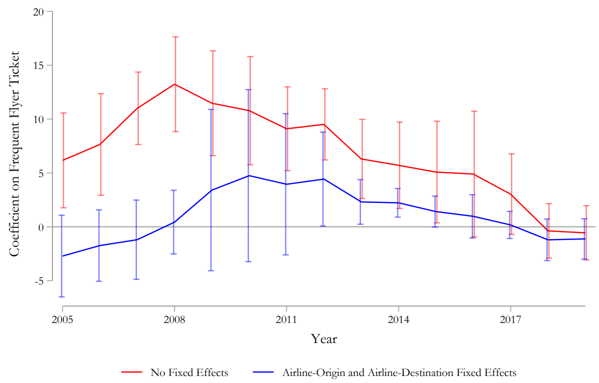
To further illustrate that FFAs are disproportionately redeemed on seasonal routes to leisure destinations, Figure 13 displays Delta's route segments with large shares of frequent flyer passengers ($\geq 15\%$) in the first (January-March) and third quarters (July-September) of 2016. In the first quarter of 2016, these segments include routes from Delta's hubs in Atlanta (ATL), Los Angeles (LAX), Minneapolis (MSP), and New York City (JFK) to winter vacation destinations such as Honolulu (HNL), Maui (OGG), Kauai (LIH), and Kona (KOA) in Hawaii in addition to ski destinations such as Aspen (ASE), Eagle County (EGE), Gunnison (GUC), and Telluride (MTJ) in Colorado (see panel (a)). In the third quarter of 2016, these segments shifted further north from ski destinations in Colorado to summer vacation destinations near Glacier, Grand Teton, and Yellowstone National Parks in Montana and Wyoming (see panel (b)). For example, these third quarter segments include routes from Delta's largest hub in Atlanta (ATL) to Bozeman Yellowstone (BZN), Glacier Park (FCA), Jackson Hole (JAC), and Missoula (MSO).

5 Characteristics of markets with large shares of frequent flyer passengers

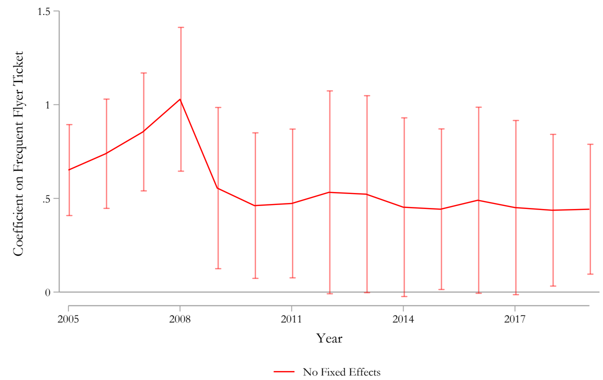
Using the FFAs we identify, our market level analysis examines how various market structure and product quality characteristics affect the share of an airline's passengers traveling on FFAs. In Section 5.1, we outline the instrumental variables strategy used to identify how market and product quality characteristics affect the share of frequent flyer passengers. In Section 5.2, we

Figure 12: Difference in Average Fare, SVID, and Carrier Share

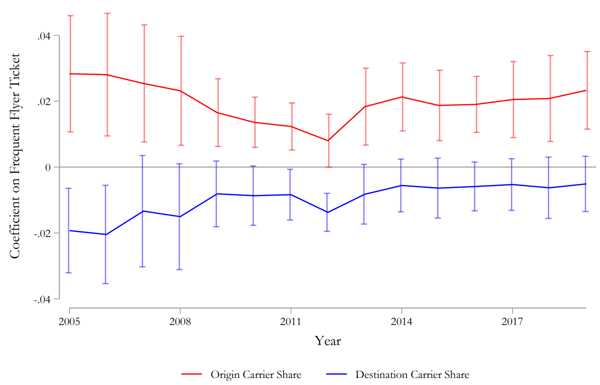
(a) Difference in the Average Fare



(b) Difference in the Seasonal Variation in Demand (SVID) of the Destination Airport



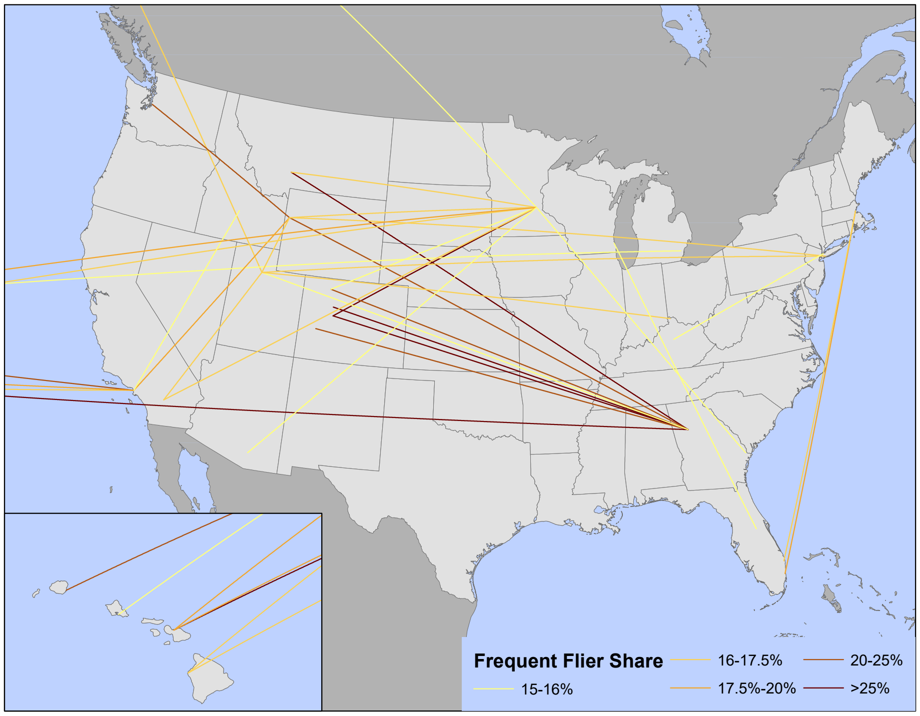
(c) Difference in Carrier Share



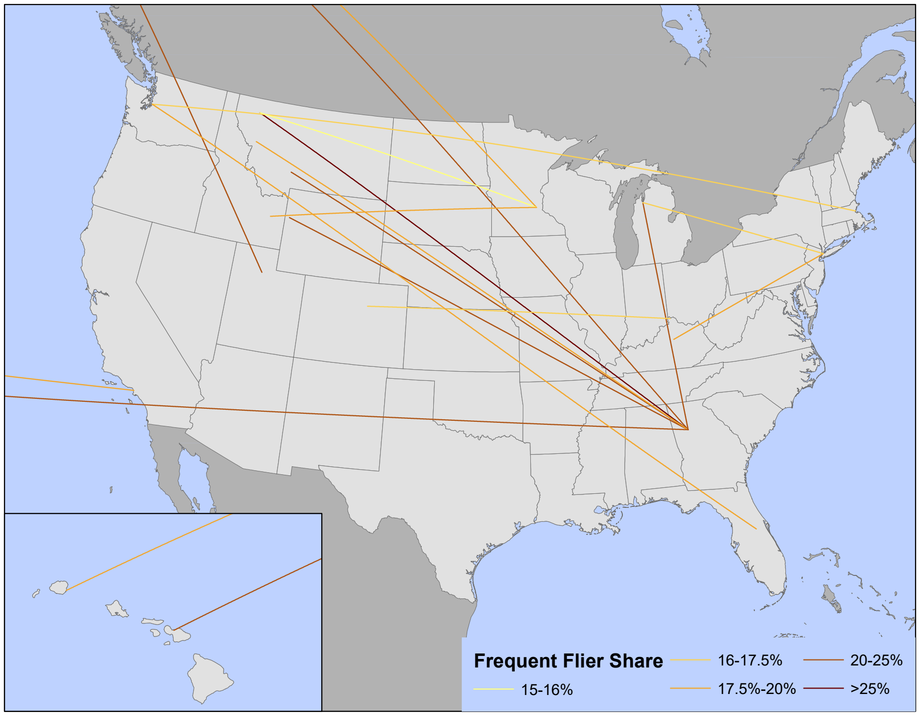
Notes: Charts display the yearly coefficients on $FrequentFlyer$ for regressions with the respective dependent variable for each panel. All regressions include controls for round-trip status and airline. The blue line in Panel (a) displays $FrequentFlyer$ coefficients when airline-origin and airline-destination fixed effects are included (our preferred specification) while the red lines in Panel (a) and (b) display $FrequentFlyer$ coefficients when airline-origin and airline-destination fixed effects are excluded. The bars stemming from the yearly coefficients indicates the 95% confidence interval. In Panel (c), the red line displays $FrequentFlyer$ coefficients for the regressions with origin airport own-carrier share as the dependent variable. The blue line in Panel (c) displays $FrequentFlyer$ coefficients for the regressions with destination airport own-carrier share as the dependent variable. Data for all panels are from the DB1B (2005-2019) and limited to one-way tickets three or fewer segments and round-trip tickets with six or fewer segments.

Figure 13: Delta Routes with Large Shares of Frequent Flyer Passengers in 2016

(a) 2016 Quarter 1



(b) 2016 Quarter 3



Notes: Data are from the DB1B and limited to one-way tickets with three or fewer segments and round-trip tickets with six or fewer segments on Delta flights. Only segments with more than 15% of Delta passengers in that quarter identified as traveling on FFAs on routes with at least 500 passengers are included.

present results from our market level analysis.

5.1 Empirical strategy

We define a market m, as a unique combination of origin airport i, and destination airport j. This definition is directional, meaning that flights between Los Angeles (LAX) and New York (JFK) constitute two different markets depending on the direction flown. We define a product p, as a specific routing (nonstop or sequence of connecting airport(s)) in a market. Letting k denote the ticketing carrier and t the quarter of travel, an observation in our market level analysis is a unique combination of product, market, ticketing carrier, and quarter.

To determine how various market and product quality characteristics affect the share of an airline's passengers traveling on FFAs, we employ a fixed effects strategy to control for unobservable confounders that may influence demand. We estimate our empirical model using two sets of fixed effects.

Foremost, we include airline-quarter-year fixed effects to control for unobserved time-varying airline-specific effects. These fixed effects control for changes in an airline's service quality or operating costs that occur during our sample period such as changes in frequent flyer programs, onboard amenities, formation of codeshare alliances, and mergers.

We also include airline-market fixed effects to control for unobserved time-invariant factors that affect an airline's demand on particular routes. Additionally, this fixed effect controls for the underlying effect of an airline's level of dominance at the origin and destination airports.

Incorporating both of these fixed effects, we estimate the following reduced-form demand equation,

$$\text{FF\_SHARE}_{pmkt} = \beta_0 + \beta_1 \cdot \text{FARE}_{pmkt} + \beta_2 \cdot \text{FARE}_{pmkt} \cdot \text{HUB\_ORIGIN}_{ikt} + \beta_3 \cdot \text{HHI}_{mt} + \beta_4 \cdot \text{NONSTOP}_{pmkt} + \beta_5 \cdot \text{ROUNDTRIP}_{pmkt} + \beta_6 \cdot \text{ROUTING\_QUALITY}_{pmkt} + \beta_7 \cdot \text{LOAD\_FACTOR}_{pmkt} + \gamma_{kt} + \delta_{km} + \varepsilon_{pmkt} \quad (2),$$

where FF_SHARE_{pmkt} is the share of passengers redeeming frequent flyer awards for product p, in market m, on ticketing carrier k, and quarter t. γ an airline-quarter fixed effect and δ is an airline-market fixed effect. FARE is the average fare paid by revenue (i.e., non-frequent flyer) passengers for product p on carrier k in quarter t. HUB_ORIGIN is an indicator specifying if origin airport i is a hub for carrier k in quarter t. NONSTOP is an indicator specifying if product p is nonstop (i.e., ticket does not involve a connection) while ROUNDTRIP is an indicator specifying if product p is for roundtrip travel. HHI is the Herfindahl-Hirschman Index for that market and quarter calculated on a scale from 0 to 1.

Following Chen and Gayle (2019), ROUTING_QUALITY is defined as the percentage ratio of the product's itinerary flight distance to the minimum flight distance in the market. Assuming that passengers prefer shorter travel times to longer travel times, the closer the product's distance to the minimum flight distance, the more desirable the product is to passengers.[37] Finally, LOAD_FACTOR is defined as the percentage ratio of seats filled to the total number of available seats for product p on carrier k in quarter t. Load factors are constructed using the T-100 domestic segment database available from the Bureau of Transportation Statistics. For products involving multiple flight segments, the maximum load factor across all trip segments is used.

FARE and FARE·HUB_ORIGIN in equation (2) are likely endogenous. For example, frequent flyer promotions that apply to a particular city will be correlated with both an airline's fares and the number of frequent flyer passengers that depart from that city. In addition, airlines may intentionally restrict frequent flyer award capacity on high fare routes to encourage passengers to redeem awards on low fare routes. In an analogous manner, passengers who wish to extract the maximum value from their frequent flyer miles may self-select into high fare routes when redeeming reward tickets. To correct for the likely endogeneity of fare, we estimate equation (2) using two-stage least squares (2SLS). Following Gayle and Xie (2019), we instrument for fare using (i) the number of competing products offered by other carriers with an equivalent number of connections and (ii) the interaction between the product's distance and the quarterly jet fuel price.[38]

Our first instrument measures the degree of market competition a product faces, which affects the size of the price markup over cost. The rationale for our second instrument is that jet fuel prices

[37] The lowest value ROUTING_QUALITY can take is 100 when the flight distance is the same as that of a nonstop flight. Higher values for ROUTING_QUALITY thus indicate more circuitous (and longer) routes.

[38] U.S. Gulf Coast Kerosene-Type Jet Fuel Spot Price (dollars per gallon) data come from the Energy Information Administration.

and distance are correlated with the marginal cost of providing airline service, which in turn affects the overall fare.

The validity of our instruments rely on the fact that the number of products offered by carriers in a market is predetermined at the time demand shocks occur, which implies that the instruments are uncorrelated with the error term. Moreover, the number of products offered and their associated non-price characteristics are not easily adjusted in the short-run, which mitigates the influence of demand shocks on the number of products and their non-price characteristics.

5.2 Results of market level analysis

Table 20 presents results from the model specified by equation (2). To prevent airline markets and airline products with low amounts of passenger traffic from disproportionately affecting our results, we limit our analysis to airline markets with at least 500 passengers in a quarter and airline products with at least 50 passengers in a quarter.

In column (1), the endogeneity of FARE and FARE·HUB_ORIGIN are ignored as equation (2) is estimated using ordinary least squares (OLS). In this specification, the coefficients on our market variables (FARE, FARE·HUB_ORIGIN, HHI) are statistically insignificant. Several of our product quality variables (ROUNDTRIP and LOAD_FACTOR) are also statistically insignificant. However, two measures of product quality display negative and statistically significant coefficients. The negative and statistically significant coefficient on NONSTOP indicates fewer frequent flyer passengers are present on direct flights. Conversely, the negative coefficient on ROUTING_QUALITY indicates that products with higher routing quality have larger shares of frequent flyer awards (FFAs).[39] These results are consistent with the idea that airlines may limit the number of FFAs that can be redeemed on direct flights to below that of market demand. Once those seats are filled by frequent flyer passengers, individuals are forced to redeem awards on connecting itineraries.

[39]Because ROUTING_QUALITY is defined as the ratio of the product's itinerary distance to the minimum distance in the market, a ratio equal to one indicates that the product's distance equals the minimum distance (i.e., highest routing quality). A ratio larger than one indicates that the product is not the highest routing quality option in the market.

Frequent flyer passengers on connecting routes still prefer more efficient routing, and thus we find higher shares of FFAs on these routes once we limit to connecting itineraries.

In column (2), we correct for the endogeneity of FARE and FARE·HUB_ORIGIN by estimating equation (2) using 2SLS. The magnitude of the coefficient on FARE is now statistically significant at the 10% level, proving suggestive evidence that unobserved factors that affect both an airline's fares and its share of FFAs are biasing the estimates presented in column (1). Moreover, the positive coefficient on FARE indicates that high fare products have larger shares of FFAs. In particular, a \$100 increase in fare increases an airline's share of FFAs on routes by 1.04%.[40]

The coefficients on NONSTOP and LOAD_FACTOR are also statistically significant at conventional levels once FARE and FARE·HUB_ORIGIN are instrumented for in column (2) of Table 20. The negative coefficient on NONSTOP implies that connecting airline products have larger shares of FFAs. This finding suggests that passengers may prefer to redeem reward tickets on routes that require connections (e.g., trips to outlying Hawaiian islands) or that carriers are effective at ensuring that passengers redeem reward tickets on lower quality connecting products. In addition, the negative coefficient on LOAD_FACTOR indicates that routes with high load factors have lower shares of FFAs. A 10% increase in load factor implies a 0.45% decrease in an airline's share of FFAs.[41] This finding suggests that airlines may restrict reward capacity on routes with high load factors. Finally, the positive and statistically significant coefficient on HHI shows that less competitive routes have higher shares of FFAs. This finding is consistent with passengers redeeming FFAs on more seasonal routes which often have fewer competitors.[42] It is also possible that passengers who routinely fly on less competitive routes may accrue rewards faster than passengers flying more competitive routes as individuals are more likely to fly the same carrier repeatedly, thereby accruing a particular airline's frequent flyer miles faster.

[40] This result is roughly a 15.7% increase over the mean frequent flyer share (1.04/6.63).

[41] This result is roughly a 6.8% increase over the mean frequent flyer share (0.45/6.63).

[42] As shown in Figure 13, many segments with high shares of frequent flyer passengers go from airline hubs to vacation destinations. These routes likely have less competition.

Table 20: Share of Frequent Flyer Passenger Results

Dependent variable: FF Share (Mean=6.63%)	OLS (1)	2SLS (2)
FARE	0.0022	0.0227*
	(0.0019)	(0.0117)
FARE·HUB_ORIGIN	0.0014	-0.0557
	(0.0017)	(0.0341)
HHI	-0.6466	1.7408**
	(0.4544)	(0.7425)
NONSTOP	-0.5653***	-1.0060***
	(0.1184)	(0.2306)
ROUNDTRIP	0.2217	-1.0765
	(1.2409)	(2.1417)
ROUTING_QUALITY	-0.0236***	-0.0301***
	(0.0050)	(0.0045)
LOAD_FACTOR	-0.0204	-0.0454***
	(0.0142)	(0.0093)
Observations	4,708,692	4,708,692

Notes: All regressions include airline-quarter-year, and airline-market (i.e., airport-pair) fixed effects. Two-way clustered standard errors at the airline and market level are reported in parentheses. The sampling period is 2005Q1-2019Q4. Fares are deflated using the Bureau of Economic Analysis Gross Domestic Product: Implicit Price Deflator (GDPDEF) indexed to quarter 4 of 2019. In column (2), FARE and FARE·HUB_ORIGIN are treated as endogenous variables. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

There is a concern that HHI is endogenous in columns (1) and (2) of Table 20. For example, markets with high fares due to a lack of competition may be attractive for entry. These markets may also be unattractive for entry if high fares are a result of entry barriers such as slot controls or limited access to gates at the endpoint airports (Luttmann and Nehiba, 2020). For two reasons, we believe that the simultaneity bias that results from an airline's decision to enter or exit a given route is not substantially biasing our results. Foremost, Gayle and Wu (2013) show that this simultaneity bias is negligible in their model that accounts for endogenous entry in U.S. airline markets. Second, the inclusion of airline-quarter-year and airline-market fixed effects likely eliminates a large degree of bias associated with the correlation between HHI and the regression error term.

Based on the descriptive analysis in Section 4, we have reason to believe that the relationship between these variables and frequent flyer shares may change over time. Table 21 breaks our fifteen year sample period into three different five year subsamples and presents results for each subsample. Comparing the 2SLS specifications across these three subsamples (columns 2, 4 and 6) indicates that the sign of the effects are consistent, but the coefficients decline in magnitude and significance over time. In the 2005 to 2009 subsample, the fare coefficients (FARE and FARE·HUB_ORIGIN) suggest that carriers are effective at steering passengers towards redeeming reward tickets on low fare routes in markets where the carrier has substantial market power (i.e., its hub airports). In markets where the carrier is not dominant, passengers are able to redeem reward tickets on higher fare routes. In later subsamples, these fare coefficients become statistically insignificant. Across all three subsamples, the coefficient on ROUTING_QUALITY is negative and statistically significant, indicating that products with higher routing quality have larger shares of frequent flyer awards (FFAs).[43] These results provide further evidence of airlines and consumers altering their approaches towards FFAs over the past fifteen years.

[43] Because ROUTING_QUALITY is defined as the ratio of the product's itinerary distance to the minimum distance in the market, a ratio equal to one indicates that the product's distance equals the minimum distance (i.e., highest routing quality). A ratio larger than one indicates that the product is not the highest routing quality option in the market.

Table 21: Share of Frequent Flyer Passenger Results By Subsample

	2005-2009		2010-2014		2015-2019	
Dependent variable:	OLS	2SLS	OLS	2SLS	OLS	2SLS
FF Share	(1)	(2)	(3)	(4)	(5)	(6)
FARE	0.0022*	0.0482*	0.0022	0.0232	0.0019	0.0074
	(0.0011)	(0.0227)	(0.0021)	(0.0179)	(0.0024)	(0.0238)
FARE·HUB_ORIGIN	0.0020*	-0.1016*	0.0010	-0.0984	0.0036	-0.0445
	(0.0010)	(0.0489)	(0.0016)	(0.06234)	(0.0037)	(0.0715)
HHI	-1.4438***	-0.2665	-0.9601	0.9860	-0.7265*	-1.2932
	(0.3814)	(0.9852)	(0.6652)	(1.0783)	(0.3981)	(0.8461)
NONSTOP	-0.9421***	-2.3758***	-0.7633***	-1.7938***	0.0641	-0.5199
	(0.1567)	(0.7191)	(0.2387)	(0.3101)	(0.2034)	(0.3387)
ROUNDTRIP	3.3500***	1.7159	-0.1564	-0.6996	-0.8425	-0.1370
	(0.7952)	(1.0148)	(1.5140)	(3.1756)	(1.0887)	(3.6122)
ROUTING_QUALITY	-0.0171***	-0.0494***	-0.0187**	-0.0202***	-0.0422***	-0.0381***
	(0.0054)	(0.0161)	(0.0077)	(0.0489)	(0.0065)	(0.0053)
LOAD_FACTOR	-0.0452***	-0.0774***	-0.0166	-0.0774**	-0.0135	-0.0234
	(0.0079)	(0.0243)	(0.0146)	(0.0059)	(0.0227)	(0.0199)
Observations	1,167,753	1,167,753	1,655,743	1,655,743	1,884,922	1,884,922

Notes: All regressions include airline-quarter-year, and airline-market (i.e., airport-pair) fixed effects. Two-way clustered standard errors at the airline and market level are reported in parentheses. Fares are deflated using the Bureau of Economic Analysis Gross Domestic Product: Implicit Price Deflator (GDPDEF) indexed to quarter 4 of 2019. In columns (2, 4 and 6), FARE and FARE·HUB_ORIGIN are treated as endogenous variables. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

6 Frequent flyer redemptions and price dispersion

As discussed in Section 1, it is ambiguous whether loyalty programs increase or decrease competition in markets where they are present. While many argue that these programs are anticompetitive because introducing switching costs generally raises prices, creates barriers to entry, and generates deadweight losses (Banerjee and Summers, 1987; Cairns and Galbraith, 1990; Fong and Liu, 2011; Kim et al., 2001; Klemperer, 1995), others argue that loyalty programs are “business stealing devices” that enhance competition and increase total surplus (Caminal, 2012; Caminal and Claici, 2007; Caminal and Matutes, 1990).

Similar to the large body of literature that examines the relationship between competition and price discrimination, we examine whether the share of passengers traveling on frequent flyer awards in a market increases or decreases an airline's ability to price discriminate. If loyalty programs enhance market power, then an increase in the share of frequent flyer passengers in a market should increase an airline's ability to price discriminate (i.e., the dispersion of that carrier's fares in the market should increase). However, if loyalty programs enhance competition, then an increase in the share of frequent flyer passengers should decrease an airline's ability to price discriminate (i.e., the dispersion of that carrier's fares in the market should decrease).

6.1 Price dispersion regression specification

Following Borenstein and Rose (1994), Gerardi and Shapiro (2009), Dai et al. (2014) and several other previous studies of airline pricing, we rely on the Gini coefficient as our measure of price dispersion.[44] Consistent with prior literature (and in an effort to focus on the effect of FFAs on paid tickets), we calculate the gini coefficient for each nonstop route excluding all fares under \$20. The median Gini coefficient we find for 2005-2019 is 0.24, which is similar to the median coefficients for nonstop routes in Gerardi and Shapiro (2009) (0.22) and Dai et al. (2014) (0.23). To determine how competition and the share of frequent flyer passengers affects price dispersion, we estimate equation (3) below,

$$ln[Gini_{mkt}/(1-Gini_{mkt})] = \beta_0 + \beta_1 \cdot \text{HHI}_{mt} + \beta_2 \cdot \text{FREQUENT FLYER SHARE}_{mkt} + \gamma_{kt} + \delta_{km} + \varepsilon_{mkt} \qquad (3),$$

where Gini_{mkt} is the Gini coefficient of ticketing carrier k's fares in market m and quarter t. To ensure that a linear estimator can be used to estimate equation (3), we employ the Gini log-odds ratio as our dependent variable to unbound the inequality index. γ is an airline-year-quarter fixed effect that controls for any unobserved time-varying airline-specific effects (e.g., carrier-specific demand shocks). δ is an airline-market fixed effect that controls for unobserved time-invariant factors that affect an airline's demand in a given market. HHI is the Herfindahl-Hirschman Index

[44] In our context, the Gini coefficient measures how far the distribution of an airline's fares on a route deviates from a completely equal distribution. Specifically, the Gini coefficient is equal to twice the expected absolute difference between two fares that are randomly drawn from the population. For example, a Gini coefficient of 0.20 for a given carrier and route implies an expected absolute difference of 40 percent of the mean fare for two randomly selected passengers traveling on that carrier and route.

for the market measured on a scale from 0 to 1. The coefficient of interest in our analysis is β_2, as this coefficient measures how the share of frequent flyer passengers (measured on a 0 to 100 scale) on a carrier in a given market affects that carrier's fare dispersion. In addition, we estimate the same equation with the natural log of the 10th and 90th percentile fares as the dependent variable to better understand how competition and the share of frequent flyer passengers affect prices across the distribution.

HHI in equation (3) is likely endogenous. For example, an airline may be more likely to enter a route that has higher levels of price dispersion as they anticipate the ability to undercut some of the prices. To address this endogeneity, we estimate equation (3) using 2SLS. Following Borenstein and Rose (1994), Gerardi and Shapiro (2009), and Dai et al. (2014), we instrument for HHI using (i) measures of passengers enplaned on the route[45] and (ii) measures of the end-point city populations.[46] Collectively, these instruments are likely unrelated to the price dispersion of a particular route, but capture the effect on competition between airlines through their entry and exit decisions.

6.2 Results of price dispersion regression

Table 22 presents results from the model specified by equation (3) in column (1) as well as the same model estimated on the log of the 10th percentile fare (column (2)) and the log of the 90th percentile fare (column (3)). Consistent with prior literature (Gerardi and Shapiro (2009) and Dai et al. (2014)) we limit our analyses exclusively to nonstop flights. To prevent airline markets with low amounts of passenger traffic from disproportionately affecting our results and to ensure a large enough dispersion of prices to calculate a gini coefficient, we limit our analysis to airline markets with at least 1000 passengers in a quarter.[47]

In column (1), our results indicate that higher levels of Frequent Flyer passengers on a route

[45] Specifically, we use the natural log of the total passenger traffic in the market across all carriers and the "genp" measure introduced by Borenstein and Rose (1994).

[46] We use the natural log of both the geometric and arithmetic means of Metropolitan Statistical Area (MSA) populations of the endpoint airports.

[47] As the DB1B is a 10% sample of all tickets, this corresponds to at least 100 observations in each quarter.

increase the price dispersion of paying passengers. Comparing columns (2) and (3) indicates that this increase in price dispersion is driven by larger fare increases in the 90th percentile of fares as opposed to the 10th percentile of fares. This finding supports the theory that loyalty programs may enhance market power, as increases in the share of frequent flyers passengers on a route increases an airline's ability to price discriminate. Additionally, these results show higher levels of FFAs increase fares across the price distribution supporting the assertion that FFAs serve as quantity discounts for customers who exhibit brand loyalty. Thus, consumers are helping to subsidize increasing their own switching costs by paying higher ticket prices in the hopes of receiving future free travel.

Table 22: Price Dispersion Results

Dependent variable:	$ln[Gini/(1-Gini)]$ (1)	ln 10th Percentile Fare (2)	ln 90th Percentile Fare (3)
FREQUENT FLYER SHARE	0.0017***	0.0054***	0.0084***
	(0.0002)	(0.0003)	(0.0003)
HHI	-0.1571***	0.8750***	0.8296***
	(0.0017)	(0.0270)	(0.0249)
Observations	405,562	405,562	405,562

Notes: All regressions include airline-quarter-year, and airline-market (i.e., airport-pair) fixed effects. The analysis sample is limited to nonstop routes with at least 100 passengers in the DB1B for that quarter. Standard errors clustered at the market level are reported in parentheses. The sampling period is 2005Q1-2019Q4. Fares are deflated using the Bureau of Economic Analysis Gross Domestic Product: Implicit Price Deflator (GDPDEF) indexed to quarter 4 of 2019. In all regressions, HHI is treated as endogenous. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

7 Conclusion

Firms introduce loyalty programs to attract and retain customers, but it is unclear whether these programs enhance or reduce competition in markets where they are present. This paper introduces a novel method to identify rewards for one of the most prominent industries that employ loyalty programs, airlines. Using this new method, we find that frequent flyer awards (FFAs) are disproportionately redeemed on less concentrated (i.e., more competitive) routes to leisure and seasonal destinations, suggesting that passengers either have a preference for redeeming award tickets for travel to vacation destinations or airlines are successful in restricting capacity on more concentrated routes where price markups are expected to be higher. In comparison to paid tickets, we also find that FFAs have more flight segments, are longer on average, and more likely to be roundtrip. However, these differences appear to have declined over time. Additionally, we find that FFAs are disproportionately redeemed on routes from origins with higher carrier specific concentrations of passengers, supporting the classic hub premium story that fares for routes leaving from hub airports are higher due to the value of frequent flyer miles.

We also show that airlines appear to limit FFAs on more competitive and more direct routes. Moreover, we find that routes with low load factors have higher shares of frequent flyer passengers, suggesting that airlines may restrict a customer's ability to redeem awards on densely traveled routes. Finally, we also find that price dispersion increases on routes with higher shares of frequent flyer passengers. This finding suggests that airline loyalty programs may enhance market power, as they appear to increase an airline's ability to price discriminate.

Our paper contributes to the literature on both loyalty programs and the airline industry in general. Our novel approach for identifying loyalty rewards could be used to shed more light on the role of loyalty programs and their impact on market structure, individual decisions, competition, and consumer welfare. For example, our method could serve as a starting point for empirical researchers interested in exploring questions such as the effects of mergers and codesharing on FFAs, how the timing and introduction of airline branded credit cards impacts reward redemptions, and how frequent flyer program devaluations over time have affected consumer welfare, among others.

References

Azar, J., Schmalz, M. C., and Tecu, I. (2018). Anticompetitive effects of common ownership. *The Journal of Finance*, 73(4):1513–1565.

Banerjee, A. V. and Summers, L. H. (1987). *On frequent flyer programs and other loyalty-inducing economic arrangements*. Harvard Institute of Economic Research.

Basso, L. J., Clements, M. T., and Ross, T. W. (2009). Moral hazard and customer loyalty programs. *American Economic Journal: Microeconomics*, 1(1):101–23.

Berry, S. and Jia, P. (2010). Tracing the woes: An empirical analysis of the airline industry. *American Economic Journal: Microeconomics*, 2(3):1–43.

Bilotkach, V. and Pai, V. (2016). Hubs versus airport dominance. *Transportation Science*, 50(1):166–179.

Borenstein, S. (1989). Hubs and high fares: Dominance and market power in the U.S. airline industry. *The RAND Journal of Economics*, pages 344–365.

Borenstein, S. and Rose, N. L. (1994). Competition and price dispersion in the US airline industry. *Journal of Political Economy*, 102(4):653–683.

Brueckner, J. K., Lee, D., and Singer, E. S. (2013). Airline competition and domestic US airfares: A comprehensive reappraisal. *Economics of Transportation*, 2(1):1–17.

Brueckner, J. K., Lee, D. N., Picard, P. M., and Singer, E. (2015). Product unbundling in the travel industry: The economics of airline bag fees. *Journal of Economics & Management Strategy*, 24(3):457–484.

Brueckner, J. K., Lee, D. N., and Singer, E. S. (2011). Alliances, codesharing, antitrust immunity, and international airfares: Do previous patterns persist? *Journal of Competition Law & Economics*, 7(3):573–602.

Brueckner, J. K. and Singer, E. (2019). Pricing by international airline alliances: A retrospective study. *Economics of Transportation*, 20:100139.

Brueckner, J. K. and Whalen, W. T. (2000). The price effects of international airline alliances. *The Journal of Law and Economics*, 43(2):503–546.

Cairns, R. D. and Galbraith, J. W. (1990). Artificial compatibility, barriers to entry, and frequent-flyer programs. *Canadian Journal of Economics*, pages 807–816.

Calzaretta Jr, R. J., Eilat, Y., and Israel, M. A. (2017). Competitive effects of international airline cooperation. *Journal of Competition Law & Economics*, 13(3):501–548.

Caminal, R. (2012). The design and efficiency of loyalty rewards. *Journal of Economics & Management Strategy*, 21(2):339–371.

Caminal, R. and Claici, A. (2007). Are loyalty-rewarding pricing schemes anti-competitive? *International Journal of Industrial Organization*, 25(4):657–674.

Caminal, R. and Matutes, C. (1990). Endogenous switching costs in a duopoly model. *International Journal of Industrial Organization*, 8(3):353–373.

Carlton, D., Israel, M., MacSwain, I., and Orlov, E. (2017). Are legacy airline mergers pro-or anti-competitive? Evidence from recent U.S. airline mergers. *International Journal of Industrial Organization*.

Chen, Y. and Gayle, P. G. (2019). Mergers and product quality: Evidence from the airline industry. *International Journal of Industrial Organization*, 62:96–135.

Ciliberto, F., Watkins, E., and Williams, J. W. (2019). Collusive pricing patterns in the US airline industry. *International Journal of Industrial Organization*, 62:136–157.

Ciliberto, F. and Williams, J. W. (2010). Limited access to airport facilities and market power in the airline industry. *The Journal of Law and Economics*, 53(3):467–495.

Ciliberto, F. and Williams, J. W. (2014). Does multimarket contact facilitate tacit collusion? Inference on conduct parameters in the airline industry. *The RAND Journal of Economics*, 45(4):764–791.

Dai, M., Liu, Q., and Serfes, K. (2014). Is the effect of competition on price dispersion nonmonotonic? Evidence from the U.S. airline industry. *Review of Economics and Statistics*, 96(1):161–170.

De Jong, G., Behrens, C., and van Ommeren, J. (2019). Airline loyalty (programs) across borders: A geographic discontinuity approach. *International Journal of Industrial Organization*, 62:251–272.

Escobari, D. (2011). Frequent flyer programs premium and the role of airport dominance. *Applied Economics Letters*, 18(16):1565–1569.

Fong, Y.-f. and Liu, Q. (2011). Loyalty rewards facilitate tacit collusion. *Journal of Economics & Management Strategy*, 20(3):739–775.

Fukui, H. (2019). How do slot restrictions affect airfares? New evidence from the US airline industry. *Economics of Transportation*, 17:51–71.

Gayle, P. G. (2013). On the efficiency of codeshare contracts between airlines: Is double marginalization eliminated? *American Economic Journal: Microeconomics*, 5(4):244–73.

Gayle, P. G. and Thomas, T. (2016). Assessing firm behavior in carve-out markets: Evidence on the impact of carve-out policy. *Journal of Economic Behavior & Organization*, 128:178–194.

Gayle, P. G. and Wu, C.-Y. (2013). A re-examination of incumbents' response to the threat of entry: Evidence from the airline industry. *Economics of Transportation*, 2(4):119–130.

Gayle, P. G. and Xie, X. (2018). Entry deterrence and strategic alliances. *Economic Inquiry*, 56(3):1898–1924.

Gayle, P. G. and Xie, X. (2019). Firms' markup, cost, and price changes when policymakers permit collusion: Does antitrust immunity matter? *Journal of Economic Behavior & Organization*, 157:680–707.

Gerardi, K. S. and Shapiro, A. H. (2009). Does competition reduce price dispersion? New evidence from the airline industry. *Journal of Political Economy*, 117(1):1–37.

Goolsbee, A. and Syverson, C. (2008). How do incumbents respond to the threat of entry? Evidence from the major airlines. *The Quarterly Journal of Economics*, 123(4):1611–1633.

He, L., Kim, M., and Liu, Q. (2022). Competitive response to unbundled services: An empirical look at spirit airlines. *Journal of Economics & Management Strategy*, 31(1):115–145.

Kim, B.-D., Shi, M., and Srinivasan, K. (2001). Reward programs and tacit collusion. *Marketing Science*, 20(2):99–120.

Kim, D., Kim, M., and Tan, K. (2021). Tacit collusion and price dispersion in the presence of southwest airlines. *Southern Economic Journal*, 88(1):3–32.

Kim, M., Shen, L., and Ge, Q. (2023). Does competition increase or decrease price dispersion? insights from one-way vs. round-trip airfares. *Oxford Bulletin of Economics and Statistics*, 85(2):435–455.

Klemperer, P. (1995). Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade. *The Review of Economic Studies*, 62(4):515–539.

Kwoka, J., Hearle, K., and Alepin, P. (2016). From the fringe to the forefront: Low cost carriers and airline price determination. *Review of Industrial Organization*, 48(3):247–268.

Lederman, M. (2007). Do enhancements to loyalty programs affect demand? The impact of international frequent flyer partnerships on domestic airline demand. *The RAND Journal of Economics*, 38(4):1134–1158.

Lederman, M. (2008). Are frequent-flyer programs a cause of the "hub premium"? *Journal of Economics & Management Strategy*, 17(1):35–66.

Lee, D. and Luengo-Prado, M. J. (2005). The impact of passenger mix on reported "hub premiums" in the U.S. airline industry. *Southern Economic Journal*, pages 372–394.

Li, S., Mazur, J., Park, Y., Roberts, J., Sweeting, A., and Zhang, J. (2022). Repositioning and market power after airline mergers. *The RAND Journal of Economics*, 53(1):166–199.

Luo, D. (2014). The price effects of the Delta/Northwest airline merger. *Review of Industrial Organization*, 44(1):27–48.

Luttmann, A. (2019). Evidence of directional price discrimination in the U.S. airline industry. *International Journal of Industrial Organization*, 62:291–329.

Luttmann, A. and Nehiba, C. (2020). The effects of employee hours-of-service regulations on the U.S. airline industry. *Journal of Policy Analysis and Management*, 39(4):1043–1075.

Morrison, S. A. (2001). Actual, adjacent, and potential competition: Estimating the full effect of Southwest Airlines. *Journal of Transport Economics and Policy*, 35(2):239–256.

Orlov, E. (2011). How does the internet influence price dispersion? Evidence from the airline industry. *The Journal of Industrial Economics*, 59(1):21–37.

Shen, Y. (2017). Market competition and market price: Evidence from United/Continental airline merger. *Economics of Transportation*, 10:1–7.

Snider, C. and Williams, J. W. (2015). Barriers to entry in the airline industry: A multidimensional regression-discontinuity analysis of AIR-21. *Review of Economics and Statistics*, 97(5):1002–1022.

Tan, K. M. (2016). Incumbent response to entry by low-cost carriers in the u.s. airline industry. *Southern Economic Journal*, 82(3):874–892.

Whalen, W. T. (2007). A panel data analysis of code-sharing, antitrust immunity, and open skies treaties in international aviation markets. *Review of Industrial Organization*, 30(1):39–61.

APPENDIX A: Number of Frequent Flyer Awards

Table A1: Number of Frequent Flyer Awards Redeemed
(millions of tickets)

Airline	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Alaska (AS)	0.8	0.9	1.1	1.4	1.4	1.7	*	*	*	*	*	*	*	*	*
American (AA)	5.2	5.2	5.2	6.2	5.2	5.6	6.0	6.0	6.1	7.9	8.3	10.0	11.0	13.0	14.0
Continental (CO)	1.4	1.5	1.5	1.6	1.3	1.6	1.9	Acquired by United (UA)							
Delta (DL)	3.3	*	*	*	*	12.0	12.0	11.0	11.0	12.5	13.3	13.4	14.9	17.2	20.0
Hawaiian (HA)	*	*	*	0.5	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.6	0.6	0.7	0.7
JetBlue (B6)	0.1	0.2	0.2	0.3	0.3	0.6	0.6	0.8	0.9	1.1	1.4	1.6	2.0	2.0	3.0
Northwest (NW)	1.5	1.5	*	*	Acquired by Delta (DL)										
Southwest (WN)	2.6	2.7	2.8	2.8	3.0	3.2	3.7	4.5	5.4	6.2	7.3	8.3	9.6	10.4	10.7
United (UA)	2.2	2.3	2.2	2.3	2.1	2.4	2.5	4.7	5.0	4.8	5.0	5.2	5.4	5.6	6.1
U.S. Airways (US)	1.3	0.5	0.9	0.9	0.8	0.8	0.8	0.8	1.8	Acquired by American (AA)					

Source: 2005-2019 Form 10-K filings for Alaska, American, Continental, Delta, Hawaiian, JetBlue, Northwest, Southwest, United, and U.S. Airways.

Notes: *Number not reported in Form 10-K filing. Numbers are reported in millions. AirTran (FL), Frontier (F9), Spirit (NK), and Virgin America (VX) are not included because they do not report the number of frequent flyer awards in their annual Form 10-K filings. Allegiant (G4) introduced their frequent flyer program in August 2023.

APPENDIX B: Continental and United

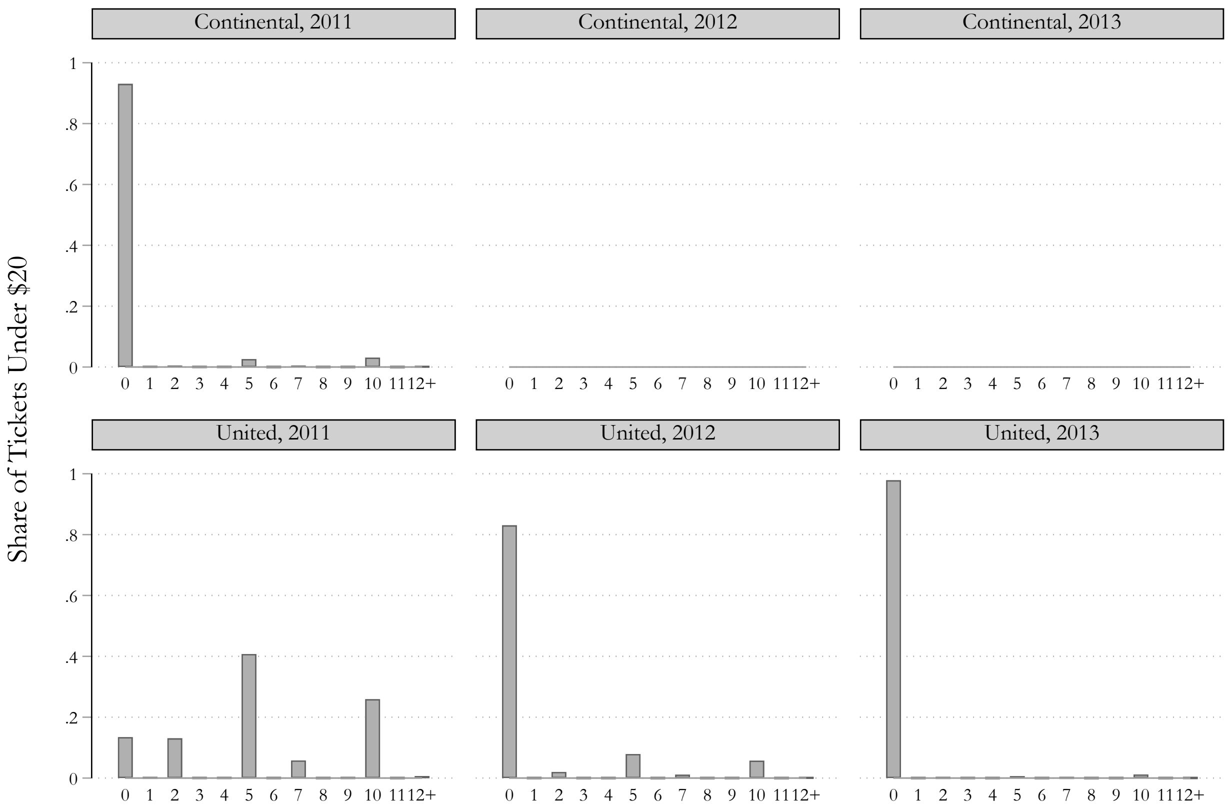
As mentioned in Table 19, Continental appears to report frequent flyer award (FFA) tickets in the Airline Origin and Destination Survey (DB1B) with an itinerary fare of zero. United also appears to report FFA tickets with a fare of zero after their merger with Continental (2012-2019). To illustrate this phenomenon, Figure B1 plots the distribution of DB1B fares under $20 for Continental and United in 2011, 2012, and 2013.

Consistent with our FFA identification strategy described in Section 3, spikes in the distribution of United's fares occur at $2, $5, $7, and $10 in 2011. However, these spikes largely disappear in the 2012 and 2013 plots after Continental's reservation system and frequent flyer program were merged into United's on March 3, 2012.[48] Additionally, the distribution of Continental's fares in 2011 has a large spike at $0 but no spikes at $2, $5, $7, and $10. These plots suggest that Continental reports FFA tickets in the DB1B with an itinerary fare of zero. Furthermore, researchers should be aware that all Continental tickets in addition to United tickets after March 3, 2012 are likely not inclusive of ticket taxes and fees since the Passenger Fee does not appear to be reflected in their DB1B observations. Accordingly, researchers should consider adding the applicable Passenger Fee to all reported Continental fares and United fares after March 3, 2012 or include the appropriate level of fixed effects (e.g., airline-time and airline-route) in their empirical analyses.

Finally, we do not classify Continental or United tickets with an itinerary fare of $0 as FFAs in the analysis presented in Sections 4, 5, and 6. However, the results presented in this paper do not materially change if all Continental tickets with a fare of $0 and all United tickets after March 3, 2012 with a fare of $0 are classified as FFAs (results from these robustness checks are available upon request).

[48] Although the United-Continental merger closed on October 1, 2011, the Federal Aviation Administration did not grant a single operating certificate to United until November 30, 2011. Continental's reservation system and OnePass Frequent Flyer miles program were officially merged into United's on March 3, 2012.

Figure B1: Distribution of DB1B Fares Under $20 for Continental and United in 2011, 2012, and 2013



Notes: Data from DB1B files for 2011, 2012, and 2013. Bars represent the share of passengers with tickets under $20 that reported that exact itinerary fare. Data are limited to one-way itineraries with three or fewer segments and round trip itineraries with six or fewer segments.

Help Really Wanted?

The Impact of Age Stereotypes in Job Ads on Applications from Older Workers*

Ian Burn
University of Liverpool

Daniel Firoozi
Claremont McKenna College

Daniel Ladd
University of California-Irvine

David Neumark
University of California-Irvine, NBER, and IZA

June 14, 2024

*This research was funded by the Sloan Foundation. We received helpful comments from Joanna Lahey and from seminar participants at Bocconi University, Cardiff University, University of Edinburgh, Hebrew University, The Hoover Institution, Iowa State University, University of Kentucky, Reichman University, Seoul National University, University of Bolzano, University College London, University of Sheffield, University of St. Andrews, Texas A&M, University of Tokyo, Uppsala University, Yonsei University, and the 2022 NBER Conference on the Labor Market for Older Workers, Spring 2023 NBER Labor Studies Meeting, and the 2023 CESifo Area Conference on Labor Economics . Any views or opinions expressed are solely those of the authors. The Pre-Analysis Plan for this project was registered on Open Science Framework on December 31, 2020. The research was approved by the UCI Office of Research Institutional Review Board on October 18, 2019: HS# 2015-2017, modification application #26404. The pre-analysis plan is available at https://osf.io/37yaq/.

Abstract

Correspondence studies have found evidence of age discrimination in callback rates for older workers, but less is known about whether job advertisements can themselves shape the age composition of the applicant pool. We construct job ads for administrative assistant, retail, and security guard jobs, using language from real job ads collected in a prior large-scale correspondence study (Neumark et al., 2019a). We modify the job-ad language to randomly vary whether the job ad includes ageist language regarding age-related stereotypes. Our main analysis relies on computational linguistics/machine learning methods to design job ads based on the semantic similarity between phrases in job ads and age-related stereotypes. In contrast to a correspondence study in which job searchers are artificial and researchers study the responses of real employers, in our research the job ads are artificial and we study the responses of real job searchers. We find that job-ad language related to ageist stereotypes, even when the language is not blatantly or specifically age-related, deters older workers from applying for jobs. The change in the age distribution of applicants is large, with significant declines in the average and median age, the 75th percentile of the age distribution, and the share of applicants over 40. Based on these estimates and those from the correspondence study, and the fact that we use real-world ageist job-ad language, we conclude that job-ad language that deters older workers from applying for jobs can have roughly as large an impact on hiring of older workers as direct age discrimination in hiring.

1 Introduction

Lengthening work lives for those able to work is a crucial part of the policy response to population aging. Because many seniors transition to "partial retirement" or "bridge jobs" at the end of their careers (Cahill et al., 2006; Johnson et al., 2009) or return to work after a period of retirement (Maestas, 2010), reducing age discrimination in hiring is critical to lengthening working lives. There is an extensive body of research that documents the extent to which employers discriminate against older workers in hiring, using correspondence studies (e.g., Bendick et al., 1997, 1999; Lahey, 2008; Farber et al., 2019; Neumark et al., 2019a, 2019b). This research focuses on measuring employer behavior – specifically, whether there is less hiring of qualified older workers – and generally finds evidence consistent with hiring discrimination against older workers. There is little work, however, that studies how workers respond to manifestations of age discrimination in the labor market, including steps employers may take to discourage older workers from applying for jobs.

In this study, we create job ads for administrative assistant, retail sales, and security guard jobs.[1] We construct these job ads using language from real job ads collected in Neumark et al. (2019a). We post these job ads while randomly varying whether the text includes language that is semantically similar to ageist stereotypes (based on a common computational linguistics metric), and that older workers perceive as age biased. We focus on stereotypes related to communication skills, physical ability, and technology skills. The innovation in this study is that the job ads are artificial, and we are studying the responses of real job searchers – in contrast to a correspondence study in which the job searchers are artificial and researchers study the responses of real employers.[2]

[1] These jobs have relatively high hiring of older workers.

[2] We obtained IRB approval to post these ads, subject to an IRB-approved protocol to do two things: (i) to quickly inform applicants that the job is not available, so as not to interrupt their job search; and (ii) subsequently, to inform those from whom we have collected data of their inclusion in an experiment (and explain why we could not use informed consent), and allow them to opt out of their data being used. (This is the standard protocol in experiments involving real job searchers; see Krause et al., 2012.) This protocol recognized the greater potential impact on job seekers than on employers in a correspondence study who receive fake applications. In our view, the critical policy issue of lengthening work lives in response to population aging poses substantial potential benefits that outweigh this impact.

The potential for age stereotypes or other language in job ads to deter applications from older job seekers is real. An extreme example is stating maximum experience levels in job ads. This occurred recently in *Kleber v. Carefusion Corp.*, where the job ad requested "3 to 7 years (no more than 7 years) of relevant legal experience," language that will clearly act to exclude many older applicants.[3] More generally, the U.S. Code of Federal Regulations covering the ADEA currently states, "Help wanted notices or advertisements may not contain terms and phrases that limit or deter the employment of older individuals. Notices or advertisements that contain terms such as age 25 to 35, young, college student, recent college graduate, boy, girl, or others of a similar nature violate the Act unless one of the statutory exceptions applies" (§1625.4).[4] Beyond that, organizations like AARP suggest that "[d]espite protections by the Age Discrimination in Employment Act of 1967 (ADEA), employers have gotten cleverer in masking what is age discrimination" by using ageist phrases in job ads (Brenoff, 2019). In our view, we add significant new evidence to the research literature on discrimination. We advance this literature with a large-scale field experiment to test the impact of exogenous inclusion of job ad language that has been previously identified by the industrial psychology literature, machine learning/computational linguistics methods, the AARP, and a panel of older workers to subtly signal employer ageism. In so doing, we provide evidence on the effect of discrimination on the behavior of workers, whereas the research literature on discrimination generally focuses on the behavior of *employers*.

The correspondence study in Neumark et al. (2019a) showed that employers discriminate against older workers.[5]Subsequently, Burn et al. (2022) found that the discriminating employers

[3]See *Kleber v. Carefusion Corp.* (http://www.aarp.org/content/dam/aarp/aarp_foundation/litigation/pdf-beg-02-01-2016/kleber-amended-complaint.pdf, viewed November 8, 2017). Surprisingly, the court ruled in favor of the defense in this case, reaching a new interpretation that the ADEA does not authorize job applicants to bring a disparate impact claim. (See Button, 2019.)

[4]As described in Pillar et al. (2022), in the Netherlands, the Dutch Equal Treatment Act regulates both explicit mention of age (e.g., "younger than 30 years") and formulations that imply age (e.g., specifically recruiting students). In contrast to the work in Burn et al. (2022), discussed below, Pillar et al. study detection in job ads of discriminatory statements that are explicitly defined and prohibited by the law.

[5]This was a very large-scale correspondence study of age discrimination, involving sending over 40,000 job applications, in triplets of applications differentiated by age, in response to over 13,000 job postings. That study was designed to test for numerous potential sources of bias in estimates of hiring discrimination in prior age discrimination correspondence studies, but still uncovered robust evidence of hiring discrimination against older women and slightly more variable but still strong evidence of hiring discrimination against older men.

in this correspondence study used language in job ads that was as identified as ageist by machine learning/computational linguistics, and was perceived by older workers as ageist against applicants aged 50 or older (Burn et al., 2023). The present paper builds on this work, addressing the question of whether job seekers respond to ageist language in job ads – which may explain the evidence that employers who discriminate against older workers in hiring use ageist stereotypes in job ads.

We find strong evidence that ageist language related to communication skills, physical ability, and technology skills, even when it is not blatantly or specifically age-related, discourages older workers from applying for jobs. Job ads that feature ageist language deter a large fraction of older applicants compared to ads that do not feature ageist language, shifting the age distribution of the applicant pool downward. For example, ads containing a machine-learning generated phrase related to each of these three stereotypes attract job applicants that are 2.5 years younger on average, and more ageist phrases we examine have even stronger impacts. To provide a bit more context, the treatment phrases we use fall at roughly the 75th percentile of "bias" (as measured by the distribution of cosine similarity scores of phrases from the job ads collected in Burn et al., 2022) with the stereotypes we study, while the control phrases are close to the median. Thus, the control phrases should not be regarded as ageist by the average job seeker reading the text, while the treatment phrases should be viewed as reflecting relatively subtle or indirect ageist language, rather than blatant language conveying disinterest in hiring older workers. When we define our treatment in terms of either the cosine similarity score for our treatment phrases, or the Likert score for the perceived ageism from our MTURK panel, we find that a 1 standard deviation increase in these measures reduces the proportion of job applicants over age 40 by about 8 percentage points, and lowers their average age by about 2 years.

We verify that this decline in the share of older workers in the applicant pool for treatment group ads is driven by a decline in the raw number of older applicants, rather than an increase in younger applicants. Moreover, the response of older applicants is not driven by the possibility that the treatment ads refer to higher skills or requirements that older workers are less likely to have (or reflect jobs they are less likely to want). Specifically, based on job application documents (mainly resumes and cover letters), we find that there is little or no age gradient in the skills or qualifications of applicants that are related to the age stereotypes we manipulate. And we find that treatment by ageist job-ad language reduces the relative number of older workers applying for our job postings, but does not increase the average skills or qualifications of older applicants. In other words, older workers select out of applying for jobs when exposed to treatment phrases in a way that is uncorrelated with either their previous occupational experience or their skills or qualifications related to the phrases in the treatment job ads. These findings reinforce the interpretation of our experimental results as showing that older workers respond to subtle shifts in the language of job ads that signal that an employer holds ageist stereotypes about older workers or is otherwise less interested in hiring older workers.

As discussed in more detail below, the simplest interpretation of the evidence is that discriminatory employers who do not want to hire older workers use ageist language in job ads to discourage them from applying, because this makes it harder to detect age discrimination in data on hiring relative to applicants. A more subtle interpretation is that some jobs have requirements, stated in ads, that older employers may on average be less likely to meet, that employers statistically discriminate in assuming that older applicants cannot fulfill these requirements, and that older workers know this and hence are less likely to apply. The treatment job-ad language does indicate somewhat higher skill requirements. But these are not requirements that are much less likely, if at all, to be met by the older job seekers for whom we see reductions in job applications (which begin around the age of 40). Moreover, the language is perceived by potential job seekers as age biased in our MTURK panel (Burn et al., 2023), and similar language is used by discriminatory employers (as measured in the correspondence study). We do not insist that there are no skill differences between younger and older workers that could, via the second hypothesis, discourage some older applicants. But given the skills we use, the age at which they set in, and the large reductions in applications from older workers, we view it as implausible that skill differences could drive our evidence. Moreover, as just noted, we find no evidence that the treatment ads lead to older applicants selecting out based on the skill or job requirements stated. As a result, we interpret our evidence as pointing to ageist-language in job ads deterring older workers from applying for jobs, and by far the most plausible explanation is that this reflects employers using such language to discourage older workers from applying for jobs – i.e., intentional discrimination.

This discouragement effect in response to ageist stereotypes in job ads illustrates a more subtle form of age discrimination in the labor market.[6] Age discrimination that deters older workers from applying for jobs has the same effect as direct age discrimination applied to job applicants; both reduce the employment of older workers.[7] Strikingly, our evidence from this new experiment, combined with prior evidence on age discrimination in hiring from the correspondence study (Neumark et al., 2019a), suggests that the effect of direct hiring discrimination may be only slightly larger than the effect of discouraging older workers from applying for jobs – although the evidence from both experiments is specific to the experimental conditions, and may not generalize to the actual incidence of age discrimination in hiring and the effects of age-related stereotypes in job ads in the broader labor market.

Our evidence has significant policy implications regarding age discrimination and its enforcement. Utilizing ageist language in job ads may be rational for employers, despite it being illegal to discriminate against older workers. Shaping the applicant pool by discouraging older applicants has a potential benefit for discriminatory employers, because of the incentives created by age discrimination laws. In particular, a lower representation of older workers in their applicant pool can justify a lower representation of older workers among employees, making it easier to rebut an allegation of age discrimination in hiring. More generally, employers who do not want to hire older workers might, in order to avoid unnecessary search costs, discourage older workers from applying by signaling their ageism. To think about this another way, in the legal system, hiring discrimination cases based on age (or, similarly, membership in other groups) typically hinge on shortfalls

[6] Our evidence should be viewed as another dimension of age discrimination in hiring – one that has not been studied or detected in the research literature that tests for hiring discrimination based on age, mainly using correspondence studies. These include Baert et al. (2016); Bendick et al. (1997, 1999); Carlsson and Eriksson (2019); Farber et al. (2017, 2019); Lahey (2008); Neumark et al. (2016, 2019a, 2019b); and Riach and Rich (2006, 2010).

[7] Yet another way to "discourage" older workers from applying for jobs is to target job ads to younger job seekers, as discussed in Ajunwa (2019), who also discusses a class action complaint against Facebook and other companies for such targeting. Moreover, the complaint alleged that Facebook used similarly discriminatory age filters in targeting its own employment ads.

of older workers among hires relative to the applicant pool. But if job-ad language deters older workers from applying, these shortfalls may be obscured, and the courts may need to weight other evidence more heavily – including both job-ad language as the source of lower applications from older workers, and comparisons with other benchmarks to assess whether hiring of older workers is notably lower at the firm in question.

To address age discrimination from stereotyped job-ad language that discourages older workers from applying for jobs, there are two tools the U.S. Equal Employment Opportunity Commission (EEOC) – or other anti-discrimination authorities – could utilize.[8]First, it could issue stronger guidance to employers on language to avoid that might be interpreted as discouraging older workers from applying. There exists a duty of care for employers to knowingly avoid using language which may deter older workers from applying. Therefore, our evidence provides a basis for further guidance regarding the usage of ageist stereotypes in job ads that may influence job application decisions. Second, the EEOC might consider flagging for potential investigation firms that use age-stereotyped language in their job ads, recognizing that, for these firms, discrimination may be occurring even in the absence of shortfalls between the share of older applicants hired and the share of older workers who apply for jobs. Thus, rigorous evidence on the role of ageist language in job ads could potentially influence policy to reduce age discrimination in hiring and contribute to lengthening work lives. Obviously, similar considerations could apply to other groups protected by discrimination laws.

2 Previous Related Literature

Very few studies in labor economics explore job ads, and fewer still focus on discrimination. Among studies of issues other than discrimination, Modestino et al. (2016) use text data from job ads to document that "downskilling" occurred during the recovery from the Great Recession,

[8]European Union law also bars age discrimination. To the best of our knowledge, it is less explicit about the forms of discrimination barred, and it also differs in not protecting older workers per se, but rather barring discrimination based on age generally. See Lahey (2010) and European Commission (2000).

with firms reducing skill requirements in their job ads. Deming and Kahn (2018) use text data in job ads to measure how ten different skills relate to wages. Marinescu and Wolthoff (2020) match text data from job ads to job application data to study the matching process between jobs and applicants. Turning to research more related to labor market discrimination, Kuhn and Shen (2013) and Kuhn et al. (2020) explore how gender preferences feature explicitly or implicitly in job ads in China, Chaturvedi et al. (2021) examine gender preferences in job ads in India, Card et al. (2021) study gender preferences in job ads in Austria, and Hellester et al. (2020) study age and gender preferences in job ads in China and Mexico.

A small number of studies are closer in spirit to ours in that they run experiments manipulating job ads and study responses of job seekers. Among these, He et al. (2021) study how job flexibility conditions influence job application behavior, and Belot et al. (2021) study the effects of experimental variation in posted wages. Flory et al. (2015) study job seeker responses to ads for jobs that differ regarding competition and uncertainty about pay. And closer to our work, Flory et al. (2019) examine how signaling interest in employee diversity affects application behavior of minority and female candidates (as well as firm selection) – although this should probably be viewed as trying to encourage job applications from a particular group, the opposite of the behavior we study.[9]

Two studies connect the text of job ads to measured discriminatory behavior of employers.[10] Tilcsik (2011) identifies words in job ads related to masculine stereotypes (decisive, aggressive, assertive, and ambitious) and links those to hiring outcomes in a correspondence study of discrimination against gay men.[11] And, in the most systematic approach, Burn et al. (2022) identify

[9]There is other research suggesting that, in other contexts, job seekers respond to job-ad language, including: Banfi et al. (2019) on posted wages; Hellester et al. (2020) on gender requests; and Ibanez and Riener (2018), Leibbrandt and List (2018), and Flory et al. (2021) on affirmative action or diversity statements in recruitment materials or job ads. A key difference between our method and the previous work is the subtler nature of the language studied. Previous work manipulates a signal of likely callback directly, as in the case of affirmative action or diversity statements, whereas we manipulate signals of likely callback indirectly through subtle shifts in the language of job requirements.

[10]Though they did not focus on job ads, Hanson et al. (2011, 2016) study language used by mortgage originators and connect this language to their behavior. Hanson et al. (2011) study subtle discrimination through "keywords" used by landlords responding to prospective tenants. Hanson et al. (2016) had research assistants subjectively (and blindly) code the helpfulness and other characteristics of mortgage loan originator responses to prospective borrowers.

[11]In an early small study, Wax (1948) found that summer resorts in Ontario, Canada, were more likely to discriminate against Jewish customers (based on names) requesting accommodations if they used phrases like "restrictive clientele" in their advertising.

common age stereotypes from the research literature in industrial psychology, use machine learning to calculate the relationship between the text of the job ads and specific age stereotypes, and then test whether job-ad language related to the stereotypes predicts hiring discrimination against older workers in a correspondence study. As already noted, the present paper builds on this prior study.

Finally, two papers study the manipulation of job ads and applicant responses. Card et al. (2021) study a quasi-experiment generated from a policy change in Austria in which the government launched a campaign to inform employers and newspapers that gender preferences in jobs ads were illegal, and find that this reduced expressed gender preferences in job ads and led to increases in women (men) hired for jobs more likely to be targeted to men (women). Delfino (2021) manipulates photos of workers in job ads to convey information on the gender composition of the workforce and finds that showing a male photo decreases the likelihood that women apply, but has no effect on whether men apply.[12]

There has been no research on how ageist language in job ads affects the decisions of older workers to apply. What is known about how job applicants read job ads for bias focuses exclusively on gender bias in job ads. Gaucher et al. (2011) found that job ads for male-dominated occupations used masculine wording (words associated with male stereotypes, such as leader, competitive, dominant) more frequently than advertisements for female-dominated occupations, and women find job advertisements less appealing when they contain more masculine than feminine wording (Bem and Bem, 1973; Gaucher et al., 2011).[13] However, these findings are based on laboratory experiments that ask how subjects perceive job-ad language, whereas our research uses a field experiment that studies the behavior of actual job seekers.

[12] Kuhn and Shen (2021) study the impact of removing preferred gender fields from ads on a Chinese job board on applicant responses. The removal of the preferred gender field was not a manipulation by any employer (the field was removed from the entire job site), but they find that "gender-mismatched" applications and callbacks increase.

[13] Chaturvedi et al. (2021) examine how words that predict the employer having a gender preference are correlated with job applications, but they do not examine the link between the words and stereotypes about men and women.

3 Conceptual Framework, Interpretation, and Model

Why might employers use stereotyped language in job ads? One hypothesis – perhaps the central one – is that employers who discriminate based on age use stereotyped language to try to shape the applicant pool. Using language that conveys positive stereotypes related to young workers might discourage older workers from applying (as might language conveying negative stereotypes related to older workers – although this is less common in our data). Employers may introduce this language via job requirements that are correlated with age, appear natural to use in job ads, and are not so blatant as to make the age discrimination clear.

This discouragement from applying would lead to the underrepresentation of older applicants in the applicant pool, and is potentially valuable to a discriminating employer because the probability of a hiring age discrimination claim and of an adverse outcome for the employer is smaller when, ceteris paribus, the ratio of older applicants to younger applicants is lower.[14] Employers could use job-ad language this way whether their discrimination is taste-based or statistical, and, in the case of statistical discrimination, whether or not the language is related to the assumptions they make about older workers (e.g., they might assume older workers will leave the firm sooner). In any of these cases, employers might use ageist language in job ads to deter older workers from applying.

A second hypothesis, which is more complex, is also related to statistical discrimination. Different jobs may have different requirements, which could be stated in job ads without any explicit intent of discouraging older applicants. But employers may hold stereotypes about older job applicants' abilities to meet these job requirements – for example, assuming that older workers are less likely to be able to do the heavy lifting that a job requires, which may well be true on average but

[14] In legal cases, the most compelling data on hiring discrimination comes from comparing hiring rates of the group in question (e.g., older workers) relative to the applicant pool. Hiring charges under the U.S. Age Discrimination in Employment Act (ADEA) made up nearly 5% of total ADEA charges in 2020 – more than double the percentage under Title VII (protecting women, minorities, etc.) or the Americans with Disabilities Act. (This is based on authors' computations using EEOC statistics available at https://www.eeoc.gov/statistics/statutes-issue-charges-filed-eeoc-fy-2010-fy-2020, viewed January 18, 2022.) The representation of hires among applicants is important in anti-discrimination enforcement, as the EEOC uses a "4/5ths" rule (the ratio of the selection rate for the group in question to the group with the highest selection rate) as "a practical means of keeping the attention of the enforcement agencies on serious discrepancies in rates of hiring, promotion and other selection decisions" (U.S. Equal Employment Opportunity Commission, 1979).

of course not of each applicant.[15] Employers may act on these assumptions, and older job seekers, expecting this, may be deterred from applying.

While social scientists are interested in the nature of discriminatory behavior, both statistical and taste discrimination are illegal under U.S. law. Not surprisingly, language in job ads that refers to age either explicitly or "mechanically" (e.g., referring to recent graduates) is illegal in the United States. The legality of less blatant job-ad language with job requirements that reflect age stereotypes and is associated with lower hiring of older workers is more complex. On the one hand, EEOC regulations state: "An employer may not base hiring decisions on stereotypes and assumptions about a person's race, color, religion, sex (including pregnancy), national origin, age (40 or older), disability or genetic information." (See U.S. Equal Employment Opportunity Commission, n.d.(a).) On the other hand, job requirements that are based on factors related to age are not necessarily illegal. The legality of job requirements related to age generally requires an employer to show that the use of these requirements is based on a reasonable factor other than age (RFOA), even if that factor is correlated with age. An RFOA is defined as "a non-age factor that is objectively reasonable when viewed from the position of a prudent employer mindful of its responsibilities under the ADEA under like circumstances." (See Federal Register, n.d.) In other words, the law recognizes that characteristics of workers that are related to age can sometimes be legitimate for employers to consider. Indeed, the law goes further, as in rare cases employers can use age as an explicit criterion if a requirement for the job is strongly related to age but hard to assess independently. This exception requires that age is a "bona fide occupational qualification" (BFOQ) that is "reasonably necessary to the normal operation of the business." (U.S. Equal Employment Opportunity Commission, n.d.(b)).[16]

[15] It is also possible, in principle, that employers use this language randomly and unintentionally, but it still deters older workers from applying. However, the evidence in Burn et al. (2022), showing that discriminatory employers used ageist stereotypes in job ads, ruled this out.

[16] A key example is Hodgson v. Greyhound Lines, Inc., where the company was sued for having a maximum hiring age. Greyhound prevailed by establishing that driving ability is essential to passenger safety, that older hires would be less safe drivers (because achieving maximum safety took 16-20 years of experience), that some abilities associated with safe driving deteriorate with age, and that these changes are not detectable by physical examination (which could otherwise be a substitute for an age criterion). (See U.S. Court of Appeals, 7th Circuit, 1974.) As discussed by Combs (1982), the issue of the rights of older workers vs. public safety have figured prominently in court decisions regarding age as a BFOQ under the ADEA.

Our evidence cannot speak directly to the question of taste vs. statistical discrimination or whether the stated job requirements would be viewed as legal. Indeed, we do not study employer behavior in our experiment, although we do use job-ad language from real employers. What our evidence does address is whether age stereotypes expressed in job ads affect the likelihood that older job seekers apply for jobs, likely by signaling to job applicants that older workers are less likely to be hired. A response could mean either that the language is perceived as directly reflecting age bias – aversion to hiring older workers – or that the language is perceived as "biased" because it puts older workers at a disadvantage because they may be less likely to satisfy the stated job requirement, or be perceived as such by employers. Thus, our evidence can reveal the potential for employers to use job-ad language to discriminate against older workers in hiring, and the potential adverse impact on older job applicants.

A simple model can describe the behavior of job seekers that we use to interpret our evidence. When deciding whether to apply to a job, potential workers read the job ad and decide whether the potential benefit outweighs the costs of applying. Suppose the utility of job j to person i is $U_{ij} = \varepsilon_{\mathrm{ij}}$ where $\varepsilon_{\mathrm{ij}} \sim N(0,1)$.[17] S indexes how age-stereotyped the job ad is. The cost of applying for a job is c. The potential benefit of a job is determined by posted wage and the probability of getting a job offer (callback).[18] For younger workers this is $p_{\mathrm{y}} = b(0 < b \leq 1)$ if young, and $p_{\mathrm{o}} = b(S)(0 \leq b(S) \leq b, b'(S) < 0)$ if old. That is, the probability of a job offer for an older worker is a decreasing function of how age-stereotyped (S) the job ad is. An example of a function meeting these conditions is:

$$(1)\ b(S) = b * e^{-\eta * S}.$$

A young person applies if $b * e > c$ (dropping the i and j subscripts) or $e < -c/b$. Given the distributional assumption, the probability of applying is $A_{\mathrm{y}} = \Phi(-c/b)$ where Φ denotes the

[17] The variance can be fixed without loss of generality.

[18] It is a simplification to focus only on the wage, of course, but these are generally low-skilled jobs that might have few opportunities for training, advancement, etc. Correspondingly, our job ads do not reference these other dimensions of jobs. Regardless, the intuition is the same if one interprets the posted wage instead as a longer-term earnings measure of the job.

standard normal distribution function. An old person applies if $b(S) * e > c$ (dropping the i and j subscripts) or $e < -c/b(S)$., so the probability of applying is $A_o = \Phi(-c/b(S))$.[19]

In this paper, we are interested in estimating $\partial A_o/\partial S$. For old applicants,

$$(2)\ \partial A_o/\partial S = \frac{-(-e)b'(S)}{b(S)^2} * \phi(\frac{-e}{b(S)}),$$

$b'(s) < 0$ implies that $\partial A_o/\partial S < 0$, thus predicting a negative response of older applicants to job-ad language with more ageist stereotypes.

4 Methods

To test whether older workers respond to ageist language in job ads (i.e., is $\partial A_o/\partial S < 0$?), we conduct an experiment where we manipulate S and observe how the applicant pool changes. In our experiment, we post job ads in three occupations in 15 U.S. cities, randomly varying the inclusion of age-related stereotypes in the text of the job ad. The job ads are artificial, and we study the responses of real job searchers. This allows us to test for differences in the applicant pool when otherwise similar ads use age-related stereotypes vs. age-neutral language.

Selecting the Cities and Occupations

We build on the experiment conducted in Neumark et al. (2019a). The 12 cities in that study were selected due to their large size, their geographic distribution across the U.S., and because they have different population age distributions. For this experiment, we added three more cities with a large online presence for the job board we use. For each city, we post our job ad on the online job board, setting the hiring firms' locations to the central business district. The cities in the experiment are shown in Figure 14 (which also provides additional information on the data collection in the experiment). We use three of the four occupations from the original study: retail sales (mixed-gender), administrative assistant (female-dominated), and security guards (male-dominated). These occupations are lower skilled, with jobs often advertised using online job

[19] We assume that $\partial A_y/\partial S = 0$, or young people do not respond to the stereotyped language. We could have the probability of an offer for a young applicant *increasing* in S – i.e., the opposite direction from old people – and the qualitative conclusion is the same.

boards. Correspondence studies typically focus on lower-skilled jobs for a number of reasons: (i) the need for a source of a large number of job ads; (ii) a job search process where submission of a simple resume results in callbacks for job interviews; and (iii) the study design requires a labor market where job applicants would not potentially be known to employers. Nonetheless, these jobs are also fairly important for hiring of older workers. As shown in Neumark et al. (2019a), all three occupations were in the top decile of jobs in terms of the proportions of older people hired.[20][21]

Selecting the Stereotypes

To select the stereotypes we use in our experiment, we start with a list of ageist stereotypes from the industrial psychology literature (see Burn et al., 2022). These are listed in Table 23. Among these, we selected stereotypes that met the following criteria. First, the stereotype is commonly expressed in job-ad language about the ideal or preferred candidate skills or attributes; we did not want to focus on stereotypes that are not often included in job ads (e.g., hearing and memory), even if employers hold these stereotypes based on the industrial psychology literature. Second, we focus on stereotypes for which we had evidence of a correlation between discrimination and the stereotype (from Burn et al., 2022).[22] Third, older workers should be aware that employers held that stereotype. As evidence, we drew on various reports put out by AARP; see Brenoff (2019) and Terrell (2019). Our final list of stereotypes is three skills or abilities for which older workers are stereotyped as deficient: communication skills, physical ability, and technological skills.

Designing the Job Ads

We used one job ad template per city-occupation combination, basing the structure and language of the template on actual ads collected in Neumark et al. (2019a) and recent real ads posted on job boards in the sample cities to capture contemporaneous patterns in their language. We selected a handful of ads to use as our base to create a template and copied their format (location

[20] As reported in Neumark et al. (2019a), looking at the distribution of the share of 62-70 year-olds hired recently (tenure less than five years) across all occupations, the percentiles for males in the occupations we use were 96.6 for retail salespersons and 93.3 for security guards and gaming surveillance officers. The percentiles for females were 100 for secretaries and administrative assistants, 96.4 for receptionists and information clerks, and 95.2 for retail salespersons.

[21] We omit the janitor jobs also included in Neumark et al. (2019a), because for them the evidence of age discrimination was less clear-cut, and there are many fewer janitor job advertisements posted online.

[22] We did not require this evidence for all three occupations or for both genders, but just for some subset.

of blocks of text, types of bullet points, and style of text) to ensure our templates were similar in appearance to others on the website. We stripped the ad of all identifying information, so there is no identifiable link between the ad posted and the template we created. The text of each ad was rewritten to give enough details about the company and the position to appear realistic, but not enough details to suggest a specific company.[23] We modified the requirements of the jobs to reduce the number of applicants they potentially exclude. All ads were written to have flexible hours, competitive pay, and the availability of part-time and full-time positions (at the employee's choice). For half of the templates, we included the requirement that applicants must have a high school diploma (randomized by template). Figure 15 provides an example of a job ad for each of our occupations.

The treatment and control ads differ in the job requirements (denoted in bold with asterisks in each template in Figure 15), with three sentences assigned to be either a treatment phrase (stereotyped) or a control phrase (not stereotyped). The requirements we manipulate have to do with a candidate's communication skills, physical ability, and technology skills. Our control phrases express job requirements that are also appropriate for the job but use age-neutral language not related to these age-stereotyped skills or abilities, and as much as possible refer to related skills, while our treatment phrases use language highly related to these ageist stereotypes.

Creating Stereotyped Job Requirements

We use two methods to generate sentences highly related to ageist stereotypes, focusing on constructing sentences that were highly related to only one of the three stereotypes we use. The first uses measures of semantic similarity generated by computational linguistics/machine learning methods. Drawing on Burn et al. (2022), we calculate the semantic (cosine) similarity of thousands of phrases to communication skills, physical ability, and technology skills.[24] From this list, we

[23] Note that about one-half of job ads on the job board we use do not specify a company name or exact location. In addition, aside from our tailoring the ads to be indistinguishable from real ads, the board monitors and filters out spam ads. For all of these reasons, we think there is virtually no reason our ads could be or would be perceived as fake and hence perhaps induce selection on who applied.

[24] These methods are explained fully in Burn et al. (2022a). However, they have a common usage with which most people are familiar. In particular, when one enters a phrase of a few words in an internet search, the first search results that are returned, which usually best match the meaning of the search phrase entered, are entries with text that is closely related to the search phrase; equivalently, these entries have a high semantic similarity score with the search

construct our treatment sentences. We iteratively edited the sentences to ensure that only the cosine similarity score of the manipulated stereotype substantively differed between the treatment and control phrases. For example, if the treatment language related to communication skills was also highly related to the stereotype about personality, we identified which words in the sentence were highly related to personality and selected synonyms that were less related to personality. Our control sentences were created to express requirements for similar jobs without referring to ageist stereotypes about skills or abilities. We iteratively removed phrases that were highly related to our stereotypes to minimize the semantic similarity. The sentences for the treatment and control groups are listed in columns (3) and (4) of Table 24.

Figure 16 shows that the treatment phrases we use fall at roughly the 75th percentile of "bias" – as measured by the distribution of cosine similarity scores of phrases from the job ads collected in Burn et al. (2022) with the stereotypes we study – while the control phrases are close to the median. Thus, the control phrases should not be regarded as ageist by the average job seeker reading the text, while the treatment phrases should be viewed as reflecting relatively subtle or indirect ageist language, rather than blatant language conveying disinterest in hiring older workers.[25]

Our second treatment conveys bias by using ageist language identified by AARP as the text related to communication skills, physical ability, and technology skills. We select three AARP examples that correspond to our respective stereotypes: "cultural fit," "energetic person," and "digital native" (Brenoff, 2019; Terrell, 2019). We adapted the language to fit our job ads and created three sentences, one for each stereotype (Table 24, column 5). Using the text about cultural fit, we created the phrase "You must be up-to-date with current industry jargon and communicate with a dynamic workforce" to reflect stereotypes about communication skills, emphasizing the commu-

phrase. This semantic similarity is determined by how commonly the phrase one enters, and the text in the returned search results, are used together in an extensive body (corpus) of English-language text – such as all of the articles on Wikipedia. More specifically, semantic similarity is measured by the cosine similarity (CS) score. The CS score varies between -1 and 1. A score of -1 means the words never appear in the same sentences or paragraphs in the corpus. As the CS score increases, the usage of the words becomes more similar; that is, they are used more often in the same sentences or paragraphs, suggesting that they are often used to discuss the same topic. A CS score of 1 means the words essentially coincide perfectly.

[25]This is somewhat lower than the types of phrases analyzed in Burn et al. (2022), who focused on phrases above the 90th percentile. But the usage of phrases closer to the 75th percentile provides greater insight into the types of phrases that are more common to observe in job ads, and are also less obviously ageist.

nication aspect of fitting in. Using the text about energetic persons, we created the sentence "You must be a fit and energetic person" to reflect stereotypes about physical ability. Using the text about digital natives, we created the sentence "You must be a digital native and have a background in social media" to reflect stereotypes about technology skills by emphasizing social media. We were interested in these more "extreme" phrases to test whether our experiment was informative, so we would be able to better interpret evidence of no significant effect of our more subtle age-stereotyped phrasing based on machine learning, if that we what we found (which was not the case).

We vary the combination of treatment and control phrases used in a job ad to create six job ads from each template: one control ad and five treatment ads. In our control ad, we use all three control phrases to express the skill requirements in language unrelated to ageist stereotypes. Four of the treatment ads utilize machine learning derived stereotyped phrases. We have three ads where we use the stereotyped phrase for either communication skills, physical ability, or technological skills (i.e., only one at a time) and the control phrases for the other two stereotypes, and there is one ad where we use all three treatment phrases. In the AARP treatment, we use all three treatment phrases.

It might seem unsurprising if the AARP phrases deter older workers from applying for jobs. However, our machine-learning generated phrases are far more subtle, and as Figure 16 shows are by no means outliers, relative to the text of job ads, in their semantic similarity with age-related stereotypes. In addition, as shown in Burn et al. (2022), the kinds of age-stereotyped phrases from the job ads that we use help predict age discrimination by employers, as measured in the correspondence study. In other words, our experiment provides evidence on the effects of real-world job ads with language that reflects ageist stereotypes relatively subtly, and that is sometimes used by employers who – based on experimental evidence – discriminate against older workers in hiring.

Validating the Treatment vs. Control Differences

One question about our treatments is how well the stereotyped vs. neutral phrases generated by the machine learning convey the intended stereotypes. In the language of epidemiology, we would like our treatment ads to have high "sensitivity" (conveying ageist stereotypes) and "specificity" (conveying information about the specific ageist stereotype intended).

Figure 17 provides an example (for administrative assistant ads) illustrating how the semantic similarity scores differ across the templates for the treatment and control job ads, and they show that our treatment job ads do activate the intended stereotypes. In this figure, words have been aggregated up to three-word phrases to ensure that we measure semantic meaning more accurately. Information on the distribution of all phrases found in the ads in Burn et al. (2022) is shown in grey, information for the treatment ads is shown with dashed black lines, and information for the neutral ads with solid black lines. The figures show the median to 99th percentile range and the average (with plotting symbols).

The figure indicates that biased (treatment) templates have a considerably higher 99th percentile than the control templates, as well as a higher mean (and median, although less so). The implication of the differences in the means and especially the upper tails of the distributions is that the treatment ads we write using the stereotyped language do, in fact, create ads with notably stronger age stereotypes. In addition, we see – importantly – that our treatment ads with single stereotyped phrases only generate a shift in similarity for the stereotypes we are seeking to activate, hence isolating those stereotypes in the job ads. Finally, note that the actual "collected" ads are more similar to the treatment ads – reflecting the fact that actual job ads often use ageist stereotypes (as documented in Neumark et al. (2019a)).[26]

The second way we assessed the validity of the treatment vs. control ads as activating the intended stereotypes was to conduct a validation exercise using Amazon MTURK.[27]We found that the control phrases were not perceived as ageist by applicants, and treatment phrases were perceived as more ageist than the control phrases. The AARP phrases were perceived as the most ageist, with the machine learning phrases intermediate between the AARP and control phrases, as

[26]Corresponding figures for other occupations are provided in Burn et al. (2023).

[27]We restricted the sample to U.S. residents, using manipulation checks to help ensure this. To guarantee that the median age of the sample was roughly 50, we used age-based quotas with a third of the sample in each of the following age bins: 25 to 35, 45 to 55, and 55+. (These age bins are pre-set by Amazon MTURK.)

we would expect. The survey and results are detailed in Burn et al. (2023).

Figure 18 presents this evidence, providing a graphical depiction of the answers from the MTURK survey participants. Across the three blocks of the survey that solicited respondents' self-assessments of age bias, their predictions of previous respondents' answers, and their predictions of the answers of respondents over the age of 50, our results were consistent. The participants, on average, strongly disagreed with the notion that anyone would perceive the control phrases as biased against workers over the age of 50.[28]Respondents rated the physical and technology-biased phrases derived from our machine learning methods as more biased than the control phrases, but viewed the communication skills stereotyped phrases as roughly identical to the controls; as shown in Table 4 of Burn et al. (2023), the differences for the physical and technology-biased phrases were strongly statistically significant. Views of the AARP-derived treatment phrases were starker, as all three were rated as far more age biased than their respective control counterparts. The absence of evidence for bias for the language related to communication skills may reflect the fact that older workers are not always stereotyped as having worse communication skills but are sometimes, as Table 23 showed, perceived as having better communication skills. In that sense, one might view the evidence of ageist ratings for the physical ability and technology-related stereotypes but not the communications stereotype as further confirmation that respondent perceptions accord with the industrial psychology literature. (Note that the cosine similarity scores from the machine learning do not detect positive vs. negative uses of the language.)

These results, like those in Figure 17, imply that our phrases capture real ageist sentiments and will be perceived as such by job applicants, so our results should be informative about the effect of ageist language on job ads on job applicants' behavior. That is, failure to find an impact would be informative about job applicant responses to age-stereotyped job-ad language, rather than reflecting a failure to convey these stereotypes in the job ads.

Posting the Job Ads

We had a total of 18 ads to post in each city, six per occupation. We staggered the posting of

[28]In the figure, values to the right are associated with less perceived bias.

ads to leave two weeks in between the taking down of one ad and the posting of the next within each city. To avoid p-hacking, we initially planned to run the experiment for 54 weeks, with the schedule of posting pre-registered.[29] To maximize the number of potential applicants, one ad was to be posted each weekday (Monday through Friday).[30] The rotation of the ads posted was staggered such that there were eight weeks between the same occupation's ad appearing in the same city with different treatment statuses.

This was a complicated process. The job board we used for the experiment makes money from fees for posting job ads, and hence is sensitive to fake ads, ads used for phishing, etc. In the course of the experiment, we encountered problems if we tried to use the same credit card to pay for ads in different cities or used the same IP address for posting ads in different cities. In addition, there seem to be human "checkers" for each city on the job board, who monitor for highly similar ads or ads that appear to be from fictitious companies.

To get around the payments problem, we used a very large number of gift cards, so we would never use one card more than four times.[31] Even this required some workarounds, as the websites for some gift cards made it difficult or impossible to register a large number of cards from the same IP addresses over a short period of time – sometimes prompting impossible to resolve "are you a robot" questions or tasks. This is apparently because gift cards are used by those who steal credit card information,[32] or others (like money launderers) who want to avoid detection.[33] We thus had to experiment to identify gift cards that did not have this constraint. To get around the problem with IP addresses, we purchased numerous cell phones and SIM cards for each city, using pay-as-you-go plans that randomize the IP address each time the service is restarted. Figure 19 conveys an idea of what was involved.[34]

[29] As discussed below, we anticipated some difficulties in placing ads in some cities, and the PAP explicitly called for a period subsequent to the initial 54 weeks when we would re-try placing these ads, in the same order.

[30] Initially, all ads were randomized to either Monday or Tuesday, but we had to switch to five days a week to avoid triggering a moderator response. The job ad board suggests not to post ads more often than every 48 hours.

[31] We used gift cards up to $100, which, depending on the city, would cover two to four ads.

[32] See, e.g., https://www.forbes.com/sites/laurengensler/2017/01/11/gift-cards-money-laundering/?sh=5498ac6f1449.

[33] See, e.g., https://www.fraud-magazine.com/article.aspx?id=4294967696.

[34] Our university grants administrator was frequently puzzled by the receipts submitted for reimbursement.

There was no way around the human checkers. A number of our ads were flagged by the job board as spam and taken down before they had been active for a week, or our payment method was rejected, leading to a delayed or skipped job posting. If the ad was taken down before we received responses, we began to repost it at the end of the study, starting in week 55. For city-occupation cells where multiple ads were taken down, we reposted them in the order that they were originally meant to be posted, still leaving one week in between each ad.

For two cities (Boston and Pittsburgh), we were unable to post many ads, due to flagging. Because of this, and because the budget allowed it, we replaced these two cities and added two additional cities (Seattle, Washington, D.C., Minneapolis/Saint Paul, and San Diego), early on in the experiment after we encountered problems. These cities were selected due to having higher numbers of job postings on the job board, which increased the likelihood that ads were being seen. Furthermore, we chose to add more than one city to replace the two problem cities in case problems emerged in other cities, based on our early experience with posting ads in Boston. Because these cities were not specified in the Pre-Analysis Plan (PAP), we also report key results for the originally proposed cities only; the results were qualitatively very similar.

Consistent with our PAP, we collect responses to our ads that we received within one week of the posting. We found, early in the experiment (when we were testing our procedures), that very few responses arrived after one week. Additionally, with our design and schedule, no two ads based on the same occupation were ever concurrently available on the job posting board.

Collecting Responses

Usually, applicants sent us their resumes when they replied to our job ad. To reduce the cost of applying for our fake job, within 24 hours of their application we informed applicants that they were not selected for an interview, via an email.[35] While it is rare for employers to inform their applicants of a negative outcome, we believe that this was important to reduce possible costs

[35] Emails read: "Thank you for your interest in this position. Unfortunately we will not be pursuing your application at this time." If someone expressed interest in applying and had a question, we provided the same response. If someone responded and only had a question about the ad, we did not reply nor did we have resume data to include (nor an email address from the resume). We did not respond to recruiters. If a job placement agency (e.g., refugee resettlement) sent a resume on someone's behalf, we sent the response addressed in the third person ("we will not be pursuing X's application").

to participants.[36] We understood, in designing this study, that there are potential ethical issues involved, and we did not have the capacity to provide real jobs to applicants – as is possible, for example, in experiments on job platforms that offer inexpensive, short-term employment for specific tasks (e.g., Pallais, 2014). We thus chose these procedures to minimize potential harm to job seekers. (And, as noted earlier, respondents were debriefed and offered the opportunity to have their data withdrawn. Only a handful of respondents chose to do so.)

Collecting Responses

Our primary outcome of interest is the age of applicants. We calculate the age of our applicants based on the available information listed on the resume. The first method to calculate age is based on the year of high school graduation (or equivalent).[37] Assuming that an individual was approximately 18 years old when they graduated high school, age is calculated as

$$Age{=}Date\ of\ Job\ Post{-}Year\ of\ HS\ Graduation{+}18.$$

We also calculate age based on the earliest date of work experience listed on the resume.[38] In this case, age is calculated as

$$Age{=}Date\ of\ Job\ Post{-}Year\ of\ First\ Job{+}16.$$[39]

Applicants were assigned the oldest age calculated across these methods.[40]

A concern is that ageist language may cause older applicants to hide their age, so the above methods to calculate age may undercount the number of older workers applying to the job ads with ageist language, creating a bias towards finding that ageist job-ad language deters older job

[36] To try to avoid spam responses and to make sure that the applicants have read the job ad, we included a manipulation check. Each ad contained a cue that applicants were asked to respond to that does not appear out of place in a job ad, such as "Please indicate which days of the week you are available to work." However, a majority (66.9%) did not respond to the cue.

[37] In a small number of cases this could be the year of getting a GED or starting post-secondary education (even if a GED was received later).

[38] In a small number of cases this could be the year of getting a GED or starting post-secondary education (even if a GED was received later).

[39] This is probably best interpreted as "minimal possible age" assuming one did not start working before age 16.

[40] E.g., for an applicant with a 2021 job posting date who had a high school graduation year of 2014 (implying age 25), an earliest working year of 2007 (implying age 30), and the earliest non-work year listed on their resume of 2010 (implying age 28), we would assign them as being 30 years old.

applicants. To address this concern, we use a binary indicator to record whether or not we can determine an applicant's age from the information on the resume. If ageist language causes older applicants to manipulate their resumes to obscure their age, we should be able to capture this effect by comparing the shares of applicants whose age we cannot ascertain for job ads with ageist language and job ads without ageist language. There was also a small number of applicants who responded to more than one job ad, and for them we can see whether the reporting of age changes in response to the job-ad language. As described later, we find no evidence of applicants manipulating the inclusion of information on age or their indicated ages in response to age stereotypes in job ads.

5 Empirical Analysis

To test whether ageist language changes the composition of the applicant pool, our primary outcome of interest is the age of applicants. We calculate three measures of the age of the applicant pool using the data on age we extracted from the applications. First, we calculate the average age of applicants (excluding, in each case, those who have not provided enough information to approximate their age). Second, we calculate the distribution of the age of applicants and identify the median and 75th percentile. Third, we calculate the share of the applicants aged 40 and over. Finally, as a check on manipulation of applications in response to job-ad language, we calculate the share of applicants to a job who do not provide information to approximate their age.

To estimate the effect of ageist language on whether older applicants apply, we estimate a series of models. We first estimate a regression equation that distinguishes between any type of age-stereotyped job-ad language treatment and the control job ads, defining the dummy variable S^A to equal one in the case of any of the five treatment arms):

$$(3)\ A = \alpha + \beta * S^A + X\delta + \varepsilon.$$

A is, alternatively, the average age of applicants to the job posting, the median, the 75th percentile, and the share of applicants over 40 years old. Observations correspond to the city, occupation, and job ad cell.

We estimate the effect of the stereotyped language in an ad (S^A) on the age of applicants conditional on controls (X) for the city and occupation the ads were posted in.[41] Because the job ads vary by city and occupation, we cluster the standard errors at the occupation and city levels (Abadie et al., 2022).

Our null hypothesis is that the presence of ageist language on a job ad has no effect on the share of older workers that apply to the job ad (i.e., $\partial A_o / \partial S = 0$, where S represents, generically, the different versions of stereotyped treatments that we use). The alternative hypothesis is that the presence of ageist language in a job ad will reduce the share of older workers that apply to the job ad (i.e., $\partial A_o / \partial S < 0$). We do not think a two-sided hypothesis test is the most meaningful in our context, but we report test results from both one-sided tests and two-sided tests (because two-sided tests are more conventional).[42] For the one-sided tests, our null hypothesis is that stereotyped language in a job ad does not deter older applicants from applying for a job. If this hypothesis is true, then β will be greater than or equal to zero. If we find that β is negative, this is evidence in favor of the alternative hypothesis that stereotyped language in a job ad deters older applicants from applying for a job (or reporting age).

As can be seen in Figure 18, there is a significant difference in the perceived age bias of the AARP treatments and the machine learning treatments. Therefore, we next estimate whether there is a difference in responses when using stereotyped language determined by the machine learning and when using the stereotyped language provided by AARP (*AARP*). To do this, we modify equation (3) to include an interaction between the dummy variable for stereotyped language in an ad and a dummy variable for using the AARP language.

$$(4)\ A = \alpha + \beta_1 * S^A + \beta_2 (S^A * AARP) + \delta X + \varepsilon.$$

Then, we test estimate the response to job ads with a single age-stereotyped phrase, by restricting the sample to the controls and, alternatively, the observations on job-ad language with a single

[41] We show that the results are robust to including the full list of controls from the experiment, which in addition to city and occupation includes month posted and day of the week posted.

[42] Our PAP generally focused on the one-sided hypothesis.

age-stereotyped phrase related to communication skills (C), physical ability (P), or technology (T). In this case, we estimate a version of equation (3) of the form:

$$(5)\ A = \alpha + \beta * S^j + X\delta + \varepsilon,\ j\text{=}C,P,or\ T.$$

We next estimate the most comprehensive model that distinguishes the five treatment arms:

$$(6)\ A = \alpha + \beta_1 * S^A + \beta_2(S^P * P) + \beta_3(S^T * T) + \beta_4(S^A * All3) + \beta_5(S^A * AARP) + \delta X + \varepsilon.$$[43]

In this model, the identification of the effect of the specific stereotypes comes from the machine learning generated phrases and not the AARP phrases because the AARP treatment only ever includes all three stereotypes, while we separately enter each stereotype for the machine learning generated phrases.

Finally, rather than estimate models with dummy variables for different treatments, we "index" the treatments by either the cosine similarity scores of the ads (see Figure 16) or the measure of perceived age bias from the MTURK survey (see Figure 18). In the latter case, we use the Likert scale elicited in the survey, but reversing the order, relative to Figure 18, so higher is more biased – consistent with the cosine similarity score. When there is more than one job ad in the treatment, we compute the average value. We thus estimate:

$$(5)\ A = \alpha + \beta * Score^j + X\delta + \varepsilon,\ j\text{=}CSS\ or\ Likert.$$

In both cases, we standardize the index so that β measures the effect of a one standard deviation increase in the index of age bias in the job ad.

6 Results

Descriptive evidence We begin by presenting the empirical cumulative distribution functions of applicant ages for the treatment and control ads, aggregating across all of the treatment ads (Figure 20). The raw data clearly show that the treatment ads that include ageist stereotypes attract fewer

[43]The PAP also calls for estimating heterogeneous effects along other dimensions. These analyses are described below, after our main results.

older applicants than the control ads. The CDF for the control ads is lower throughout nearly all of the distribution, consistent with the treatment arms – combined – leading to fewer older applicants. In Figure 21, we disaggregate the different treatment arms. There is a good deal of heterogeneity, but one can discern that the CDF for the control ads is generally lower throughout the distribution, again consistent with all of the treatment arms involving ageist stereotypes leading to fewer older applicants. One can also see that the treatment ads with three ageist phrases together, and the AARP treatment, are most pronounced in attracting fewer older applicants. Finally, among the ads with one ageist phrase, ads with an ageist phrase related to communication skills or physical ability tend to attract fewer older applicants than ads with an ageist phrase related to technology.

In Figure 22, we examine the empirical density functions of applicant ages by treatment arm. Displaying the data this way gives a clearer indication of the ages at which job searchers are less likely to apply when faced with ageist phrases in job ads. Looking at any treatment, in the upper-left panel, the under-representation is evident between about ages 40 and 60, whereas there are more younger applicants for the treatment job ads. For the individual communications and physical stereotypes, the same pattern is evident. In the lower-left pane, for the treatment including all three machine-learning generated phrases, the lower application rate from older workers is more marked and extends down to about age 35, indicating a stronger response when the job ad contains three stereotyped phrases. In the lower-right panel, for the AARP phrases, the separation between about ages 40 and 60 is stronger still, as is the greater representation of younger job applicants.

Regression results, combined treatments

We first present the regression results for the simplest specification (equation (3)), comparing the ages of respondents to job ads with any type of age-stereotyped language to the controls. These estimates are reported in the top panel of Table 25. Job applicants to ads with ageist language are on average 2.7 years younger than applicants to the control ads. The median age for these ads was also 2.7 years younger, and the 75th percentile was 3.1 years younger. The share of applicants over 40 was lower by 9.4 percentage points. All of these estimates are statistically significant at the 5% level or less, in both one-sided and two-sided tests.

The bottom panel introduces an interaction with the AARP treatment (equation (4)). In this case, the estimated coefficients of Any Treatment correspond to the machine learning phrases, the estimated effects of the AARP dummy variable are the differences between that treatment and the others combined, and the total effect of the AARP treatment is the sum of the coefficients. The first and perhaps most important result is that the effects of Any Treatment – which now exclude the AARP treatment – remain large. Job applicants to the ads with machine-learning generated ageist phrases are on average 2.2 years younger than the applicants to the control ad. The median age for these ads was 2.4 years younger, and the 75th percentile was 2.5 years younger. The share of applicants over 40 was lower by 7.9 percentage points. Except for the effect on the 75th percentile, all of these estimates are statistically significant at the 5% level or less, in both one-sided and two-sided tests. The second result is that the incremental effect of the AARP treatment is also large and statistically significant. For example, average age of applicants was lower by an additional 2.3 years, and the share of applicants over 40 was lower by an additional 7.7 percentage points. All these differentials are statistically significant at the 5% level or less, in both one-sided and two-sided tests.

Underlying changes in applications The decline in the average age of applicants or the share of applicants over age 40 could in principle be generated solely from increases in the number of people under age 40 who apply for jobs. That could happen because the ageist stereotypes are presented – as is typical in job ads – in a positive rather than a negative light: that is, as skills or attributes an applicant should have. This would still imply that the age composition of the applicant pool is affected by ageist stereotypes in job ads and the effects of the job-ad language would serve to obscure shortfalls in hiring of older workers. But if the only effect was to increase applications from younger workers, it would be incorrect to interpret our evidence as indicating that ageist language in job ads discourages older job seekers from applying for jobs.

However, we confirmed that in the analyses just reported, as well as those in the tables that follow, the experimental treatments generate declines in applications from both younger and older applicants, but the relative decline for older applicants is about twice as large (and the change for younger applicants is not significant). The key result is that applications for younger job seekers do not increase. This is not counter-intuitive, because despite the treatment phrases (see Table 24) expressing language that favors the young, they also generally also express higher skill demands. Hence, the evidence indicates the job-ad language related to ageist stereotypes does discourage applications from older job seekers.[44]

Manipulation of age reporting?

Before moving on to the other specifications, we consider evidence on two other issues. First, because workers could, in principle, strategically mask their age in responding to ageist job ads, we test the effect of ageist language on the share of applicants providing no age-identifying information. This analysis can be viewed as studying selection into reporting age.[45] As reported in the last column of Table 25, we find no evidence of this kind of selection, and hence do not need to be concerned with this potential source of bias. The same is true in the tables that follow; since the result is the same, we include the corresponding column in each table, but do not discuss this result anymore.

A related possibility is that job applicants report information on age but manipulate this information to appear younger in response to age stereotypes in job-ad language. We do not regard this as very likely, given that at the time of job interviews, employers learn about a job applicant's age and can decline to hire older applicants if they do not want to do so.[46]

We can actually garner some evidence on this question from repeat applicants. In particular, there were about 400 observations on individual applicants who applied to more than one job ad.[47] For these observations, we can test the effect of the experimental manipulation of the job-ad lan-

[44]This analysis was not in the PAP but was suggested subsequently. Results are available upon request.

[45]In a different context, Kang et al. (2016) report interview evidence that minority job seekers (of university age) sometimes conceal or downplay racial cues in job applications, and in a lab experiment do so less in response to job ads that suggest the employer values diversity.

[46]As evidence of the importance of age as revealed at interviews, Neumark (forthcoming) studies data on hiring of older vs. younger job applicants at a single company hiring across multiple stores. In the study, hiring procedures changed from in-person interviews at stores, to age-blind online assessments followed by an interview for candidates selected after these assessments. The in-person interviews, in which the interviewer could assess age immediately, led to lower hiring of older applicants. Older applicants fared better in the age-blind online assessments, getting more interviews – likely because of more work experience – but adverse treatment of older applicants emerged after the interviews, when age became apparent.

[47]Most of these applied to two ads, but some applied to three, four, or five.

guage to see if the reported age information responded. Perhaps the most obvious hypothesis is that they would report information indicating a younger age or be less likely to provide age information for ads with ageist stereotypes. We found no evidence of this. In estimates corresponding to the specification in Table 27 (equation (6), where we distinguish the different treatments, discussed below), only two of the 15 estimates were statistically significant, and only at the 10% level. Moreover, these estimates were inconsistent with the age manipulation we might expect – with a positive effect of the physical stereotype on age, and a negative effect of the *All 3* treatment on reporting no age information. Thus, we conclude that age manipulation in response to the experimental treatment is not an issue.[48]

Multiple testing

Second, we consider multiple hypothesis testing. Given that we pre-registered our analysis plan, we are less concerned with the issue of searching for statistically significant results. Still, we are estimating multiple effects. We use the Simes procedure for the "false discovery rate" (Benjamini and Hochberg, 1995). Controlling the false discovery rate (for example, at the 5% significance level) means that we are 95% confident that at least some of the rejected null hypotheses are false. Given that our hypothesis is somewhat general – that ageist language in job ads may deter older job applicants – and that we do not have a strong hypothesis about either a specific treatment among those we consider, or a specific measure of the age of applicants, the false discovery rate is appropriate (as opposed to more conservative multiple testing methods). The procedure results in a q-value, which has the same interpretation as a p-value, but with the multiple testing adjustment.

We apply the multiple testing correction to each panel of Table 25 and the main tables that follow (Tables 26-28), for the different measures of the age of applicants, for each treatment included in a set of specifications. For example, in Table 25, it is applied to the first four columns of the top panel (for the estimates of equation (3)), and then to the first four columns of the bottom panel, which has two treatment variables. In Table 25, we highlight in boldface those estimates that remain significant at the 5% level in two-sided tests, and in italics those that remain significant

[48]These results are available upon request.

at the 10% level. In this table, the statistical conclusions are very robust to multiple testing.[49]

Evidence for single stereotype treatments

Table 26 reports estimates of versions of equation (3) estimate for job ads with a single machine-language generated stereotyped phrase vs. the control arm (equation (5)). In each case, the sample is smaller because only one treatment arm is included. The three separate panels of Table 26 consider each stereotype in turn. For the individual stereotypes, and the smaller samples considered in Table 26, the results are statistically weaker. However, the key result, in our view, is that for every way that we measure the age of applicants – average age, median age, 75th percentile, and the share over 40 – for all three stereotypes, the sign of the estimate indicates that the ageist stereotype reduces applications from older job seekers.

For ads featuring an ageist phrase related to communication skills, the average age of applicants was 2.6 years younger than the control ads. The median and 75th percentile were lower by a similar amount, and the share over 40 lower by 7 percentage points. For ads featuring an ageist phrase related to physical ability, the average age of applicants was 1.8 years younger than the control ads. The effects on the median and 75th percentile were somewhat larger, and the share over 40 was lower by 8.2 percentage points. For ads featuring an ageist phrase related to technology, the average age of applicants was 1.9 years younger than the control ads, and the median was lower by 2.1 years, while the effects on the 75th percentile and the share over 40 were lower than for the other two stereotypes. Some of these estimates are significant at only the 10% level, and none after the correction for multiple testing. But the evidence that all 12 estimates in the first four columns are negative is nonetheless quite compelling.

Regression estimates for all treatment arms

In Table 27, we estimate a single model that uses all of the data and includes separate variables for each of the different treatment arms (equation (6)). Note, first, that every single estimated effect on the age of applicants in this table (shown in the first four columns) is negative, implying that no matter how we measure age, and for every treatment and stereotype, the age-stereotyped job ads

[49]Table A1 in online Appendix A reports the p-values and q-values for Table 25, and our other main analyses (Tables 26, 27, and 28).

attract fewer older applicants. The estimates for the job ads with a single stereotype are similar to those in Table 26, although a bit stronger statistically for the communications and physical ability stereotypes. We would not expect any difference except because the city-occupation dummy coefficients are estimated from different (more) observations, and the residual variance changes.

The treatment arms including all three machine learning phrases or all three AARP phrases generate large and strongly significant reductions in the ages of job applicants. For ads that feature all three machine learning treatment phrases, job applicants were 2.5 years younger than the applicants to the control ad. The median age for these ads was also 2.5 years younger, and the 75th percentile was 4.2 years younger. The share of applicants over 40 was lower by 11.7 percentage points. All of these estimates are statistically significant at the 5% level or less, in both one-sided and two-sided tests, and at the 5% or 10% level after considering multiple testing.

For ads that feature all three AARP treatment phrases, the estimates are even larger and more strongly statistically significant. Note that in this specification, the *AARP* effect is the full effect of this treatment, not the incremental effect (as in equation (4) and Table 25). For example, the average age of applicants was 4.8 years younger than applicants to the control ads, and the share of applicants over age 40 was lower by 15.6 percentage points. All of these estimates are strongly statistically significant.

The evidence in Table 27 for the *All 3* treatment gives some indication of a "dosage" response, with a job ad that reflects more than one stereotype more strongly signaling to job applicants an employer is less likely to hire older workers and hence reducing applications from older workers by more. However, this is not consistent; it is apparent for the 75th percentile and the proportion over 40, but not for the average or median age. Moreover, the point estimate of the effect of the All 3 treatment is well below the sum of the effects of the single stereotype treatments, indicating that even a single stereotyped phrase related to one skill or characteristic does not go unnoticed by job applicants and can have a sizable negative effect on job applications from older workers.

Regressions using cosine similarity score or Likert scale

Finally, we report the estimates substituting indices of age bias in the job-ad language for the treatment dummy variables (equation (7)). The top panel of Table 28 reports the estimates using the cosine similarity score (or scores, when averaged across multiple stereotypes). The evidence points strongly to age-biased job-ad language reducing the age of job applicants. When the index was one standard deviation higher, average age was lower by 1.7 years, median age by 1.5 years, and the 75th percentile of age by 3.1 years. The proportion over age 40 was lower by 8.3 percentage points. All of the estimates are statistically significant at the 5% level or less in two-sided tests, and after correcting for multiple testing. The evidence in the bottom panel, using perceived age bias of the language from the MTURK survey, is very similar. The point estimates are very close to those in the top panel, and the evidence is even stronger statistically.

Summary

Overall, we find significant evidence that ageist stereotypes substantially reduce the likelihood that older workers apply. For example, when all three machine-learning generated stereotypes are used in the model with all treatment arms, average age across cities was lowered by about 2.5 years (on a mean of 32.7 for the control group), and the share of applicants over age 40 was lowered by 12 percentage points (on a mean of 20.0% for the control group).

Robustness analyses

We next consider two robustness analyses.[50] In general, we do these analyses for the specifications with all treatment arms (Table 27), and for the indexes of age bias (Table 28). However, in cases where we introduce interactions, to avoid an excessive number of parameters we substitute the simpler specifications from Table 25 for those in Table 27. Given that the structure of the tables is similar to the preceding ones, we discuss these results briefly. The bottom line, though, is that none of these alternative analyses change our conclusion that ageist language in job ads deters older job applicants.

First, we test the effect of varying our definition of older workers by raising the threshold to 50 and then 65 years old. As shown in Tables 29A and 29B, the estimates for the age 50 threshold

[50] These were specified in the PAP. Online Appendix B presents additional robustness checks, as well as analyses of heterogeneous effects by gender (of which we find no statistical evidence). Most of these were specified in the PAP. We note the exceptions. None of these alter our conclusions.

are quite similar to those for age 40. The estimates for the age 65 threshold are very small and statistically insignificant – although there were very few applicants above age 65, as Figure 20 shows.

Of course, these estimates measure the effect in terms of the change in the proportion of applicants above a particular age as a result of the treatment. Since the baseline proportion of applicants declines the higher the age threshold, a smaller estimated change in the proportion above a higher age threshold can represent a larger relative effect. To provide more detail on the effect of changing the age threshold, but in terms of relative effects, we estimate the model for *Any treatment* (like Table 25) for the effect on the proportion above each age threshold from age 30 through age 70, rescaling the estimated effects by the baseline proportion. (We similarly rescale the confidence intervals, so the statistical inferences are based on the original regressions.) The results are reported in Figure 23. The estimates show that in fact the estimate at age 50 is a bit larger in relative terms (compared to 40). Moreover, the estimates show an increasing effect through about age 53, after which things get a bit less consistent. And at the oldest ages, although imprecise, the estimates are largest.

Second, we also add fixed effects for city and occupation, in Table 30. The estimates are little changed from Tables 27 and 28.

Comparing effects of age-stereotyped job-ad language vs. direct age discrimination in hiring

As noted in the Introduction, age discrimination that deters older workers from applying for jobs can have the same effect as direct age discrimination applied to job applicants; both reduce the employment of older workers. We can compare the estimated effects on the share of older job seekers hired from the discouragement of older applicants from age-stereotyped job-ad language, estimated in this paper, and the direct impact of age discrimination in hiring in the closely related correspondence study (Neumark et al., 2019a). It is important to keep in mind, though, that the evidence from both experiments is specific to the experimental conditions, and may not generalize to actual incidence of age discrimination in hiring and age-related stereotypes in job ads in the broader labor market. In addition, we have no way to link directly the discouragement of applications from older workers, which we find in this experiment, to hiring decisions. Rather, we are doing a back-of-the-envelope calculation in which the discouragement effect maps directly into an effect on who gets hired. By the same token, in this calculation we assume that callback differences from the correspondence study translate into hiring differences.

In our experiment in this paper, the share of applicants over 40 in the control group is 20.00%; the use of ageist language reduces the share of job applicants over age 40 by 4.41 percentage points.[51] In the correspondence study, the overall callback rate for the over 40 group (averaging across those near age 50 and near age 65) was 13.78%, compared with 18.69% for those under 40, a shortfall of 4.91 percentage points. Clearly these effects are of a similar magnitude. However, it is more instructive to calculate the implied effects on the share of older "hires" among all "hires."[52]

- · If there were no age discrimination in hiring and no discouragement of applications from older job seekers, the percentage of older workers among all hires would be the same as this percentage in the control group, or 20.00%.
- · Age discrimination reduces the percentage of older workers among hires to 15.56%.[53]
- · The discouragement of older applicants reduces the percentage of older workers among hires to 16.31%.[54]

Thus, the two effects are very similar, with the discouragement effect only slightly smaller.[55]

---

[51] We use the individual-level data to be comparable across the two experiments.

[52] We equate "hiring" with "callback" for these calculations.

[53] This is computed by applying the actual hiring rates for older and younger applicants to the shares of applicants in each age group in the control group, to eliminate the discouragement effect:

$$= (0.20000.1378)/[(0.20000.1378) + (1–0.2000)0.1869)].$$

[54] Since the hiring rate for older and younger applicants would be the same, this is computed by simply adjusting downward the proportion of older applicants and recomputing the share of older applicants:

$$= (0.2000–0.0441)/[(0.2000–0.0441) + (1–0.2000)].$$

[55] Note that the summed effects exceed the overall effect by a little bit. That is because there is a negative interaction from the lower callback rate for older applicants being applied to a reduced number of older applicants. The percentage of older workers among all hires resulting from both effects is 12.56%, computed as (0.1631 × 0.1378)/[(0.1631 × 0.1378) + (1 – 0.1631) × 0.1869)]. This is a reduction of 7.44 percentage points, vs. the sum of the two effects adding to an 8.31 percentage point reduction, or [(0.2000 – 0.1556) + (0.2000 – 0.1631)].

If these numbers roughly generalize to the actual labor market, the implication is that enforcement that focuses only on hiring shortfalls could conceivably miss nearly half of age discrimination – subject also to the caveat discussed earlier that job-ad language that reflects age-related stereotypes may not solely reflect age discrimination. One might argue that there is no incremental effect of ageist language in job ads, in the real world, from employers using ageist language in job ads, given the evidence (Neumark et al., 2019; Burn et al., 2022) that the employers who use this language tend to discriminate against older workers in hiring. But the prior evidence is based on hiring behavior relative to experimental applications, which are by construction balanced by age. Deterring applications from older workers is an additional effect.

One might object that this discouragement effect is not so strong in the real world, and that ageist language in job ads does not ultimately impact older workers' employment much because they can just apply to other jobs. But that assumes there is a very large supply of potential jobs to which older workers might apply, and they can simply pick and choose among them. However, the supply is not that large, especially in the smaller markets. Across city and occupation cells, the median number of job ads in the same category (which remain up for 30 days) was 123, the 25th percentile was 39, and the 10th percentile was 12. (Outside of administrative assistant jobs, the numbers are considerably lower.) Moreover, recall from Figure 16 that our treatment phrases were generally at around the 75th percentile of the distributions of CS scores for the three stereotypes, based on the job ads used in Burn et al. (2022). The implication is that there are many job ads with language as stereotyped or more stereotyped than our treatment job ads. In other words, job ads with the kind of language we use in our treatment ads are not easily avoidable.

Are older applicants deterred from applying because of perceived age discrimination, or required skills or qualifications?[56]

Our evidence implies that the use of ageist stereotypes in job ads deters older workers from applying for jobs. A natural interpretation of this evidence is that job searchers perceive such job-ad language as signaling that employers discriminate against older workers in hiring, and hence

[56]This section presents analyses not in our original pre-analysis plan, but suggested by seminar and conference participants.

are deterred from applying. We have to be a little cautious in this interpretation, however, because it is conceivable that age-stereotyped language in job ads may signal job requirements that older applicants are less likely to meet (or conceivably job characteristics that older workers are less likely to prefer). While it is not possible to disentangle job applicants' thought processes, we are quite confident these interpretations do not fully explain our findings, for a number of reasons.

First, there is evidence from our previous work ruling out these alternative interpretations. In Burn et al. (2023), we observe that both older and younger respondents perceive that the machine learning and AARP phrases that we use to describe job requirements are biased against older workers. Moreover, the AARP requirements were explicitly billed as "phrases employers use to mask ageist discrimination" (Brenoff 2019). In addition, the evidence from Burn et al. (2022) indicates that employers who used physical and technologically-biased language sometimes discriminated against older men. Thus, older applicants with experience looking for jobs would be expected to learn from experience that callback rates are lower for ads that include biased language, and hence be less likely to apply to such ads. Finally, the evidence in Neumark et al. (2019) did a good deal to rule out explanations of age differences in callbacks based on statistical discrimination.

Second, the older workers deterred by our treatment phrases are largely between the ages of 40 and 60. Previous research indicates that age discrimination begins in one's early 40s (Carlsson and Eriksson 2019), which suggests that discrimination begins to appear before an age group becomes obviously less qualified to fulfill the job requirements.[57]The AARP phrases, for example, do not convey any specific or objective skill requirements. In addition, the computer programs listed as required in our machine learning treatments (see Table 24) have been available for over 30 years, so many of our deterred applicants have been familiar with these programs for much of their working lives. Indeed some (like Quickbooks) might be more likely to be unfamiliar to younger workers. Given these considerations, we regard it as implausible that the differences in skill requirements

[57]Note that the evidence in this paper of discouragement of older applicants from ages as young as 40, and the evidence in Neumark et al. (2019a) of age discrimination in hiring at around age 50, is still quite relevant to population aging and extending work lives past conventional retirement ages. Real or perceived diminution of labor market opportunities in one's 40s or 50s can deter human capital investment and job search, and perhaps make reliance on other programs like SSDI more attractive, hence reducing an individual's employment long term.

could play much role in the roughly 50% decline in applications from older job seekers.

There is additional evidence from our experiment that skills or qualifications related to the stereotypes we manipulate in the experiment do not decline with age. Specifically, we took the set of resumes (and cover letters) submitted in response to our control ads. Following procedures similar to those in Burn et al. (2022), we computed the cosine similar score of all three-word phrases in the resumes (excluding "stopping words" like "the," "in," etc.) with each of the three stereotypes ("physical ability," "communication skills," and "technology skills"). To characterize how strongly a resume reflected skills or qualifications related to these stereotypes, for each stereotype we computed the 95th percentile of the cosine similarity score of each phrase with that stereotype.[58] So a resume for which this 95th percentile is higher has language describing skills or qualifications more closely related to this stereotype. We then computed the average 95th percentiles at each age. We also, as additional information, computed the average previous experience (as a share of total experience) in the corresponding occupation.

The results are reported in Figure 24. The cosine similarity scores are quite flat by age, indicating little if any meaningful decline in skills or qualifications related to the job-ad language we manipulate (in the treatment arms). Thus, based on the resumes there is little if any evidence of an age gradient in skills or qualifications related to the stereotypes we manipulate.

There is more general evidence consistent with this conclusion, as the research evidence more broadly does not provide clear evidence of productivity declines with age during the usual working years. Research on cognitive performance points to some declines in "fluid" intelligence (memory, processing speed, etc.), but stable or perhaps increasing "crystallized" intelligence (knowledge acquired over time, vocabulary, etc.). Moreover, the fluid declines are largely preserved through

[58] As shown in Burn et al. (2022), when a job ad contains few stereotyped phrases, the phrase with the median cosine similarity score with a stereotype in the ad is often unrelated to that stereotype. In the context of this experiment, this implies differences in the median cosine similarity scores in the resumes are unlikely to reflect significant differences in skill/qualification content since the differences are driven by phrases unrelated to the stereotype. Since higher cosine similarity scores indicate a stronger relation to the stereotype, selecting a higher percentile of the cosine similarity score distribution ensures that the phrases being analyzed are more related to the stereotypes. The 95th percentile is preferred to the maximum phrase to avoid the extreme tails of the distribution while still capturing differences among highly-related phrases.

the working years, and when they occur may be subtle and not affect job performance.[59] One meta-analysis finds that age is positively related to job performance (Waldman and Avolio, 1986). Another finds no relationship (McEvoy and Cascio, 1989), a result also confirmed by a narrative review (Posthuma and Campion, 2009). A set of meta-analyses by Sturman (2003) finds that conclusions are sensitive to objective vs. supervisor assessments and the complexity of the job. Ng and Feldman (2008) find that age is unrelated to core job performance but positively related to other performance measures like "organizational citizenship." A recent study by Quinby et al. (2023) studies the relationship between profitability and age composition of workforce, instrumenting with the age composition of commuting zone workforce (since declining firms will have older workers), and finds no clear relationship with profitability. The most detailed studies of age and performance are for a truck factory (Börsch-Supan and Weiss, 2016) and a service firm (Börsch-Supan et al., 2021). The first finds productivity increases until age 65 (when everyone retires). The second finds productivity declines for routine and undemanding jobs but not for more complex jobs (and increases in the most challenging jobs). While we do not have direct knowledge of productivity changes with age in the jobs we study, these research findings indicate that there is no good reason to assume that productivity does decline, especially over the age ranges we study and the younger ages at which job searchers respond to our experimental treatments. Correspondingly, there is no basis for believing that older workers are more likely to self-select out of the jobs with language related to age stereotypes that also indicate more demanding jobs.

Finally, we extend our analysis of the resume language by examining the effect of the treatment on the resumes – measured in the same way as the cosine similarity scores for job-ad language and the age-related stereotypes. As shown in Figure 25, there is no evidence that older workers select out of applying based on their skills or qualifications in response to the job-ad language in the treatment arms – i.e., the skill or qualifications composition does not change materially.[60] The

[59]These conclusions come from a recent report for the National Academies of Sciences, Engineering, and Medicine (Committee on Understanding the Aging Workforce and Employment at Older Ages, 2022). See specific findings in: Charness and Czaja, 2006; Peng et al., 2018; Salthouse, 2018 and 2019; Schaie, 2013; and Wang et al., 2013.

[60]Similar to what we did to see if applicants manipulate age on their resumes in response to the treatments, we also examined this evidence for applicants who responded to more than one job ad. We found no evidence of "within-person" treatment effects.

age gradients of the plotted lines are the same for the treatments and controls. In contrast, if we saw the lines shifting upward and sloping upward for the treatment arms, we would conclude that job searchers were responding to the specific skills or requirements in the job ads, with older workers responding to them more strongly.[61] Moreover, the line for the treatment arms is not lower, consistent with little response of younger applicants to the treatment.[62]

We believe the most natural interpretation of this evidence plus the earlier evidence is that older workers are deterred from applying to jobs with age-stereotyped job-ad language because they perceive the language as ageist and indicating employers are less likely to hire older workers, rather than because the job-ad language indicates skills or qualifications that older workers are less likely to have. That is, workers – older workers in particular – focus mainly on the ageist "cues" in job-ad language rather than the specific skill or other requirements.

There is one other piece of evidence consistent with this interpretation. In particular, we are able to assign gender to a large share of our sample (91%) based on names and Social Security data.[63] We estimate whether the physical ability stereotype (lifting 40 or more pounds) reduces the share female among applicants. If job seekers were responding to this actual requirement, or it was binding, then we would expect women to be less likely to apply to jobs with this requirement. We find no impact (the estimate is a tiny -.01 percentage point, and not significant). Rather, women and men respond similarly to the treatment phrase, implying it is the age cue rather than the actual physical requirement that influences applicant behavior.

There is one closely related implication of these findings. In light of our earlier discussion about RFOAs, the use of ageist language, per se, in job ads, might not necessarily imply age discrimination. Job ads that feature bona fide job requirements that happen to be related to age (and

[61] We verified this by estimating regressions for those 40 and over and those under 40, for the effect of any treatment on the number of applicants whose skills/qualifications based on the resumes were below vs. above the median 95th percentile cosine similarity scores for each stereotype. The results indicated that treatment induces similar reductions in the number of older applicants who are more qualified for our job ads based on their skills and qualifications (i.e., above median), or less qualified (i.e., below median).

[62] The possible slight exception is the slight downward shift of the line for the treatment arms for technology skills at most ages in the bottom panel of Figure 25. This may indicate the job-ad language regarding technology skills in the treatment arms does signal skills that some workers do not have. Still, the response does not differ by age.

[63] A name is assigned a gender based on Social Security data if more than 80% of babies born in the United States since 1950 were assigned the corresponding sex at birth.

hence ageist stereotypes) might deter older workers from applying for jobs, and similarly – per the evidence in Neumark et al. (2019) – might be associated with lower hiring of older applicants. From that perspective, were job-ad language to be added to the tools of anti-discrimination enforcement, it might be most appropriate to use it only as a potential flag for discriminatory behavior – prompting further investigation, including whether employers who use such language are still less likely to hire older job applicants, and whether that is because of bona fide occupational qualifications.[64] However, at least in the context of the jobs and labor markets we study, there is no evidence indicating that older job searchers are less likely to have the skills or qualifications associated with the age stereotypes in job-ad language. Thus, an RFOA interpretation appears to be inapplicable to the evidence we have documented of hiring discrimination against older workers (Neumark et al., 2019), discriminatory employers using age stereotypes in their job ads (Burn et al., 2022), and the evidence in this paper that these age stereotypes in job ads deter older workers from applying for jobs.

7 Discussion and Conclusion

In this paper, we conducted the first field experiment that examines how older job seekers respond to the presence of ageist language in job ads. We manipulated the language of online job ads to feature control phrases that had low relatedness to ageist stereotypes or treatment phrases that were highly related to ageist stereotypes or flagged as such by AARP. The treatment and control phrases were validated using two methods. The first shows that the machine learning phrases are only related to the specific stereotypes and are not related to any of the other stereotypes about older workers. The second method showed the phrases to individuals on MTURK and asked them to rate how ageist they perceived them to be. We found that the treatment job-ad language was viewed as significantly more ageist than the control job-ad language.

We study job ads posted in three occupations in 14 cities, with six job postings in each city-

[64]Moreover, some employers may be using this language unintentionally. In this case, guidance from the EEOC about how subtle language may deter older workers from applying can be a useful policy intervention.

occupation cell. The results indicate that older workers, when faced with ageist language in job ads, are less likely to apply for jobs, with measures like average age and the share of applicants over the age of 40 (as well as other measures) falling. The results indicate that there may be additive or "dose-response" effects of ageist language, with the effect growing with additional ageist phrases in the job ads. Job ads with multiple ageist phrases led to strong declines in applications from older job seekers. For example, when job ads included three machine-learning generated phrases with ageist stereotypes related to communication skills, physical ability, and technology skills, the share of applicants over 40 declined by 12 percentage points, and the average age of applicants fell by 2.5 years. The decline was particularly sharp in the upper parts of the age distribution, with the 75th percentile falling by 4.2 years.

We believe that the most plausible interpretation of our findings is that discriminatory employers who do not want to hire older workers use ageist language in job ads to discourage them from applying, because this makes it harder to detect age discrimination in data on hiring relative to applicants. We showed that there is not evidence that older workers are less likely to have the skills or meet the job requirements stated in the treatment ads, or to select out of applying based on a treatment vs. a control ad. Moreover, the correspondence study in Neumark et al. (2019a) showed that employers discriminate against older workers, and Burn et al. (2022) found that the discriminating employers in this correspondence study used language in job ads that was as identified as ageist by machine learning/computational linguistics techniques (language that was also perceived by older workers as ageist against applicants aged 50 or older (Burn et al., 2023).)

Our evidence has significant policy implications regarding age discrimination. We show that ageist stereotypes in job ads discourage older applicants from applying for jobs. The effects of this discouragement of applications from older job seekers can have as deleterious an impact on the hiring of older workers as can direct age discrimination in hiring; indeed, our evidence suggests the discouragement effect may be nearly as large as the direct discrimination effect. As a result, these results suggest the need for further guidance from the EEOC to employers to avoid age-stereotyped job-ad language that deters older workers from applying for jobs. Using language that explicitly deters older workers from applying is already illegal under the ADEA, but the subtler usage of ageist language that we study suggests that job-ad language that would not be flagged as explicitly illegal can still have pernicious effects on older workers in the labor market, and possibly facilitate age discrimination. Moreover, the EEOC might consider flagging for potential investigation firms that use age-stereotyped language in their job ads, recognizing that, for these firms, discrimination may be occurring even in the absence of shortfalls between the share of older applicants hired and the share of older workers who apply for jobs. The findings also imply that, in assessing evidence of age discrimination in hiring, courts may need to put more weight on evidence aside from differences between the shares of older workers among hires and among job applicants, as the share of older workers among job applicants may itself reflect the discrimination that occurs through job-ad language. Finally, of course, these same considerations may apply to discrimination against other protected groups, but such an assessment awaits research on these groups using methods similar to ours.

References

[1] Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge. 2022. "When Should You Adjust Standard Errors for Clustering? *Quarterly Journal of Economics* 138: 1-35.

[2] Adams, Scott. 2004. "Age Discrimination Legislation and the Employment of Older Workers." *Labour Economics* 11: 219-41.

[3] Ajunwa, Ifeoma. 2019. "Age Discrimination by Platforms." *Berkeley Journal of Employment and Labor Law* 40: 1-27.

[4] Baert, Stijn, Jennifer Norga, Yannick Thuy, and Marieke Van Hecke. 2016. "Getting Grey Hairs in the Labour Market. An Alternative Experiment on Age Discrimination." *Journal of Economic Psychology* 57: 86-101.

[5] Banfi, S. and Villena-Roldán, B. 2019. "Do High-Wage Jobs Attract More Applicants? Directed Search Evidence from the Online Labor Market." *Journal of Labor Economics* 37: 715-746.

[6] Belot, Michele, Philipp Kircher, and Paul Muller. 2021. "How Wage Announcements Affect Job Search – A Field Experiment." *American Economic Journal: Macroeconomics* 14: 1-67.

[7] Bem, S.L., and Bem, D.J., 1973. "Does Sex-biased Job Advertising "Aid and Abet" Sex Discrimination?" *Journal of Applied Social Psychology* 3: 6-18.

[8] Bendick, Marc, Jr., Lauren E. Brown, and Kennington Wall. 1999. "No Foot in the Door: An Experimental Study of Employment Discrimination Against Older Workers." *Journal of Aging & Social Policy* 10: 5-23.

[9] Bendick, Marc, Jr., Charles W. Jackson, and J. Horacio Romero. 1997. "Employment Discrimination Against Older Workers: An Experimental Study of Hiring Practices." *Journal of Aging & Social Policy* 8: 25-46.

[10] Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B* 57: 289-300.

[11] Börsch-Supan, Axel, Christian Hunkler, and Matthias Weiss. 2021. "Big Data at Work: Age and Labor Productivity in the Service Sector. *Journal of the Economics of Ageing* 19: 100319.

[12] Börsch-Supan, Axel, and Matthias Weiss. 2016. "Productivity and Age: Evidence from Work Teams at the Assembly Line." *Journal of the Economics of Ageing* 7(C): 30-42. Burn, Ian, Patrick Button, Luis Munguia Corella, David Neumark. 2022. "Does Ageist Language in Job Ads Predict Age Discrimination in Hiring?" *Journal of Labor Economics* 3: 613-67.

[13] Burn, Ian, Daniel Firoozi, Daniel Ladd, and David Neumark. 2023. "Stereotypes of Older Workers and Perceived Ageism in Job Ads: Evidence from an Experiment." *Journal of Pension Economics and Finance* 22: 463-89.

[14] Brenoff, Ann. 2019. "5 Ageist Phrases to be Aware of." AARP: Washington, DC.

[15] Cahill, Kevin E, Michael D. Giandrea, and Joseph F. Quinn. 2006. "Retirement Patterns from Career Employment." *The Gerontologist* 46: 514-23.

[16] Card, David, Fabrizio Colella, and Rafael Lalive. 2021. "Gender Preferences in Job Vacancies and Workplace Gender Diversity." NBER Working Paper No. 29350.

[17] Carlsson, Magnus, and Stefan Eriksson. 2019. "Age Discrimination in Hiring Decisions: Evidence from a Field Experiment in the Labor Market." *Labour Economics* 59: 173-83.

[18] Combs, Deborah H. 1982. "Striking a Balance between the Interests of Public Safety and the Rights of Older Workers: The Age BFOQ Defense." *Washington and Lee Law Review* 39: 1371-95.

[19] Charness, Neil, and Sara J. Czaja. 2006. *Older Worker Training: What We Know and Don't Know*. Washington, AARP Public Policy Institute Report No. 2006-22, Washington, D.C.

[20] Chaturvedi, Sugat and Mahajan, Kanika and Siddique, Zahra. 2021. "Words Matter: Gender, Jobs and Applicant Behavior." IZA Discussion Paper No. 14497.

[21] Committee on Understanding the Aging Workforce and Employment at Older Ages. 2022. Understanding the Aging Workforce: Defining a Research Agenda. Washington, D.C.: National Academies of Sciences, Engineering, and Medicine.

[22] Delfino, Alexia. 2021. "Breading Gender Barriers: Experimental Evidence on Men in Pink-Collar Jobs." IZA Discussion Paper No. 14083.

[23] Deming, David, and Lisa B. Kahn. 2018. "Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals." *Journal of Labor Economics* 36 (S1): S337-69.

[24] European Commission. 2000. *Council Directive 2000/78/EC of 27 November 2000 Establishing a General Framework for Equal Treatment in Employment and Occupation*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32000L0078.

[25] Farber, Henry S., Silverman, Daniel., and Till von Wachter. 2017. "Factors Determining Callbacks to Job Applications by the Unemployed: An Audit Study."*RSF: The Russell Sage Foundation Journal of the Social Sciences* 3: 168-201.

[26] Farber, Henry S., Herbst, Chris M., Silverman, Daniel, and Till von Wachter, T. 2019. "Whom Do Employers Want? The Role of Recent Employment and Unemployment Status and Age." *Journal of Labor Economics* 37(2): 323-49.

[27] Federal Register. (n.d.). *Disparate Impact and Reasonable Factors Other Than Age Under the Age Discrimination in Employment Act.* https://www.federalregister.gov/documents/2012/03/30/2012-5896/disparate-impact-and-reasonable-factors-other-than-age-under-the-age-discrimination-in-employment (viewed September 15, 2019).

[28] Flory, Jeffrey A., Andreas Leibbrandt, and John A. List. 2015. "Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job Entry Decisions." *Review of Economic Studies* 82: 122-55.

[29] Flory, Jeffrey A., Andreas Leibbrandt, Christina Rott, and Olga Stoddard. 2021. "Increasing Workplace Diversity: Evidence from a Recruiting Experiment at a Fortune 500 Company." *Journal of Human Resources* 57: 73-92.

[30] Gaucher, Danielle, Justin Friesen, and Aaron C. Kay. 2011. "Evidence that Gendered Wording in Job Advertisements Exists and Sustains Gender Inequality." *Journal of Personality and Social Psychology* 101: 109-28.

[31] Gordon, Randall A., and Richard D. Arvey. 2004. "Age Bias in Laboratory and Field Settings: A Meta-Analytic Investigation." *Journal of Applied Social Psychology* 34: 468-92.

[32] Hanson, Andrew, Zachary Hawley, and Aryn Taylor. 2011. "Subtle Discrimination in the Rental Housing Market: Evidence from email Correspondence with Landlords." *Journal of Housing Economics* 20: 276-84.

[33] Hanson, Andrew, Zachary Hawley, Hal Martin, and Bo Liu. 2016. "Discrimination in Mortgage Lending: Evidence from a Correspondence Experiment." *Journal of Urban Economics* 92: 48-65.

[34] He, Haoran, David Neumark, and Qian Weng. 2021. "Do Workers Value Flexible Jobs? A Field Experiment. *Journal of Labor Economics* 39: 709-38.

[35] Hellerstein, Judith K., David Neumark, and Kenneth Troske. 1999. "Wages, Productivity, and Worker Characteristics: Evidence from Plant-Level Production Functions and Wage Equations." *Journal of Labor Economics* 17: 409-46.

[36] Hellester, Miguel D., Peter Kuhn, and Kailing Shen. 2020. "The Age Twist in Employers' Gender Requests." *Journal of Human Resources* 55: 482-69.

[37] Ibanez, Marcela, and Gerhard Riener. 2018. "Sorting through Affirmative Action: Three Field Experiments." *Journal of Labor Economics* 36:437-78.

[38] Johnson, Richard W., Janette Kawachi, and Eric K. Lewis. 2009. "Older Workers on the Move: Recareering in Later Life." AARP Public Policy Institute. Washington, DC: AARP.

[39] Kang, Sonia K., Katherine A. DeCelles, and András Tilcsik. "Whitened Résumés: Race and Self-Presentation in the Labor Market." *Administrative Science Quarterly* 61: 469-502.

[40] Krause, Annabelle, Ulf Rinne, and Klaus Zimmerman. 2012. "Anonymous Job Applications of Fresh Ph.D. Economists." *Economics Letters* 117: 441-4.

[41] Kuhn, Peter, and Kailing Shen. 2013. "Gender Discrimination in Job Ads: Evidence from China." *Quarterly Journal of Economics* 128: 287-336.

[42] Kuhn, Peter, and Kailing Shen. 2021. "What Happens When Employers Can No Longer Discriminate in Job Ads." NBER Working Paper No. 29116.

[43] Kuhn, Peter, Kailing Shen, and Shuo Zhang. 2020. “Gender-Targeted Job Ads in the Recruitment Process: Evidence from China.” *Journal of Development Economics* 147: 102531.w

[44] Lahey, Joanna. 2010. “International Comparisons of Age Discrimination Laws.” *Research on Aging* 32: 679-97.

[45] Lahey, Joanna. 2008. “Age, Women, and Hiring: An Experimental Study.” *Journal of Human Resources* 43: 30-56.

[46] Leibbrandt, Andreas, and John List. 2018. “Do Equal Employment Opportunity Statements Backfire? Evidence from a Natural Field Experiment on Job-Entry Decisions.” NBER Working Paper No. 25035.

[47] Maestas, Nicole. 2010. “Back to Work: Expectations and Realizations of Work after Retirement.” *Journal of Human Resources* 45: 718-48.

[48] Marinescu, Ioana Elena, and Ronald Wolthoff. 2020. “Opening the Black Box of the Matching Function: The Power of Words.” Journal of Labor Economics 38: 535-568.

[49] McEvoy, Glenn M., and Wayne F. Cascio. 1989. “Cumulative Evidence on the Relationship between Employee Age and Job Performance.” Journal of Applied Psychology 74: 11-17.

[50] Modestino, Alicia Sasser, Daniel Shoag, and Joshua Ballance. 2016. “Downskilling: Changes in Employer Skill Requirements over the Business Cycle.” Labour Economics 41: 333-47.

[51] Neumark, David. “Age Discrimination in Hiring: Evidence from Age-Blind vs. Non-Age-Blind Hiring Procedures.” Forthcoming in Journal of Human Resources.

[52] Neumark, David, Ian Burn, and Patrick Button. 2016. “Experimental Age Discrimination Evidence and the Heckman Critique.” *American Economic Review Papers and Proceedings* 106: 303-8.

[53] Neumark, David, Ian Burn, and Patrick Button. 2019a. “Is It Harder for Older Workers to Find Jobs? New and Improved Evidence from a Field Experiment.” *Journal of Political Economy* 127: 922-70.

[54] Neumark, David, Ian Burn, Patrick Button, and Nanneh Chehras. 2019b. “Do State Laws Protecting Older Workers from Discrimination Reduce Age Discrimination in Hiring? Evidence from a Field Experiment.” *Journal of Law and Economics* 62: 373-402.

[55] Ng, Thomas W.H., and Daniel C. Feldman. 2008. “The Relationship of Age to Ten Dimensions of Job Performance.” *Journal of Applied Psychology* 93: 392-423.

[56] Pallais, Amanda. 2014. “Inefficient Hiring in Entry-Level Labor Markets.” *American Economic Review* 104: 2565-99.

[57] Peng, Yishang, Steve M. Jex, and Mo Wang. 2018. “Aging and Occupational Health.” In Aging and Work in the 21st Century, 2nd Edition, Shultz and Adams, eds., New York: Routledge, 213-33.

[58] Pillar, Anna, Kyrill Poelmans, and Martha Larson. 2022. "Regex in a Time of Deep Learning: The Role of an Old Technology in Age Discrimination Detection in Job Advertisements." *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pp. 13-18.

[59] Posthuma, Richard A., and Michael A. Campion. 2009. "Age Stereotypes in the Workplace: Common Stereotypes, Moderators, and Future Research Directions." *Journal of Management* 35: 158-88.

[60] Quinby, Laura D., Gal Wettstein, and James Giles. "Are Older Workers Good for Business?" Working Paper 2023-19. Chestnut Hill, MA: Center for Retirement Research at Boston College.

[61] Riach, Peter A., and Judith Rich. 2006. "An Experimental Investigation of Age Discrimination in the French Labour Market." IZA Discussion Paper No. 2522.

[62] Riach, Peter A., and Judith Rich. 2010. "An Experimental Investigation of Age Discrimination in the English Labor Market." *Annals of Economics and Statistics* 99/100: 169-85.

[63] Ryan, Ellen Bouchard, Sheree Kwong See, W. Bryan Meneer, and Diane Trovato. 1992. "Age-Based Perceptions of Language Performance Among Younger and Older Adults." *Communication Research* 19: 423-43.

[64] Salthouse, Timothy A. 2018. "Why is Cognitive Change More Negative with Increased Age?" *Neuropsychology* 32: 110-20.

[65] Salthouse, Timothy A. 2019. "Trajectories of Normal Cognitive Aging." *Psychology and Aging* 34: 17-24.

[66] Schaie, K. Warner. 2013. Developmental Influences on Adult Intelligence: The Seattle Longitudinal Study, 2
Oxford University Press.

[67] Sturman, Michael C. 2003. "Searching for the Inverted U-shaped Relationship between Time and Performance: Meta-analyses of the Experience/Performance, Tenure/Performance, and Age/Performance Relationships." *Journal of Management* 29: 609-40.

[68] Terrell, Kenneth. 2019. "Age Bias That's Barred by Law Appears in Thousands of Job Listings." AARP: Washington, DC.

[69] Tilcsik, András. 2011. "Pride and Prejudice: Employment Discrimination against Openly Gay Men in the United States." *American Journal of Sociology* 117: 586-626.

[70] Waldman, David A., and Bruce J. Avolio. 1986. "A Meta-analysis of Age Differences in Job Performance." *Journal of Applied Psychology* 71: 33-38.

[71] Wang, Mo, Deborah A. Olson, and Kenneth S. Shultz. 2013. Mid and Late Career Issues: An Integrative Perspective. New York: Routledge.

[72] Wax, S. L. 1948. "A Survey of Restrictive Advertising and Discrimination by Summer Resorts in the Province of Ontario." *Canadian Jewish Congress.* pandoc version 3.2

Figure 14: Map of Cities in Experiment

Note: This map shows the cities in the experiment. The relative size of the symbol corresponds to the total number of applicants in each city, ranging from 23 in Salt Lake City to 643 in New York City. The total number of applicants is 2,646.

Figure 15: Examples of Job Ad Templates

Administrative Assistants Template 1 (Admin Assistant)
Psychiatric office is in need of a full or part time Administrative Assistant to assist in front/back office general clerical duties. This individual will work on a several tasks and stay on course at all times. The Administrative Assistant we hire will be trained in various duties that cover the entire office.

This individual MUST possess the following:

-Exceptional customer service background to greet and register patients, answer phones, schedule appointments.
-Can multitask.
-High School diploma or GED.
-Professional attitude.
-***Communication Skill Requirement***.
-***Technology Requirement***.
-***Physical Requirement***.
-Available for flexible hours.
(Schedule hours and days will alternate every other week)

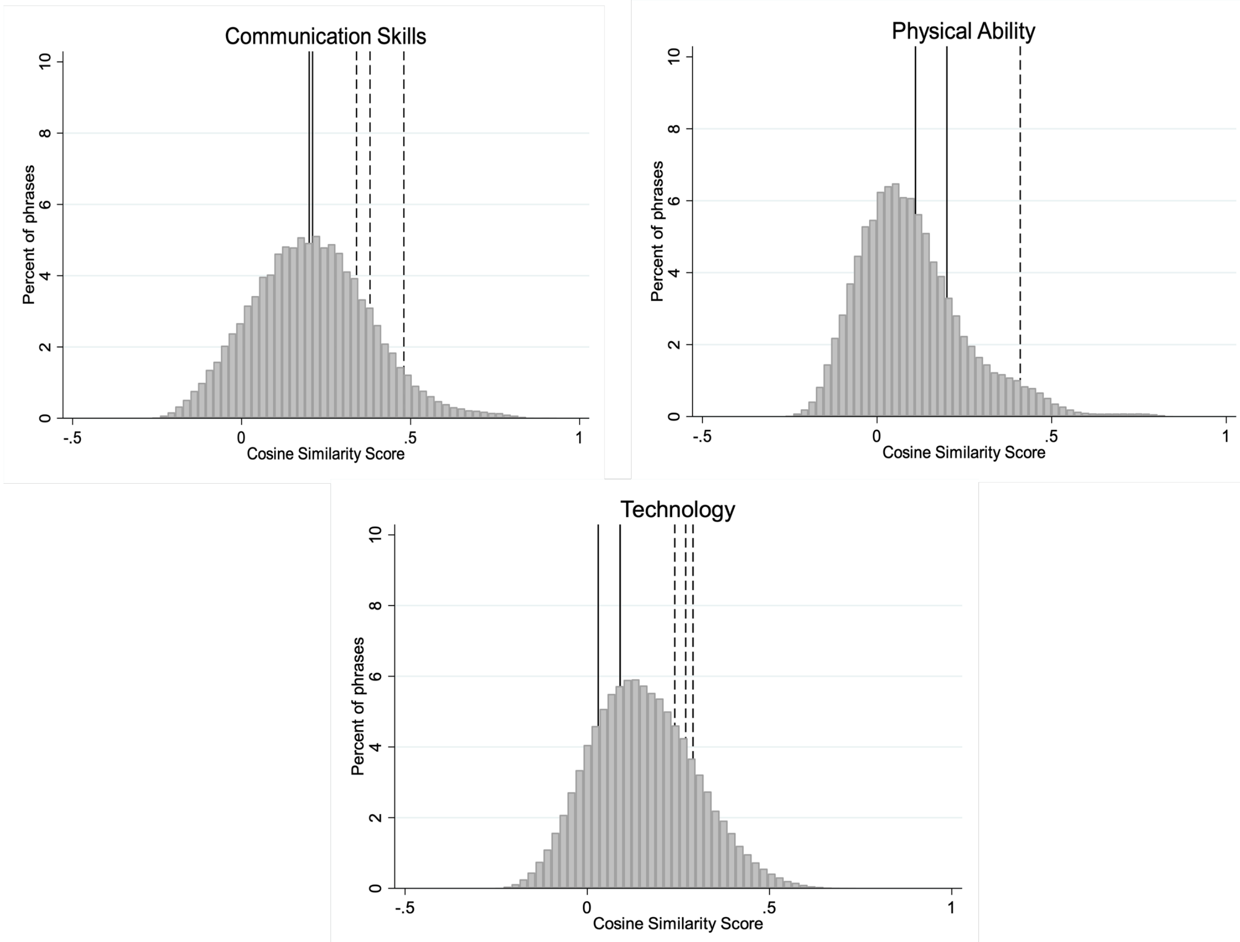
Please email us a CV or resume and put “full-time” or “part-time” in the subject line.

Retail Sales Associate Template 1 (Retail Sales Job)
Our women’s clothing store in ***City*** is looking for a sales associate to help us out weekday afternoons. We are pretty busy store and you must ***Physical Requirement***. We are looking for someone with open to working in retail, who ***Communication Skill Requirement***. We need you to ***Technology Requirement***. So if this sounds like you, send us your resume and your earliest possible starting date and we will be in touch.

Security Guard Template 1 (HIRING UNARMED SECURITY GUARDS)
We currently have a position for a full-time or part-time security officer available. Training and uniforms will provided. We offer flexible working hours and have shifts any day of the week. Our pay scale is competitive. Email your resume and potential work hours to apply.

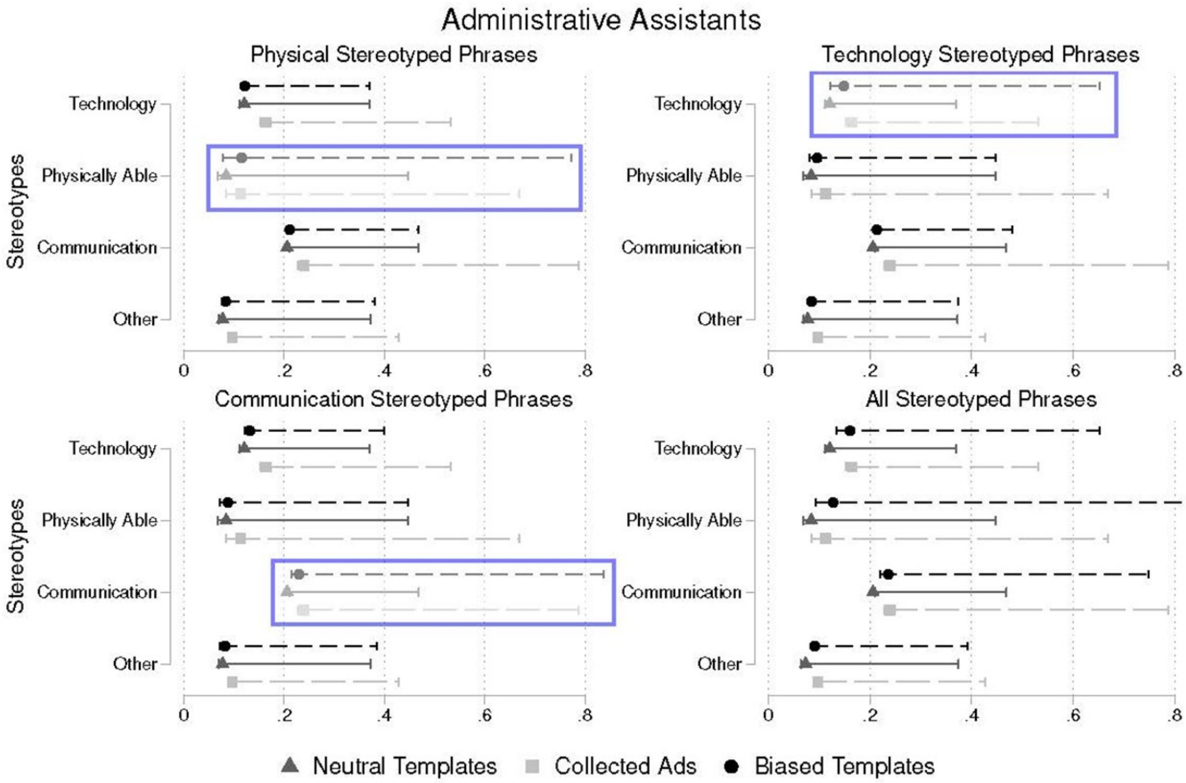
Requirements
-Professional appearance & attitude
-Detail oriented
-***Communication Skill Requirement***
-***Physical Requirement***
-***Technology Requirement***
-At least 18 years of age
-Access to transportation

Figure 16: Locations of Treatment and Control Phrases in the Cosine Similarity Score Distribution of Job Ad Phrases



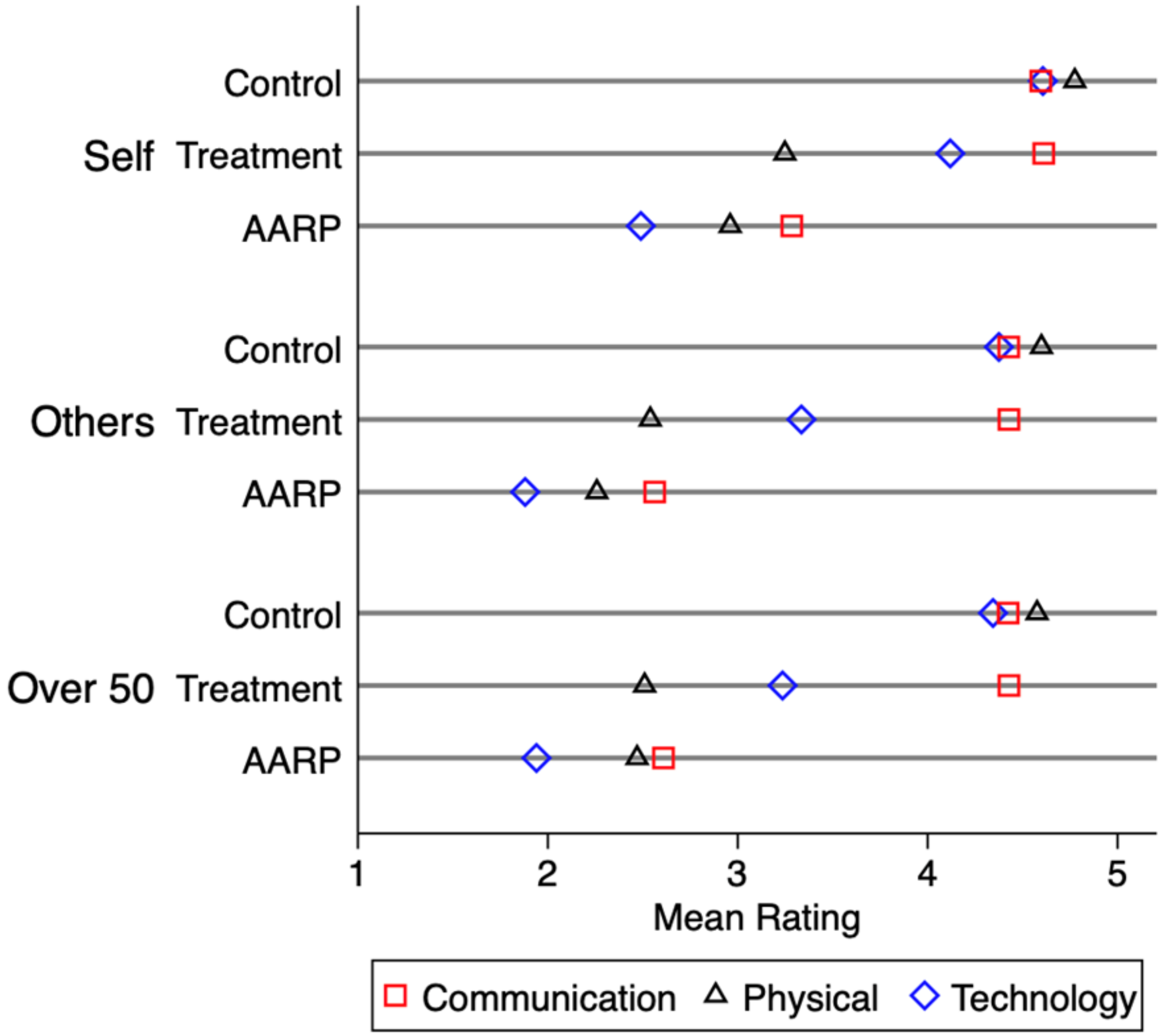
Note: Figure reports the distributions of cosine similarity scores for all trigrams from the job ads with the indicated stereotypes. The higher the cosine similarity score, the more related the trigram is to the stereotype, with a minimum of -1 and a maximum of 1. Solid lines indicate the location of a control sentence in the cosine similarity score distribution. Dashed lines indicate the location of a treatment phrase (for the Machine Learning Treatments shown in Table 24).

Figure 17: Cosine Similarity Score of Administrative Assistant Templates



Note: Graphs display median to 99th percentile range of trigram semantic similarity scores for each stereotype for Administrative Assistant ads. The average trigram semantic similarity score for each stereotype is represented by the respective shape for each template. The category "Other" shows the averages for the remaining stereotypes listed in Table 23. Control ("neutral") templates contain trigrams from the created ad templates with only non-stereotyped phrases included. Collected ads comprise trigrams from all Administrative Assistant job ads. Treatment templates contain trigrams from the created ad templates with the respective stereotyped phrase or phrases included.

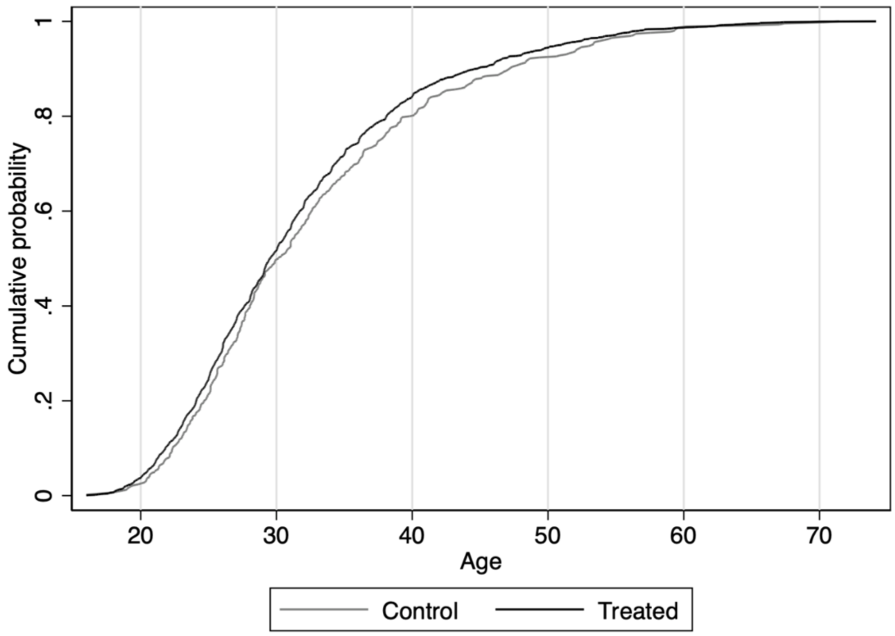
Figure 18: Survey Results



Note: These numerical ratings reflect the degree to which survey respondents rated phrases as age-biased or not age-biased, with lower numbers indicating a greater bias against older workers. Likert Scale ratings were translated to a numerical value such that "Strongly Agree" mapped to 1, "Somewhat Agree" mapped to 2, "Neither agree nor disagree" mapped to 3, "Somewhat Disagree" mapped to 4, and "Strongly Disagree" mapped to 5. The three categories "Self," "Others," and "Over 50" refer to which group's opinions the MTURK respondents were asked to provide or predict in a given survey block. The average bias rating was collapsed on the treatment status of phrases (control, treatment, and AARP) as well as the category of the stereotype (communication, physical, or technology). Hence, each point in the figure above reflects the average bias rating MTURK respondents gave to a given treatment status for a specific stereotype from the perspective of a given group of people. For example, the triangle in the first row of the figure indicates that when respondents were asked for their self-assessment of whether or not the physical stereotype control phrases were age-biased, they, on average, stated that they strongly disagreed.

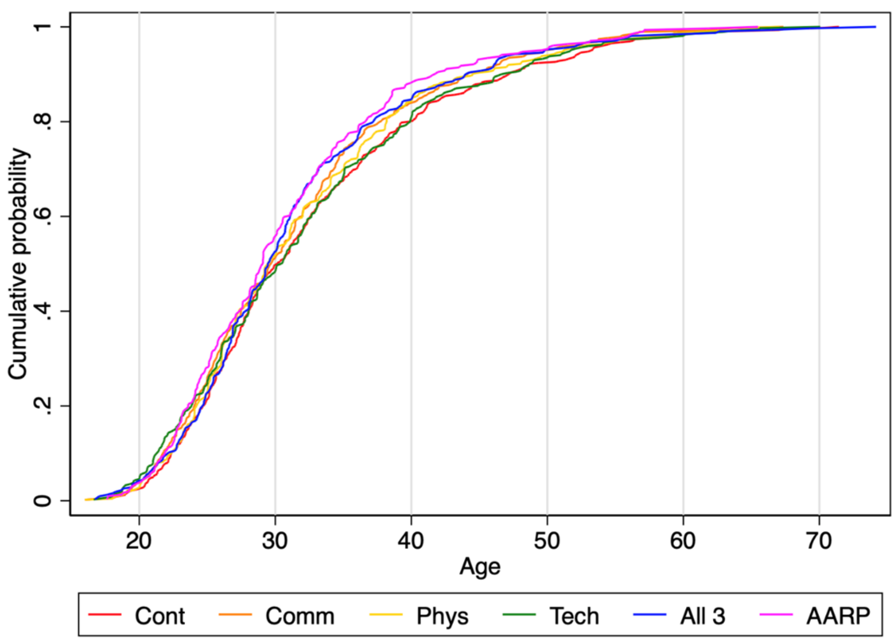
Figure 19: Posting Job Ads Was Not Easy!

Figure 20: Empirical Cumulative Density Functions, Any Treatment



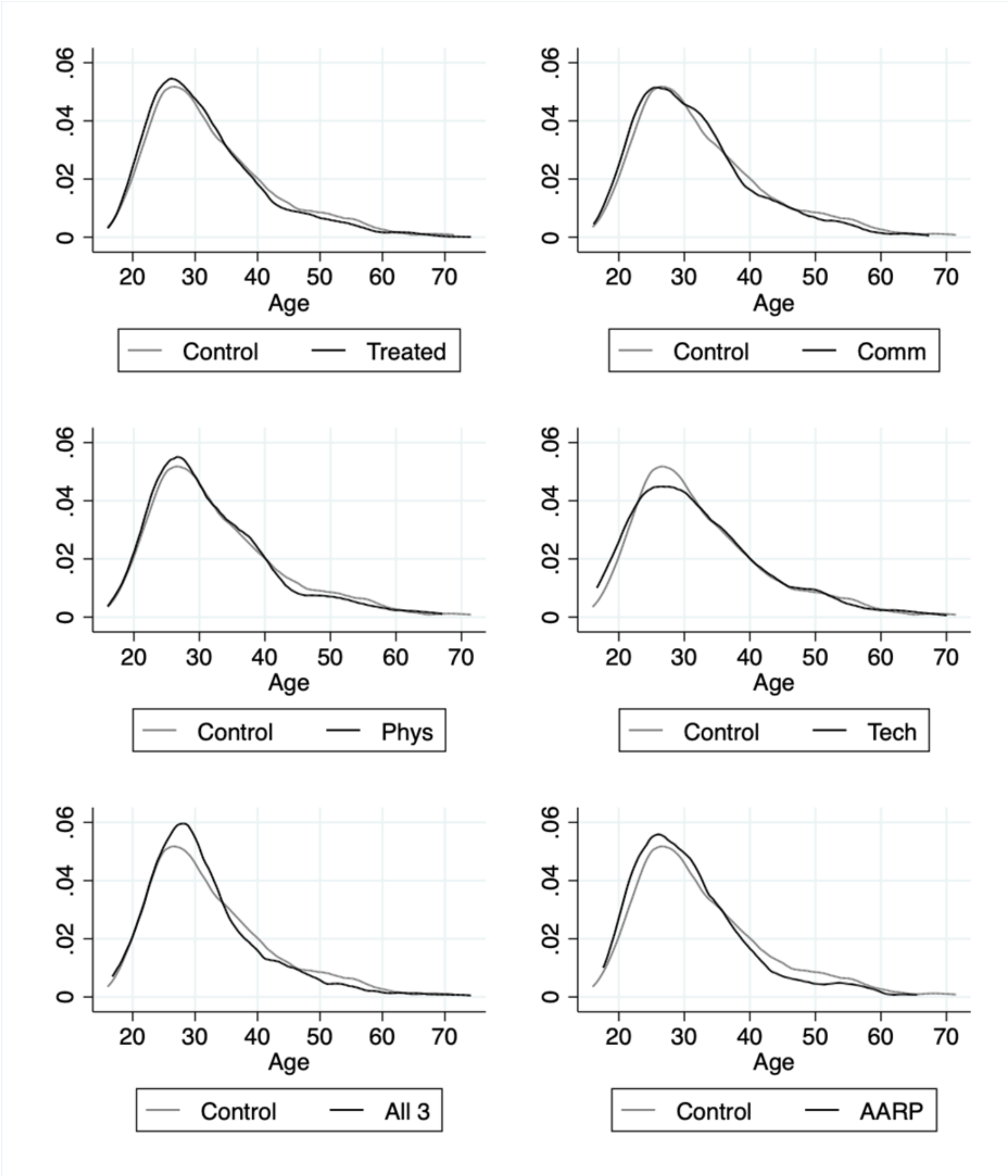
Note: “Treated” refers to any treatment (individual stereotypes, *All 3*, or *AARP*).

Figure 21: Empirical Cumulative Density Functions by Ad Type



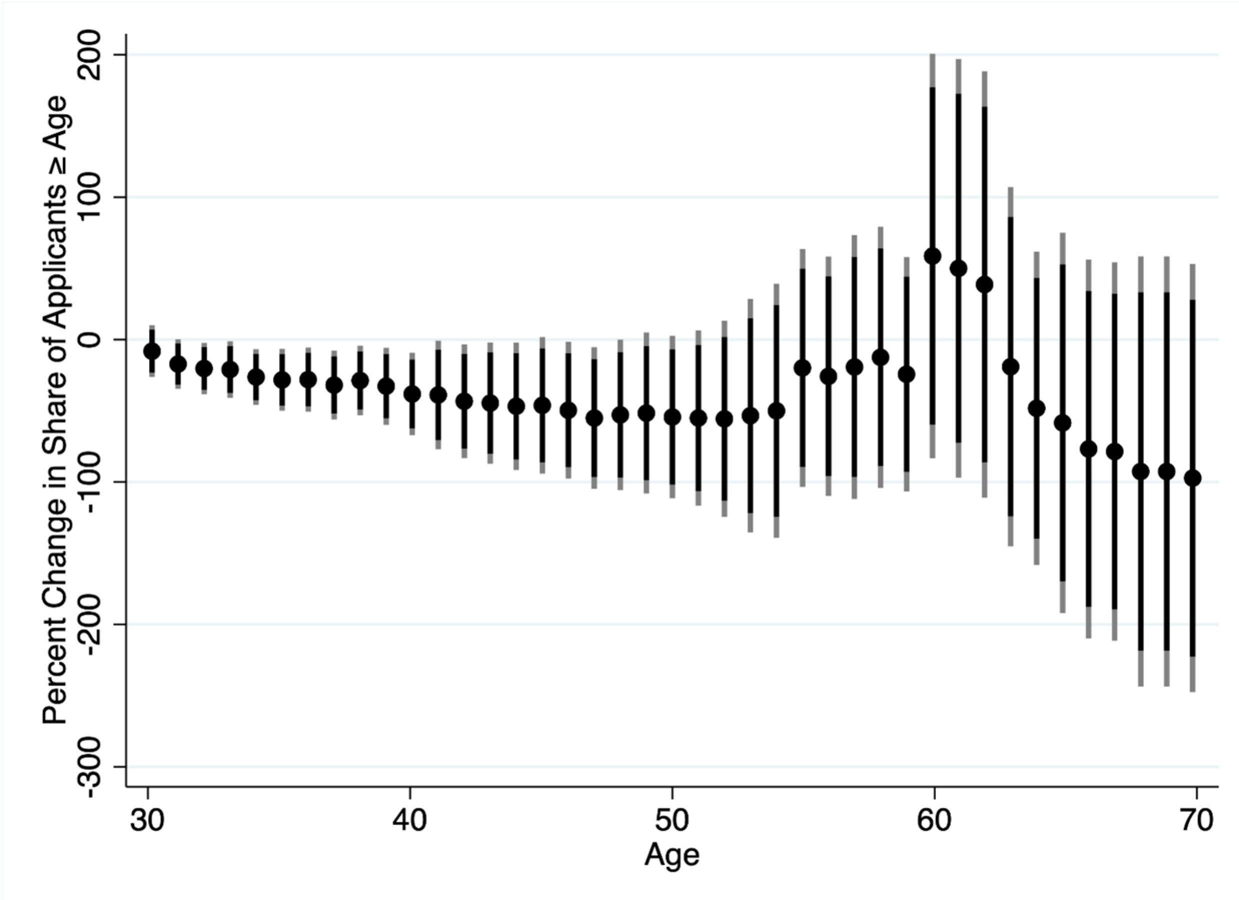
Note: “Cont” refers to controls; “Comm” to communications skills stereotypes; “Phys” to physical ability stereotypes; “Tech” to technology stereotypes”; “All3” to the ads with all three stereotypes reflected in the text; and “AARP” to the ads with AARP ageist language/stereotypes.

Figure 22: Empirical Probability Density Functions for Age



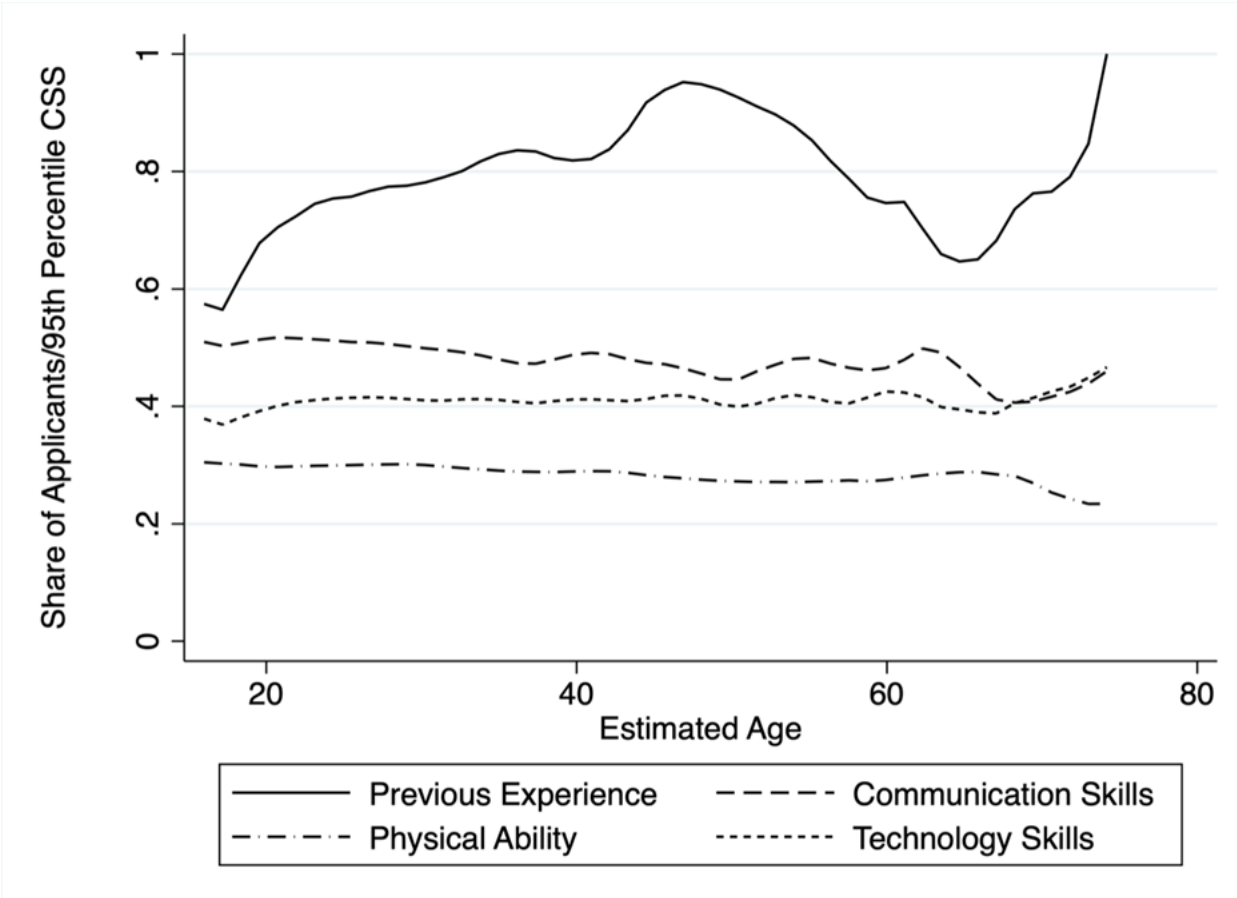
Note: In the upper-left panel, “Treated” refers to any treatment (individual stereotypes, all 3, or AARP). The other labels are explained in the notes to Figure 8.

Figure 23: Estimated Effects of Any Treatment on Proportion Above Each Age Threshold, Scaled by Proportion of Total Applicants Above Age Threshold



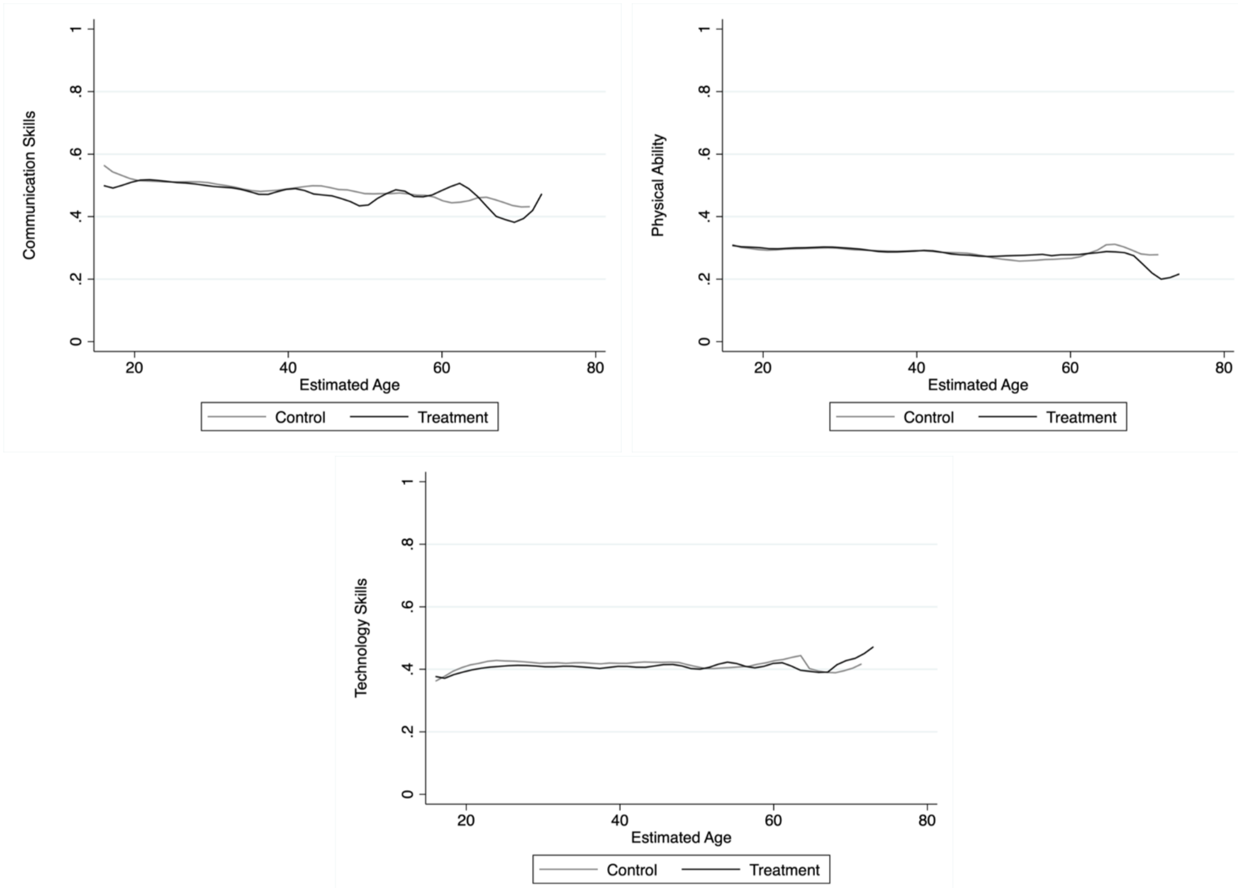
Note: 90% and 95% confidence intervals are shown, based on the regression estimates.

Figure 24: Average 95th Percentiles of Cosine Similarity Score (CSS) of Resume Phrases with Stereotypes, and Average Share of Experience in Occupation, by Age



Note: See text for explanations. These are local polynomials fitted to single-year age estimates.

Figure 25: Average 95th Percentiles of Co-
sine Similarity Score (CSS) of Resume Phrases with Stereotypes, Treatments and Controls, by Age



Note: See text for explanations. These are local polynomials fitted to single-year age estimates.

Table 23: Age Stereotypes from Industrial Psychology Literature

Health	Personality	Skills
Less Attractive	Less Adaptable	Lower Ability to Learn
Hard of Hearing	Careful	Better Communication Skills
Worse Memory	Less Creative	Worse Communication Skills
Less Physically Able	Dependable	More Experienced
	Negative Personality	More Productive
	Warm Personality	Less Productive
		Worse with Technology

Note: See Burn et al. (2022).

Table 24: Control and Treatment Phrases by Occupation

Occupation	Stereotype	Control	Machine Learning Treatment	AARP Treatment
(1)	(2)	(3)	(4)	(5)
Administrative Assistants	Communication skills	You must be good at working without supervision	You must have good communication and teamwork on tasks	You must be up-to-date with current industry jargon and communicate with a dynamic workforce
Administrative Assistants	Physical ability	You must enter bills and keep track of invoices	You must be able to lift 40 pounds	You must be a fit and energetic person
Administrative Assistants	Technical skills	You must produce and distribute documents such as correspondence memos, faxes and forms	You must use accounting software systems like Netsuite, Freshbook, and QuickBooks	You must be a digital native and have a background in social media
Retail sales	Communication skills	You must be good at working without supervision	You must have good communication with customers and staff	You must be up-to-date with current industry jargon and communicate with a dynamic workforce
Retail sales	Physical ability	You must enter bills and keep track of invoices	You must be able to lift 40 pounds	You must be a fit and energetic person
Retail sales	Technical skills	You must help to clean and organize the store	You must use software such as Microsoft Office/Excel or Google Sheets	You must be a digital native and have a background in social media
Security guard	Communication skills	You must follow instruction from supervisors	You must maintain communication about tasks with supervisors	You must be up-to-date with current industry jargon and communicate with a dynamic workforce
Security guard	Physical ability	You need to carry a flashlight	You must be able to lift 50 pounds	You must be a fit and energetic person
Security guard	Technical skills	You must write patrol records in journal notebook	You must type patrol entries into a journal application on a computer system	You must be a digital native and have a background in social media

Note: See text for a description of how each sentence was created.

Table 25: Estimated Effects on Age Composition of Applicants,
Any Stereotype Treatment, and Distinguishing AARP Treatment, All Cities

	Average Age	Median Age	75^{th} Percentile	Over 40	No Age Information
Any Treatment	**-2.687*****†††	**-2.680****†††	**-3.133****††	**-0.094****†††	0.030
	(0.990)	(1.028)	(1.523)	(0.035)	(0.038)
N	228	228	228	228	237
Any Treatment	**-2.241****††	**-2.391****††	-2.474†	**-0.079****††	0.030
	(0.990)	(1.031)	(1.560)	(0.037)	(0.037)
AARP	**-2.318*****†††	**-1.497****††	**-3.420*****†††	**-0.077*****†††	-0.003
	(0.745)	(0.709)	(1.226)	(0.026)	(0.040)
N	228	228	228	228	237

Note: The regressions include all treatment arms and the control arm. In the second panel, the AARP variable is equivalent to the interaction between Any Treatment and AARP. Bold horizontal lines distinguish separate regressions. All specifications include fixed effects for both city and occupation. Standard errors clustered at the city-occupation level are reported in parentheses. Data are collapsed to the city-occupation-job ad level. ***, **, or * indicates statistically significant at the 1%, 5%, or 10% level in a two-sided test. †††, ††, or †indicates statistically significant at the 1%, 5%, or 10% level in a one-sided test. In the first four columns, boldface estimates indicate statistical significance at the 5% level in a two-sided tests, correcting for multiple testing across the estimates in those columns, using the Simes False Discovery Rate (q-values ≤ 0.05). Italicized estimates indicate q-values > 0.05 and ≤ 0.1. (See Table A1 in online Appendix A.)

Table 26: Estimated Effects on Age Composition of Applicants, Separate Stereotype Treatments, All Cities

	Average Age	Median Age	75^{th} Percentile	Over 40	No Age Information
Communication	-2.632**††	-2.948*††	-2.583	-0.070†	0.014
	(1.232)	(1.460)	(2.043)	(0.050)	(0.040)
N	79	79	79	79	81
Physical	-1.828†	-2.062†	-2.454	-0.082†	0.022
	(1.403)	(1.394)	(2.134)	(0.050)	(0.053)
N	79	79	79	79	81
Technology	-1.930†	-2.105†	-0.728	-0.044	0.051
	(1.221)	(1.285)	(1.945)	(0.042)	(0.047)
N	79	79	79	79	81

Note: Each regression includes a single machine learning stereotype arm and the control arm. Bold horizontal lines distinguish separate regressions. All specifications include fixed effects for both city and occupation. Standard errors clustered at the city-occupation level are reported in parentheses. Data are collapsed to the city-occupation-job ad level. ***, **, or * indicates statistically significant at the 1%, 5%, or 10% level in a two-sided test. †††, ††, or †indicates statistically significant at the 1%, 5%, or 10% level in a one-sided test. In the first four columns, boldface estimates indicate statistical significance at the 5% level in a two-sided tests, correcting for multiple testing across the estimates in those columns, using the Simes False Discovery Rate (q-values $\leq$ 0.05). Italicized estimates indicate q-values > 0.05 and ≤ 0.1. (See Table A1 in online Appendix A.)

Table 27: Estimated Effects on Age Composition of Applicants, All Treatment Arms, All Cities

	Average Age	Median Age	75^{th} Percentile	Over 40	No Age Information
Communication	*-2.683***††	*-2.986***††	-2.704†	-0.075†	0.014
	(1.147)	(1.350)	(1.898)	(0.046)	(0.039)
Physical	-1.879†	-2.071†	-2.385	-0.083*††	0.034
	(1.288)	(1.276)	(2.000)	(0.045)	(0.051)
Technology	-1.889†	-2.002†	-0.707	-0.041	0.056
	(1.165)	(1.222)	(1.822)	(0.041)	(0.045)
All 3	*-2.516***††	*-2.504***††	*-4.156***††	**-0.117*****†††	0.016
	(1.122)	(1.104)	(1.754)	(0.041)	(0.048)
AARP	**-4.559*****†††	**-3.888*****†††	**-5.896*****†††	**-0.156*****†††	0.027
	(1.222)	(1.212)	(1.799)	(0.038)	(0.057)
N	228	228	228	228	237

All specifications include fixed effects for both city and occupation. Standard errors clustered at the city-occupation level are reported in parentheses. Data are collapsed to the city-occupation-job ad level. ***, **, or * indicates statistically significant at the 1%, 5%, or 10% level in a two-sided test. †††, ††, or †indicates statistically significant at the 1%, 5%, or 10% level in a one-sided test. In the first four columns, boldface estimates indicate statistical significance at the 5% level in a two-sided tests, correcting for multiple testing across the estimates in those columns, using the Simes False Discovery Rate (q-values ≤ 0.05). Italicized estimates indicate q-values > 0.05 and ≤ 0.1. (See Table A1 in online Appendix A.)

Table 28: Estimated Effects on Age Composition of Applicants, Cosine Similarity Score (or Average), or MTURK Likert Scale (or Average) of Treatment (Standardized), All Cities

	Average Age	Median Age	75th Percentile	Over 40	No Age Information
CSS	**-1.722****†††	**-1.534****††	**-3.139*****†††	**-0.083*****†††	0.008
	(0.652)	(0.637)	(0.972)	(0.023)	(0.031)
N	228	228	228	228	237
Likert score (perceived age bias)	**-2.048*****†††	**-1.524****†††	**-3.354*****†††	**-0.085*****†††	0.004
	(0.652)	(0.637)	(0.972)	(0.023)	(0.031)
N	228	228	228	228	237

Note: In the top panel, the treatment is the cosine similarity score for the arm (averaged when there were multiple treatments) with corresponding stereotype or stereotypes; the score (or average) is standardized so the table reports the estimated effects of a 1 standard deviation in the score. In the bottom panel, the treatment is average MTURK Likert score for the arm (averaged when there were multiple treatments) with corresponding stereotype or stereotypes; the score (or average) is standardized so the table reports the estimated effects of a 1 standard deviation in the score. The sign is switched from Figure 18 so that a higher value implies job-ad language perceived as more biased against older workers. Bold horizontal lines distinguish separate regressions. All specifications include fixed effects for both city and occupation. Standard errors clustered at the city-occupation level are reported in parentheses. Data are collapsed to the city-occupation-job ad level. ***, **, or * indicates statistically significant at the 1%, 5%, or 10% level in a two-sided test. †††, ††, or †indicates statistically significant at the 1%, 5%, or 10% level in a one-sided test. In the first four columns, boldface estimates indicate statistical significance at the 5% level in a two-sided tests, correcting for multiple testing across the estimates in those columns, using the Simes False Discovery Rate (q-values $\leq$ 0.05). Italicized estimates indicate q-values $>$ 0.05 and $\leq$ 0.1. (See Table A1 in online Appendix A.)

Table 29A: Estimated Effects on Age Composition of Applicants, Different Age Cutoffs, All Treatment Arms, All Cities

	Over 40	Over 50	Over 65
Communication	-0.075†	-0.073*††	-0.005
	(0.046)	(0.037)	(0.010)
Physical	-0.083*††	-0.042	-0.009
	(0.045)	(0.040)	(0.007)
Technology	-0.041	-0.038	-0.004
	(0.041)	(0.036)	(0.009)
All 3	-0.117***†††	-0.081**††	-0.006
	(0.041)	(0.034)	(0.007)
AARP	-0.156***†††	-0.090**†††	-0.008
	(0.038)	(0.035)	(0.008)
N	228	228	228

Note: All specifications include fixed effects for both city and occupation. Standard errors clustered at the city-occupation level are reported in parentheses. Data are collapsed to the city-occupation-job ad level. ***, **, or * indicates statistically significant at the 1%, 5%, or 10% level in a two-sided test. †††, ††, or †indicates statistically significant at the 1%, 5%, or 10% level in a one-sided test.

Table 29B: Estimated Effects on Age Composition of Applicants, Cosine Similarity Score (or Average), or MTURK Likert Scale (or Average) of Treatment (Standardized), Different Age Cutoffs, All Cities

	Over 40	Over 50	Over 65
CSS	-0.083***†††	-0.051***†††	-0.004
	(0.023)	(0.019)	(0.004)
N	228	228	228
Likert Score (Perceived Age Bias)	-0.085***†††	-0.040***†††	-0.004
	(0.017)	(0.014)	(0.004)
N	228	228	228

Note: In the top panel, the treatment is the cosine similarity score for the arm (averaged when there were multiple treatments) with corresponding stereotype or stereotypes; the score (or average) is standardized so the table reports the estimated effects of a 1 standard deviation in the score. In the bottom panel, the treatment is average MTURK Likert score for the arm (averaged when there were multiple treatments) with corresponding stereotype or stereotypes; the score (or average) is standardized so the table reports the estimated effects of a 1 standard deviation in the score. The sign is switched from Figure 18 so that a higher value implies job-ad language perceived as more biased against older workers. Bold horizontal lines distinguish separate regressions. All specifications include fixed effects for city, occupation, month of posting, and day-of-week of posting. Standard errors clustered at the city-occupation level are reported in parentheses. Data are collapsed to the city-occupation-job ad level. ***, **, or * indicates statistically significant at the 1%, 5%, or 10% level in a two-sided test. †††, ††, or †indicates statistically significant at the 1%, 5%, or 10% level in a one-sided test.

Table 30: Estimated Effects on Age Composition of Applicants, with Month and Day-of-Week Fixed Effects, All Treatment Arms, Cosine Similarity Score (or Average), or MTURK Likert Scale (or Average) of Treatment (Standardized), All Cities

	Average Age	Median Age	75^{th} Percentile	Over 40	No Age Information
Communication	-3.155**†††	-3.094**††	-3.291†	-0.087*††	-0.003
	(1.270)	(1.482)	(2.163)	(0.046)	(0.044)
Physical	-1.873	-1.769	-2.620	-0.087*††	0.017
	(1.492)	(1.460)	(2.249)	(0.048)	(0.049)
Technology	-2.222*††	-2.145†	-1.017	-0.056	0.046
	(1.288)	(1.293)	(2.056)	(0.048)	(0.041)
All 3	-2.383*††	-2.160*††	-3.786*††	-0.105**††	0.010
	(1.209)	(1.233)	(1.960)	(0.045)	(0.050)
AARP	-4.232***††	-3.281**†††	-5.601**†††	-0.159***†††	0.031
	(1.357)	(1.321)	(2.118)	(0.046)	(0.056)
N	228	228	228	228	237
CSS	-1.369*††	-1.083†	-2.606**††	-0.068**†††	0.010
	(0.690)	(0.688)	(1.098)	(0.026)	(0.032)
N	228	228	228	228	237
Likert Score (Perceived Age Bias)	-1.636**†††	-1.035*††	-2.941***†††	-0.080***†††	0.010
	(0.642)	(0.589)	(1.001)	(0.021)	(0.030)
N	228	228	228	228	237

Note: In the second panel, the treatment is the cosine similarity score for the arm (averaged when there were multiple treatments) with corresponding stereotype or stereotypes; the score (or average) is standardized so the table reports the estimated effects of a 1 standard deviation in the score. In the third panel, the treatment is average MTURK Likert score for the arm (averaged when there were multiple treatments) with corresponding stereotype or stereotypes; the score (or average) is standardized so the table reports the estimated effects of a 1 standard deviation in the score. The sign is switched from Figure 18 so that a higher value implies job-ad language perceived as more biased against older workers. Bold horizontal lines distinguish separate regressions. All specifications include fixed effects for city, occupation, month of posting, and day-of-week of posting. Standard errors clustered at the city-occupation level are reported in parentheses. Data are collapsed to the city-occupation-job ad level. ***, **, or * indicates statistically significant at the 1%, 5%, or 10% level in a two-sided test. †††, ††, or †indicates statistically significant at the 1%, 5%, or 10% level in a one-sided test.