Two Essays on Specialized Labor Markets

by

Yulia Chhabra

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Business Administration)
in the University of Michigan
2019

Doctoral Committee:

Adjunct Professor Margaret C. Levenstein, Co-Chair
Professor Jagadeesh Sivadasan, Co-Chair
Professor Scott E. Masten
Assistant Professor Amy Nguyen-Chyung
Professor Jason D. Owen-Smith

Yulia Chhabra

muzyrya@umich.edu

ORCID iD: 0000-0003-3643-1876

ACKNOWLEDGEMENTS

I want to thank my advisor Maggie Levenstein for her acceptance of my ideas. All committee members contributed their time and advice over the years and this dissertation is better because of it.

A separate thank you is to my mentor Pai-Ling for her optimism, my officemate Jacob for good conversations and being real, and to the Balaporias for becoming our family. Dr. Rachelle Wilcox and Katie Flanner are exceptional at their jobs and help researchers thrive.

I am most grateful to my husband who is my rock and helps be my best.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

ABSTRACT

The dissertation studies two specialized labor markets: fashion and science. The essays examine two separate topics in the context of a specialized market. The topics are hiring practice and its consequences for diversity and inclusion and the role of federal investment in a specialized market and its effect on total employment.

The first essay provides empirical evidence of race-based hiring quotas in a competitive labor market. Using novel data on the hiring of fashion models for high-end runway shows, I find that Black and Asian models are less likely to be hired if a designer retains more models of the same race from previous shows. The substitution effect between minority newcomers and incumbents of the same race is larger than the substitution effect between a random newcomer and incumbent or a newcomer and incumbent who are White. The substitution effect for minority models is also larger than the same effect derived from a simulation of race-blind hiring. The credentials of Black and Asian newcomers improve with the increase in the percentage of rehired models of the same race in a show. The findings suggest that designers choose a set number of minority models per show and become more selective of minority newcomers when they retain more models of the same race from previous shows.

In the second essay, based on the joint work with Margaret C. Levenstein and Jason Owen-Smith, I use the American Recovery and Reinvestment Act (ARRA), a large stimulus package passed into law to combat the Great Recession, to estimate the effect of R&D and science spending on local employment. Unlike most fiscal stimuli, the R&D and science portion of ARRA did not target counties with poor economic conditions but rather was awarded following a peer review process, or based on innovative potential and research infrastructure. We find that, over the program's five-year disbursement period,

each one million USD in R&D and science spending was associated with twenty-seven additional jobs. The estimated job-year cost is about \$15,000.

CHAPTER I

"We've Already Hired A Black Girl": Empirical Evidence of Quotas in the Fashion Industry

1.1 Introduction

Does an increase in minority representation mean less biased employment? A more diverse workforce is viewed by many as an indicator of fair hiring practices. Consistently, many employers face increasing pressure to raise the proportion of minorities hired. This paper suggests that this pressure creates incentives for race-based hiring. It emphasizes that increasing minority representation in the workforce is not sufficient to guarantee fair hiring practices and can generate a quota effect.

This paper empirically tests whether the race of retained workers affects the selection of a new employee. First, I examine whether the racial composition of an existing workforce changes the probability of a minority hire. If an employer has a certain number of "minority" spots in mind, the number of retained workers reduces the probability of a new hire of the same race. Then I compare my findings with race-blind hiring. Under the race-blind hiring, the number of minority incumbents in a firm has the same effect on all newcomers, regardless of race, and it does not differ from the effect of White incumbents on all newcomers. I simulate hiring patterns in a race-blind environment and compare the effect of minority incumbents on minority newcomers under two scenarios – race-blind hiring in the simulated data and hiring patterns in the fashion industry data. Finally, I examine whether the credentials of newly hired employees vary with the racial composition of the incumbent workforce. If the hiring is race-blind, then, all other things being equal, the credentials of newcomers should not vary among employers with different racial compositions of incumbent workers. If race matters in hiring, the employer with many minority incumbents has fewer "minority vacancies" and becomes more selective of minority candidates. In other words, if race matters, the variation in availability of "minority spots" drives the competition minority applicants face. Thus, the employers with few "minority spots" select minority workers with stronger credentials.

I test these predictions using a novel data set on the hiring of models for high-end fashion runway shows. The fashion industry hiring displays some unique features that allow me to address whether quota-based hiring can arise in a competitive market under non-legislative pressure from the public. The dissimilarity between the labor market for fashion models and the general labor market is a virtue because it provides an opportunity for examining the quota effect, which is harder to measure in other settings. First, hiring choices of designers for runway shows are easily observed. Second, the labor market for fashion runway shows is open and competitive. It is not subject to affirmative action legislation, and most jobs last less than a few days, making firing costs negligible. Third, education and experience are not crucial for runway hiring, negating the effect of premarket discrimination and group differences in education and experience on hiring decisions.[1] Fourth, the skills are observable, alleviating concerns about the use of race by employers as a proxy for unobservable skills. Fifth, networks, referrals, and in-group preferences of workers – all-powerful social phenomena affecting racial composition of a workforce – are less prominent in runway hiring due to large power distance between a designer and a model.[2] Modeling for a fashion runway does not involve teamwork or extensive communication among workers further reducing concerns about the effect of in-group preferences on hiring.[3] Finally, sequential recruitment and importance of relational contracts allow me to use the number of rehired minority workers as a predictor

[1] "Fashion modeling...has no formal entry criteria based on credentials such as diplomas or certificates. There are very few informal barriers like cultural or social capital..." [Mears, 2011a].

[2] Mears [2008] points in her study that "models hold the least power in this market; they have the weakest sense of judging criteria and the greatest replaceability."

[3] "Modeling, like artistic careers, consists of short-term contractual ties, in which employment is on a per-project basis" [Mears, 2008]. See Mears [2011a] for detailed description of modeling work.

for the number of vacancies open to minority newcomers.

The hiring of models for a runway show takes place in two stages: the rehiring[4] of models from previous seasons who performed well and are available for the current season, followed by the search for newcomers to fill the remaining spots. The combination of the uncertainty of customer demand, the uncertainty in the initial quality of workers, and the lack of signaling mechanisms, such as education, training, or references result in a form of path dependency: the number of models retained by a designer sets a constraint on the number of newcomers.

I examine whether, in a runway show, the number of newly hired minority models differs systematically depending on the number of retained models of the same race. This result is contrasted against a simulation of race-blind hiring in the same context. Then, I test whether the credentials of minority newcomers differ based on the percentage of rehired models of the same race in a show.

I find that the number of minority models retained from previous seasons disproportionately affects newcomers of the same race. The average reduction in Asian newcomers associated with an additional Asian incumbent is almost six times larger than a similar effect from a rehired White model. The effect of Black and Hispanic rehires on Black and Hispanic newcomers is two and a half and four times larger than the effect of an additional rehired White model. In contrast, White newcomers are affected equally by a rehired model of any race. The magnitude of this effect is similar to that of an additional White incumbent on Asian, Black, and Hispanic newcomers. Next, I simulate race-blind hiring in the same context and compare the estimates from simulated data to the main result. The main empirical result from the actual data is substantially higher than the substitution effect for minority newcomers and rehires under the simulated race-blind hiring. The difference between these results alleviates concerns about the non-replacement and other supply issues that can create a disproportionately large substitution effect between minority newcomers and incumbents. Finally, I compare the portfolios of new-

[4] I use the words "rehire" and "retain" interchangeably. Although runway contracts are show-specific, previous collaborations between a designer and a model reduce uncertainty about the model's future performance and give her an advantage over new hires [Mears, 2011a].

comers in shows with higher and lower percentages of retained models of the same race. I find that Black and Asian newcomers in the shows with high percentage of Black and Asian rehires consistently outperform their peers in the shows with low percentage of Black and Asian rehires. The former group dominates in almost all portfolio items.[5] The results for White newcomers are mixed, with no clear relationship between the average quality of a newcomer portfolio and the percentage of rehired White models in a show.

My study contributes to the personnel economics literature on organizational demographics and hiring.[6] Multiple studies tie demographic composition of the current workforce to firms' hiring choices. Dustmann et al. [2015] use the share of minority workers in a group as a proxy for referrals and show that this share is positively related to the probability of hiring from the same minority group. Oyer and Schaefer [2012] find that U.S. law firms hire disproportionately from law schools that are geographically close and have more of its alumni as partners. Giuliano et al. [2009] find that managers disproportionately hire employees of the same race. These results are due to social networks and in-group preferences and three studies find positive associations between the demographic characteristics of an existing workforce and those of new hires.

I find the opposite: an increase in the number of retained workers of a particular minority group reduces the share of new hires from that group. The discrepancy can be explained by the differences in settings and relationships between current workers and new hires. Designers do not use referrals or networks of retained models to hire new models because of the large power differences. In-group preferences are also muted because worker communication and teamwork are limited: from initial casting to runway show, models do not depend on interactions with fellow models to do their job. In addition, the difference between retained and newly hired workers is not as substantial as it is in other occupations. In the aforementioned studies, a worker has some decision-making power over new hires (he is either a hiring manager or a worker suggesting a candidate), whereas in runway hiring, a rehired model does not gain any knowledge of or authority

[5] I compare models based on the number of lookbooks (a collection of photographs representing a clothing line), editorials (artistic fashion photography in fashion magazines), magazine covers, catalogs, and advertising campaigns.

[6] For a survey of the personnel economics literature on hiring, see Oyer and Schaefer [2010].

over the hiring of her peers.[7]

This study complements the broad organizational demographics literature in sociology,[8] which focuses on the demographic composition of teams or organizational units in relation to performance, turnover, innovation potential, conflict, and so on. For example, Sørensen [2004] shows that turnover increases when same-race representation in the workplace declines. In general, this strand of literature analyzes relatively stable teams and organizational units. The decision about the addition of a new member to a team or organizational unit is either designed as random (in lab experiments) or assumed to be exogenous to the race composition of the current group. I find that this assumption does not hold in runway hiring; rather, the race of a new addition to a group depends on the racial composition of the existing workforce. Moreover, the racial composition of the existing workforce affects the credentials of the new hires (i.e., a larger number of retained minority workers raises the bar for the newcomers of the same race).

In addition to the personnel economics literature on hiring and organizational demography, this study is also related to the literature on labor market segmentation [Reich et al., 1973], the research on affirmative action as quotas [Welch, 1976], the policy evaluation of affirmative action legislation [Leonard, 1990; Oyer and Schaefer, 2002; DeLeire, 2000; Acemoglu and Angrist, 2001], and the theory of tokenism [Kanter, 1977a,b]. Similar to the studies on tokenism, I examine the ratios of minority workers in an organization and the differential treatment of a worker based on minority status. Many of the core topics in tokenism literature – including heightened visibility of a minority worker, exaggerated differences between minority and majority workers, and the assimilation of minority workers into stereotypes about their race [Kanter, 1977b] – are present in fashion. However, I focus instead on employment outcomes between minority workers who face a prospective employer with many retained workers of the same race and those who face an employer with few or no such workers.

[7]Mears [2011a] writes about the hiring process. Clients (editors, photographers, casting directors, designers, and stylists) "are charged with the selection of models for designers and fashion companies... stylists and casting directors wield considerable clout, makeup artists the least, and designers usually make the final verdict on which models to hire."

[8] Williams and O'Reilly III [1998] review field and lab studies in organizational demography carried out by sociologists.

The narrow context of the study differentiates it from the labor segmentation literature, as well. Given the average pay, job security, returns to experience, and length of tenure, modeling belongs to the secondary market for everyone, regardless of race.[9] While my findings hint at a labor market segmentation with minority models competing in small niche markets with peers of the same race, no evidence of a minority overrepresentation in fashion exists. Contrary to the predictions in the labor market segmentation literature, White workers dominate this secondary labor market.

The hiring of models for runway shows is not subject to affirmative action legislation, making it an interesting case of race-based quotas arising in response to public pressure rather than to a threat of legal action. While my findings suggest that the allocation of minority workers in fashion is reminiscent of self-enforced quotas, the evidence does not support other fundamentals used in affirmative action papers, such as the difference in skills between minority and majority applicants and the allocation of low-skilled minority applicants to high-skilled jobs. The analysis in this study focuses on a narrow subset of hires in the fashion industry that does not provide significant variation in skill.

This paper is closely related to the small number of studies on non-legislated quotas [Fryer, 2009]. The non-legislated quotas are notoriously difficult to show. Dezső et al. [2016] document patterns in top management team composition that are consistent with a quota for women but, critically, do not address non-replacement, self-selection via fertility decisions, labor supply (hours and occupational choice), skill gap - all important alternatives that can generate similar team composition. Even when data on the supply side are available, as in the recent case of Harvard v. Students for Fair Admissions, economics experts can come to different conclusions regarding the existence of quotas.[10]

Finally, quite important is the fact that market participants share a belief that designers enforce quotas for minority models.[11] Mears [2011a] and Mears [2010] present

[9] The details about the working conditions, risks, contractual arrangements, and the costs of "looking like a model" are in Chapters 2 and 3 of [Mears, 2011a].

[10] See expert reports in Arcidiacono [2018] and Card [2017].

[11] Teeman [2013] quotes Chanel Iman, a Black model, "Designers have told me, "We already found one black girl. We don't need you any more.""

Mears [2011a] quotes David, New York casting director, "There always has to be at least one because they don't want to offend any group, you know. So I always try to get one Asian, one Black, and also I think it does a service to a designer if we are trying,.."

descriptive evidence of quotas and interviews with minority models, e.g. Teeman [2013], provide insightful anecdotes.

1.2 The Fashion Industry

The fashion industry is an important part of the global economy, spanning the production of textiles, the design and manufacturing of clothes, advertising, and retail. This study focuses on a commonly known form of advertising in fashion – runway shows. A fashion show is an expensive[12] and elaborate advertising event that promotes the sales of a collection of garments designed for a brand. Runway shows take place in many cities across the world, but historically the most established and well-known fashion brands show their collections twice a year in New York, London, Milan, and Paris. Each of the four cities hosts runway shows for one week in September for Spring/Summer collections and in February for Fall/Winter collections. The month-long series of events is known as the Big Four Fashion Weeks. Traditionally, brands show their collections in one of the four cities, which is also their headquarters.[13]

The majority of fashion brands are eponymous and have had the same head designer since inception. A change of head designer is rare and usually reflects succession after a designer's death or retirement. For example, Sarah Burton, Alexander McQueen's main design aide, stepped in as the head designer for his eponymous brand after McQueen's suicide in 2010. Smaller brands tend to disappear once the designer leaves the profession. Many prolific designers head several brands; for example, Karl Lagerfeld is the head designer for Chanel, Fendi, and the Karl Lagerfeld brand, though each brand has its own unique aesthetics.

A designer is involved in many aspects of a runway show because runways generate visibility, carry their name, and are a major display of their work. The designers of smaller brands select runway models themselves. Larger brands employ casting directors

[12] The cost of a basic runway during New York Fashion Week is around US$200,000. The top brands can spend upward of US$1 million [Mau, 2014].

[13] There are some exceptions when the show is moved to a different fashion capital for cost-saving purposes or to advertise a forthcoming store opening in that city.

and supporting designers who are involved in model screening and casting. However, all personnel of large brands are working toward a common goal of implementing the head designer's vision, which leads to continuity in preferences for runway models, even if changes to the brand's team occur over time.

1.2.1 Why Fashion Models?

Runway hiring represents a small share of the total labor market and has unique features that set it apart from traditional forms of employment. I use these unique features to isolate the quota effect and shed new light on labor markets in general.

The labor market for fashion models is fast-paced and displays high churn rates. On average, modeling careers span six seasons, or three years, and are to a labor economist what fruit flies are to a geneticist. High entry-exit rates and multiple employment events within one's career generate sufficient variation for estimating hiring patterns by designers in relatively short panels.

Group differences is the fundamental problem in estimating the effect of race on hiring. Runway hiring stands out, since formal education is not required and experience is often an impediment. Few newcomers have networks that can guarantee a spot on a high-end runway, thus alleviating concerns about the effect of networks on hiring [Oyer and Schaefer, 2012]. The power distance between designers and models make referrals uncommon and not an important source for finding new employees, unlike evidence in Dustmann et al. [2015]. The setting, therefore, reduces concerns about group differences in education, experience, training, or networks.

The setting also alleviates concerns about the employer's use of race as a proxy for unobservable skills [Phelps, 1972]. The skills required for the runway modeling are observable during castings. A fashion model does not search for, schedule, or negotiate payment in any of her jobs, further reducing the importance of education, training, and employer's use of race for inference [Mears, 2011a]. Runway employment lasts, at maximum, several days. It does not carry any firing costs or long-term commitments that may affect differentially the hiring of employees of different races [DeLeire, 2000; Acemoglu and Angrist, 2001]. Furthermore, fashion runways do not require teamwork or communication among models, reducing the importance of "common language" for performance outcomes [Lang, 1986].

1.2.2 The Hiring of Models During Fashion Weeks

The hiring process differs for models who have never worked for the brand before and those who have. New hires go through several rounds of screenings, castings, and fittings before they walk the runway. First, a modeling agency sends a stack of composition cards (comp cards)[14] for all the models who have the potential to be employed by the designer. The comp cards of all newcomers are generally sent as designers are always seeking new faces and the agencies are betting on newcomers' potential (they would not be signed with an agency otherwise). As the models' careers advance, the agency reassesses their "employability" based on the industry demand for them.

After screening thousands of comp cards, designers invite selected models to castings. Some brands begin with "cattle-call" castings in which all models are invited. When models are selected, the designer contacts their agencies to "place an option"[15] and to inform the agents about follow-up castings and fittings. The options are either converted into booking or canceled in days or even hours before the show [Godart and Mears, 2009]. After multiple castings, from go-sees[16] to fit-to-confirm, the runway lineup is selected.

The hiring process differs for models who have worked for the designer before. The designer "requests"[17] a model directly from their agency, who then join at a later stage in the screening process, usually during fittings.[18]

Designers face a trade-off when selecting models for the runway between the uncertainty about any one model's performance and the quest for new faces. They turn to

[14]A composition card for a runway show includes the pictures of a model (usually a headshot and a full-length body shot) and her statistics (height, dress size, shoe size, and measurements). It also includes contact details of the agent. A comp card of Samantha Archibald from Major agency is in Figure 1.1.

[15] According to Godart and Mears [2009], an option is "a hold on the model's future availability in rank order of interest, from first(strong) to third (weak)." For more information on option mechanism in runway hiring, see pp.681-683 op.cit.

[16]A meeting with the client, not necessarily for a specific job [Mears, 2011a].

[17]An invitation for a casting extended to specific models.

[18] For details about model bookings, casting process, "requests", and "options", see Mears [2008], Chapter 3 of Mears [2011a], and Safronova et al. [2017].

the new faces when "requested" models retire, change careers, take a break for school, or start a family. The new faces also fill in the spots of models who did not do well the previous season, since they have more upward potential.

Monetary compensation[19] is not the main driver of model participation in runway show, designers pay in prestige. Aspiring models walk the runway to gain reputational capital, while established models do so to maintain their relevance and renew a relational contract with a designer, which is a prerequisite for high-paying advertising jobs. There is a large variation in pay based on model status. For instance, an established model who brings the publicity and spotlight to the runway can be paid more than US$10,000 per show, but these instances are infrequent. On average, during New York Fashion Week, big brands pay models approximately $1,000, and small brands pay $200, before taxes and agency commission (usually 20%), for all castings, fittings, and runway work [Mears, 2011b]. Designers in the other fashion capitals pay less [Mears, 2011a]. This pay scale is fairly recent, as five years ago designers did not pay models for runway work, or they paid in trade (clothes they wore on the runway) [Odell, 2013]. Some designers still do not pay, pay less than what was agreed on, or delay payments [Mears, 2011a]. The modeling agencies transfer all risks of a nonpayment to the model [Mears, 2011a], as they are in a difficult position in negotiations with the designer about model pay. The commission to the agency from a model's runway booking is small compared with the monetary and reputational value of a business relationship with the designer.

The wages paid by a designer to fashion models constitute a small percentage of overall runway costs. A simple calculation suggests that a designer holding a basic show featuring 25 models who are paid US $200 will spend less than 3% of the total runway cost of US $200,000. The combined pay of all models will be less than what a make-up artist will receive for styling models in the same show [Mau, 2014].

Fashion modeling displays a typical tournament structure, with many contenders at the bottom and few incumbents at the top in terms of earnings and prestige. The supply of new models exceeds demand, especially for the top brands. The anecdotal evidence in

[19] For details about model earnings, see Chapter 2 in Mears [2011a], Mears [2011b], Nisita [2013] and Odell [2013].

Mears [2011a], fashion forums (e.g., The Fashion Spot), and popular press suggests that designers have difficulty in choosing among many worthy candidates.

1.2.3 The Lack of Diversity on the Runway

The fashion industry is often criticized for its labor practices, including the employment of underage models, the pressure on workers to lose weight, inconsistent and low pay, and bad working conditions. This study specifically addresses the lack of diversity on fashion runways.

The fashion industry as a whole, and many designers in particular, are often criticized by the media about the lack of minority models on runways and in other forms of advertising.[20] This issue is so prominent that it spurred biannual reports about diversity on the runways during major Fashion Weeks (e.g. Tai [2017], Sauers [2013]). Twice a year, the Fashion Spot and Jezebel websites count models of color on every major runway. They then publish simple statistics and commentary about the diversity on the runway, often naming “the best” and “the worst” designers in terms of minority representation. Major news outlets often use these reports to single out designers who do not hire or hire too few minority models to walk in their runway shows, thus creating negative publicity for a fashion brand. While the intentions of the editors are good, the focus on counting minority models oversimplifies the problem and its solution. If having no fewer models than other designers makes bad press go away, the industry evolves by introducing an acceptable number of minority models per runway as an unspoken guideline.

The emphasis on numbers is echoed by industry insiders. The Council of Fashion Designers of America (CFDA), a trade association of American designers, included Diversity Guidelines in its regular address issued before the New York Fashion Week [von Furstenberg and Kolb, 2016]. Balance Diversity, a small grass-roots organization with a mission to increase diversity on fashion runways, wrote the guidelines in 2014. Balance Diversity drew the attention of industry insiders after its founder, Bethann Hardison, a former model, mailed letters to the organizers of Big Four Fashion Weeks with a list of

[20] Safronova et al. [2017], Pilkington [2007], and Moore-Karim [2016] are a few of the many articles on the issue.

brands that had hired none or few models of color for their Fashion Week runways. A year later, the organization sent a letter commending the industry and designers for their efforts in improving diversity (see Figure 1.2). Its metric of success is the increase in the count of minority models walking in shows. The Model Alliance, the Black Models Matter campaign, and other efforts to increase diversity also focus on increasing minority representation on the runway [Pilkington, 2007; Moore-Karim, 2016]. These campaigns, too, focus on targeting the brands with few minority models in their shows. The emphasis on the numbers creates incentives for hiring by the numbers, or the quota effect, without any legislation mandating it.

1.3 Data

I collect data on runway hiring from several online sources. Figure 1.3 shows the structure of the data. The main data comprise a list of models who walked in the runway shows during the Big Four Fashion Weeks from 2000 to 2017. The first source of data is the runway section of Vogue.com. The website provides complete picture galleries of designer collections shown during Fashion Weeks. A model name is tagged in the picture, listed in the caption, or both. The show locations are also included in the source code of a collection's webpage. Vogue.com has data on the runway shows held as early as 1991, but the data before 2000 is incomplete. I supplement and cross-check the data on show-model pairs from Vogue.com with the data from Models.com. Figure 1.4 shows the screenshots of the model data for the Prada Spring 2017 collection from Models.com (on the left) and Vogue.com.

The dataset represents the middle and the right tail of model distribution in terms of prestige. The aspiring models who never made it to a single show are not in the main dataset. I do not have the data on the selection at the scouting[21] or the casting stage. I collect recent data on the listings by the top modeling agencies to partially alleviate this problem. The data approximate the racial composition of a pool of aspiring models and

[21] The discovery by an agent (often on the street, in a cafe, or at a mall), followed by the model's signing of a contract with an agency.

provide a conjecture about the supply side of the modeling labor market.

The model-show pairs are supplemented with the portfolio data from the Fashion Model Directory (FMD), a collection of model profiles, similar to LinkedIn in its purpose, serving people with careers in fashion. The site lists the details of past employment and other credentials of fashion models, including physical characteristics (dress size, shoe size, height, eye color, hair color), date and place of birth, nationality, and work experience (lookbooks, covers, advertisements, catalogs, and editorials). I match the FMD data to the main data using the model's name. The names are not standardized in the raw data, so I created a cross-walk to link the two datasets. The data on agencies representing models come from the FMD and the picture captions from Vogue.com.

I collected the data on the models' race with the help of workers on Amazon Mechanical Turk (MTurk). The answers from MTurk are linked to model names by design.

1.3.1 Defining Race

Race is a complex construct. This paper employs a narrow view of race that does not include self-identification or racial attributes not evident in physical appearance. It treats race as a brief and subjective evaluation of observable physical traits by a collective of people other than the self. This view of race is limited and skewed. It is also prevalent in social interactions and important economic decisions, including hiring. A minority, in this context, is someone who is different than the rest of the workforce in the way that they look, to an employer and customer.

The setting, fashion runways, limits phenotypical traits of race even further. Little variability in height, weight, and, because I only sample women, gender leaves skin color and facial features as the only characteristics determining race in this context.[22] Yet these narrow attributes of race are sufficient to generate important differences in hiring decisions.

I do not take part in determining race; rather, the MTurk workers determine the race of all fashion models in the data. I instructed the workers to open a link with a picture

[22]Figure 1.5 illustrates this point by showing models Yue Han, Julia van Os, and Jasmine Tookes walk in the Carolina Herreira Spring/Summer 2017 runway show.

of a model and to key in race as Asian, Black, Hispanic,[23] White, or Other. The models who are visually non-White but do not look Black, Asian, or Hispanic, are classified as Other. Many models are biracial. It is up to MTurk workers to choose one of the five categories for models of two or more races.

I cross-check MTurk data using several sources. The FMD provides data on nationality, hair color, and eye color as a part of the model portfolio. I also use articles in popular press, such as the Top 50 Black Models in *Vogue Italia* [D'Angelo, 2016], to verify race.

1.3.2 Descriptive Analysis

The main data contain 113,929 appearances of a model in a runway show. The appearances span 4,329 shows by 395 brands in four locations (New York, London, Milan, and Paris) over 35 seasons and feature 3,034 unique models. The main data does not include the first season of each brand; that is, the first season is omitted from analysis because, by construction, every model in the first show is new to a brand. Because I am interested in the relationship between the number of newcomers by race and the racial composition of retained models, the inaugural shows do not provide any variation.

Table 1.1 summarizes the location of shows in the data. The largest number of shows, 1,581 (37%), take place during New York Fashion Week. New York is closely followed by Paris, with 1,345 (31%) shows. Milan and London hosted 886 (20%) and 517 (12%) shows in 2000-2017.

Table 1.2 shows the summary statistics on models' race. White models dominate the sample, accounting for 2,544 (84%) of unique names. The same statistics for Black, Asian, Hispanic, and Other models are 236 (8%), 185 (6%), 53 (2%), and 16 (1%), respectively.

Tables 1.3 and 1.4 highlight the difference in models' portfolios based on race. I regress a number of portfolio items, such as the number of magazine covers, editorials, or

[23] Hispanic is an ethnicity, not a race. Still, I separate Hispanic models from the rest of the sample. My definition of Hispanic is narrow and differs from traditional definitions based on Latin American heritage. The MTurk instructions specify that Hispanic, in this context, refers to Hispanic models who cannot be visually classified as White, Black, or Asian. For the purposes of this paper and in the eyes of a casting director and a customer White Hispanics Gisele Bündchen and Adriana Lima are White and Black Hispanics Joan Smalls and Sessilee Lopez are Black.

advertising campaigns on model's race. The general form of these regressions is:

$$P_{i,t} = Asian_i + Black_i + Hisp_i + Other_i + \tau_t + \gamma_c + \alpha_b + \varepsilon_i,$$

where $P_{i,t}$ is the number of editorials, advertising campaigns, or other items in model i's portfolio in season t, τ_t is season fixed effect, γ_c is city fixed effect, α_b is brand fixed effect, and ε_i is an error term clustered by model ID.

All regressions in Table 1.3 are at the model-season level and include city, season, and brand fixed effects, except for the last regression, which is at the model-city level and includes city fixed effects. There is no statistical difference between White and Asian models in terms of age, number of lookbooks, covers, catalogs, ads, shows per season, or the representation by a top agency.[24] Asian models have more editorials per season than White models. Black and Hispanic models have fewer editorials, covers, and shows per season and are approximately 20% less likely to be represented by a top agency than White models. Models classified as Other are younger and have fewer lookbooks, covers, catalogs, and advertising campaigns per season than White models.

Table 1.4 contains a comparison of models of different races at the model level using characteristics that are time-invariant in the data. Similar to the model-season regressions, Asian models are similar to White models on many characteristics; the only exceptions are smaller shoe and dress size,[25] more editorials over the length of their careers, and shorter average career. Black models differ from White models in that they have a smaller dress size, larger shoe size, and fewer shows during their careers. Hispanic models have fewer covers and longer careers and models classified as Other race have fewer lookbooks, covers, catalogs, and shows over the length of their careers than White models. Like Asian models, models classified as Other have shorter average careers. There are no statistically significant differences in height or rehiring rate across models of all races.

Figure 1.6 shows the percentage of appearances by White, Black, and Asian models

[24] IMG, Women, Elite, and Next are classified as top agencies. They summarily represent models in 45% of all appearances in the data. They have offices in all four fashion capitals, with the exception of Women which does not have a London office. The number of appearances managed by the next largest agency, Viva, is less than a half the number represented by Next, the smallest top agency.

[25] In fashion, smaller shoe size and dress size are considered more advantageous.

by city over time. The minority representation on the runways increases incrementally over time in all cities with the most gains take place in the mid-2000s.

Table 1.5 presents the summary statistics on race at the show level. It breaks down the count and percentage of models in the show by race and type of employment (newcomer or rehire). The statistics show high levels of turnover from one season to the next. More than half the models walking in an average show, regardless of race, are new to the brand. Fashion is a fast-paced industry preoccupied with novelty and youth, and the labor market for models reflects this. The other reasons for high turnover is a tournament structure of the labor market with few high earners and many hopefuls who are dropped by the agencies if they do not “make it” after a few seasons. The majority of shows are dominated by White models (i.e. 86% of models in an average show are White), and 11% of shows have a White-only cast. Only 20 shows in the sample have minority majority casts, all of which took place in 2008 or later.

1.4 Empirical Strategy and Results

1.4.1 Empirical Strategy

I show two main results. First, I find that the substitution effect, a negative relationship between the number of newcomers and the number of retained models of the same race, is several times larger for minority models than for White models. I proceed to show that a substitution effect of this magnitude is inconsistent with race-blind hiring. Second, I show that minority newcomers in shows with many rehired models of the same race have a stronger portfolio than newcomers in the shows with few rehired models of the same race.

Runway hiring exhibits strong path dependency, especially within a brand. When selected to walk for a brand, a model is likely to be invited to walk for the brand again. The outcome of the first appearance reduces the uncertainty about customer demand and model ability. She also skips initial screening and “cattle call” castings in the upcoming seasons. Her subsequent participation, however, is not known to the public until the next season's runway. Even if she did well on the runway the first time, there are any number of reasons why she may not be rehired. By the next season, she may have quit modeling because of career change, fertility decisions, or college. If she was successful the last season, she may have conflicts in her Fashion Weeks schedule and less-known designers may not get a chance to rebook her in the current season.

In general, the models do not know who will join them on the runway and can not adjust their labor supply accordingly. The number of models retained from previous seasons reduces the demand for new hires. I examine whether the large number of minority rehires disproportionately affects newcomers of the same race.

I estimate the effect, at a show level, as the percentage change in newcomers of a race r in response to the percentage change in rehired models of race r:

$$\frac{New_{r,b,t}}{All_{b,t}} = \beta_r \frac{Rehire_{r,b,t}}{All_{b,t}} + \sum_{i \neq r} \beta_i \frac{Rehire_{i,b,t}}{All_{b,t}} + \tau_t + \gamma_c + \alpha_b + \varepsilon_b,$$

where $New_{r,b,t}$ is the count of all new models of race r walking for brand b in season t, $All_{b,t}$ is the count of all models walking for brand b in season t, $Rehire_{r,b,t}$ is the count of all rehired models of race r walking for brand b in season t, τ_t is season fixed effect, γ_c is city fixed effect, α_b is the brand fixed effect, and ε_b is an error term clustered by brand.

1.4.2 Main Results

Table 1.6 reports the main results. The coefficients estimate the change in the number of newcomers, by race, in response to retaining an additional model from previous seasons. The retained models are counted separately for each race. I scale all the explanatory variables by the mean of the dependent variable to interpret coefficients as percentage changes and allow for comparison of estimates between regressions. The number of Asian, Black, and Hispanic newcomers decreases with an additional retained model of their own race and an additional White model. White models constitute the majority of all hires, with retained models taking 38% of all spots and the newly hired models accounting for another 49%. If a designer replaces one model in a show at random, this model is most likely White.

Because the number of spots in a show is limited, a substitution effect inevitably occurs between all models. The main result is that the size of the substitution effect varies across races. The baseline substitution rate (i.e., the change in the number of newcomers, regardless of race, in response to an additional rehire of any race) is estimated in Column 1 of Table 1.6 and equals 7%. Once I keep the total number of models in a show fixed, by adding a denominator to both sides of the equation, one rehired model takes the place of one newcomer. An average number of newcomers per show is 14.3, which translates to a 7% decrease in newcomers in response to an additional rehired model.

The estimates in the last column of Table 1.6 are very close to the baseline substitution rate in Column 1. The effect of an additional rehired model on the number of White newcomers is close to 7%, regardless of race; the effect is 7.1%, 7.6%, and 7.2% for an additional rehired Black, Hispanic, and White model, respectively. The effect is larger for non-White rehired models classified as Other, 9.2%, and smaller for rehired Asian models, 5.8%.

The estimate in Column 2 of Table 1.6 shows that an additional rehired Asian model reduces the number of newly hired Asian models by 33.5%. An additional rehired White model reduces the number of newly hired Asian models by 5.8%. The effect of a rehired White model on Asian newcomers is almost six times smaller than that of a rehired Asian model. The rehired models who are Black, Hispanic, or classified as Other, have no effect on Asian newcomers.

The estimates for Black and Hispanic newcomers exhibit a similar pattern. An additional rehired Black model reduces the number of newly hired Black models by 13.7%. The same estimate for Hispanic models is 29%. Both magnitudes are larger than the size of the effect of an additional rehired White model. This effect is 5.2% for Black newcomers and 6.7% for Hispanic newcomers. Asian and Hispanic rehires as well as rehires classified as Other have no effect on Black newcomers. Black, Asian, and non-White newcomers classified as Other have no effect on Hispanic newcomers.

Only 16 models are classified as Other in the data. The pattern for them is different. While I still observe a strong substitution effect with the rehired models of their own race, 204%, there is also a complementarity between the newcomers classified as Other and rehired Black models. This pattern may be due to designer preferences for the minority models not accounted for in the brand fixed effect.

In Table 1.7, I test the difference across coefficients from the regressions in Table 1.6. The first column in Table 1.7 lists the p-values from the tests of difference between the coefficient on Rehired Asian and coefficients on the estimates for rehired models of other races from Column 2 ("Asian") in Table 1.6. The coefficient estimate on Rehired Asian in Column 2 of Table 1.6 is statistically smaller than coefficients on Rehired Black, Rehired Hispanic, Rehired Other, and Rehired White, suggesting that the substitution effect from an additional rehired Asian model on the number of Asian newcomers is larger than the substitution effect from rehired models of any other race. Similarly, the coefficient estimate on Rehired Black from Column 3 in Table 1.6 is statistically smaller than the coefficient estimates on Rehired Asian, Rehired Hispanic, and Rehired White, but not smaller than the coefficient on the Rehired Other. The coefficient estimates from Columns 5 and 6 in Table 1.6 ("Hispanic" and "White") are not statistically different from other coefficient estimates in the respective regressions, suggesting that the effect of Rehired Hispanic and Rehired White on Hispanic and White newcomers respectively are not statistically different from the effect of the rehired models of other races. One exception is the effect of Asian rehired models on White newcomers, which is statistically smaller than the effect of rehired models of all other races on White newcomers.

Table 1.8 presents the results of the baseline regression split by city. For Asian newcomers, the substitution effect from Asian rehired models is strongest in Milan (45.3%), followed by New York (40.5%) and Paris (27.6%). For Black newcomers, the substitution effect (21.2%) from Black rehired models is only present in New York shows. Milan and Paris samples do not display a statistically significant effect of Black rehired models on Black newcomers. The London Fashion Week has the largest average number of Asian and Black models per show but the effect of Asian and Black rehired models on Asian and Black newcomers is not estimated precisely.

The regression estimates of the effect of White rehired models on Asian and Black newcomers are similar across cities and close to a baseline substitution effect of 7% in their magnitude. The effect of White rehired models on Asian newcomers is between 5.2% and 7.5%. The effect of Black rehires on Black newcomers is 3.7-6%.

Table 1.9 presents the results of the baseline regressions split by designer preferences in hiring Asian models (odd-numbered columns) and Black models (even-numbered columns). The substitution effect is significant across six of eight split sample regressions. For Black models, the substitution effect is the strongest for the designers who generally hire the fewest Black models (first quartile) - 44.8%. The effect is decreasing for the designers in the second (24%) and third quartile (22.4%). There is no effect for designers in the fourth quartile. If we look at the average number of Black models per show, the substitution effect disappears for designers who, on average, hire 2 or 3 Black models per show. For Asian models, the magnitude of substitution effect is similar across all quartiles (25.3%-34.5%). The coefficient is not significant for the first quartile of designers - those hiring, on average, none or one Asian model per show. The lack of statistical significance of this coefficient may be caused by a large number of designers who do not hire any Asian models for their shows in this quartile.

1.4.3 Race-Blind Hiring

I simulate race-blind hiring in the same context to compare the main result to the substitution effect under race-blind hiring. I collected data on the model listings from the London, Milan, New York, and Paris offices of the top four modeling agencies. $N_{a,c,r}$ is the number of models of race r listed on the website of a top modeling agency a in their office in city c. A pool $P_{c,r}$ of models of race r in city c is approximated by $\frac{N_{c,r}}{\psi_c}$,[26] where ψ_c is the percentage of top agency models in the runway data. Each city has D_c designers who show their collections for 35 seasons. In the first season, $t = 1$, a designer d chooses $M_{c,d,1}$ models to walk in the runway from the distribution M_c. Each designer chooses models randomly from the pool, P_c,[27] of aspiring models in a city c without observing their race. In the first season, the probability of a selection, $\phi_{c,1,n}$, by designer

[26] $N_{c,r} = \Sigma_a N_{a,c,r}$
[27] $P_c = \Sigma_r P_{c,r}$

d is the same for all models in a city c and equals $1/P_c$. In subsequent periods, the hiring probability, $\phi_{c,t,o}$, is increased by the rehiring rate δ_d for the models hired by designer d previously. The hiring rate for the models who have never been hired by the designer before, $\phi_{c,t,n}$, is adjusted to make the hiring probabilities of all models sum up to one for every designer in every period. Every period rehired models exit and enter a pool of aspiring models in a city c at a rate ζ_c.[28] $N_{a,c,r}$ comes from top agency listings data. ψ_c comes from model agency data on Vogue.com and the FMD. D_c, M_c, δ_d, ζ_c come from the main data on model-show pairs from Vogue.com and Models.com.

I construct one hundred random samples under the condition of race-blind hiring and run baseline specification. I sort the coefficient estimates and plot them in Figure 1.7 against the main result from Table 1.6. The coefficient estimates in Table 1.6 are inconsistent with race-blind hiring. The most striking result is on the graph in the upper left corner of Figure 1.7. It shows a large gap in the substitution effect between Asian newcomers and Asian rehires estimated in the main data, -33.5%, and the estimates of the same effect from the race-blind simulation. The smallest coefficient from the simulation does not reach the dotted line suggesting that the data reject the hypothesis that the substitution effect between Asian newcomers and Asian rehires arises from race-blind hiring. The substitution effect between Black newcomers and Black rehires is shown on the graph in the center. Only one draw out of one hundred has an estimate below -13.7%, a substitution effect from the data. Most of the other coefficient estimates also do not fall in line with the race-blind hiring. The substitution effects between Asian newcomers and Black rehires, Black newcomers and Asian rehires, and Black newcomers and White rehires are smaller than expected under race-blind hiring. The effect of White rehires on Asian and White newcomers are consistent with race-blind hiring.

The averages of coefficients plotted in Figure 1.7 are in Table 1.10. The substitution effect between Asian newcomers and Asian rehires is 11.4%, compared to the main result of 33.5%. The same effect for Black models is 10%, compared to the main result of 13.7%.

[28] The entry-exit rate of newcomers is irrelevant in this setting because there is no observable difference between a model who entered the market this year and was not hired by any designer and a model who entered the market the previous year but also was not hired by any designer.

In addition, I observe a substitution effect of 6.7-6.8% between Asian newcomers and Black incumbents and Asian incumbents and Black newcomers. There is no statistically significant substitution effect among these groups of models in the baseline specification. All the substitution effects with White models, incumbents and newcomers, are similar to the baseline results and range between 6.5% and 6.8%.

Under the condition of race-blind hiring, simulated data generate on average 0.66 Asian newcomers per show (a decrease of 19%), 1.34 Black newcomers per show (an increase of 35%), and 12.77 White newcomers per show (an increase of 4%). If the pool of aspiring models from the top four agencies is representative of the race composition of all aspiring models, these numbers suggest an oversupply of Black newcomers and an undersupply of Asian newcomers.

1.4.4 Newcomers' Portfolio Comparison

The main results indicate that minority newcomers are less likely to be hired if designers retain many models of the same race from previous seasons. However, some newcomers do get hired, even if a show features many rehired models of the same race. Are these models different? I estimate whether portfolio characteristics of newcomers change in response to the percentage increase in the rehired models of the same race in a show. The general form of these regressions is:

$$P_{i,t,r} = \alpha_r + \beta_r \frac{Rehire_{r,b,t}}{All_{b,t}} + \varepsilon_{b,t},$$

where $P_{i,t}$ is the number of editorials, advertising campaigns, or other items in model i's portfolio in season t, and $\varepsilon_{b,t}$ is an error term clustered by show (brand-season).

Table 1.11 shows that the quality of portfolios of Asian newcomers improves as the percentage of Asian rehires in a show increases. Asian newcomers in the shows with higher percentage of Asian rehires have more editorials and advertising campaigns in the last six months; and have more lookbooks, editorials, and advertisements over the length of their careers than Asian newcomers in the shows with a lower percentage of rehired Asian models. Other coefficient estimates are positive but not statistically significant, except for the number of catalogs in the past six months, which is negative but not statistically significant. Altogether, these results suggest that Asian newcomers in the shows with a higher percentage of Asian rehires have better portfolios than Asian newcomers in shows with a lower percentage of Asian rehires.

The results are similar for Black newcomers. Table 1.12 shows that Black newcomers in the shows with a higher percentage of Black rehires have more editorials and lookbooks in the last six months; and have more lookbooks, editorials, catalogues, and advertisements over the length of their careers than Black newcomers in shows with a lower percentage of rehired Black models. Other coefficient estimates are positive but not statistically significant. Similarly to Asian models, Black newcomers in the shows with a higher percentage of rehires of the same race have better portfolios than Black newcomers in shows with a lower percentage of rehires of the same race.

Table 1.13 shows the estimates for White newcomers, and the results are mixed. White newcomers in the shows with a higher percentage of retained White models have more covers and advertisements in the past six months and career-wise, but also have fewer lookbooks and editorials in the last six months and career-wise, and fewer catalogs over the course of their career. The results for White newcomers suggest that there is no clear relationship between the quality of a portfolio and the percentage of rehired models of the same race.

1.5 Conclusion

Many industries face public pressure to reduce race and gender gaps in hiring. Because the hiring process is hard to monitor, attention is shifted to its outcomes, which are easily observed. It is common to judge the fairness in hiring by the racial composition of an existing workforce. In this paper, I argue that the focus on counting minority hires creates incentives for quota-based hiring, even if legal enforcement or guidelines are absent.

I use data on high-end runway hiring to examine whether employers use quotas to achieve a desired racial composition of their workforce. This unique setting, which is not constrained by long-term contracts, teamwork, firing costs, worker interaction, or networks, but is very visible and under constant public scrutiny, demonstrates how quotas emerge in a competitive labor market. The results show that designers revert to a fixed number of minority hires per show. If they retain more minority models from previous shows, newcomers of the same race are either not hired at all or have stronger credentials than if they are hired for a show with fewer or no rehires of their race.

Rather than undermining the notion that the fashion industry needs a more diverse workforce, this study highlights that current diversity efforts are misguided. Current hiring practices force minority models into niche labor markets, in which they compete in minor leagues of their own. One may think that smaller markets are not necessarily more competitive. The results of simulation suggest that there is an undersupply of Asian models and an oversupply of Black models, making the niche market for Asian models less competitive and niche market for Black models more competitive in comparison to the larger market for White models.

More diversity on the runway achieved by reserving "minority" spots increases the labor supply among young minority women because they see an opportunity to "make it" in a market that was previously sealed off to them. However, they do not necessarily compete in the same market as White models. The quotas lock them into a smaller, niche market, where they compete among themselves for the few "minority" spots on a runway, because they are less "comparable" to the White models.

More uncertain demand faced by minority workers is another implication of quota-based hiring. Because a worker does not know whether the quota has been filled or not, she spends effort on getting a job with an employer who does not consider hiring her.

The findings herein can be extended to the entertainment industry (music, television, film) and advertising, as appearance plays a role in hiring decisions in those industries as well. In addition, the implications of the results in this paper are especially important in occupations which face public scrutiny about minority representation. The results serve as a caution to many industries that face criticism about diversity, including the tech industry and science. If employers practice race-based or gender-based hiring to fend off criticism about the lack of diversity in the workforce, they are negating the main principle that underlies the calls for diversity – an unbiased and fair hiring process.

Figures and Tables

Figure 1.1: A Composition Card

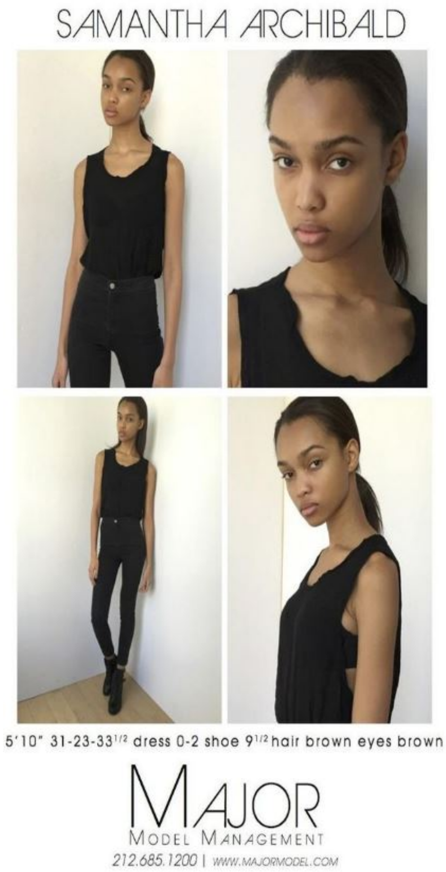


Figure 1.2: Balance Diversity letters to the CFDA

September 4, 2013

Council of Fashion Designers of America
65 Bleecker, Floor 11
New York, NY 10012

To: Diane Von Furstenberg/Steven Kolb

Eyes are on an industry that season after season watches fashion design houses consistently use of one or no models of color.

No matter the intention, the result is racism.

Not accepting another based on the color of their skin is clearly beyond "aesthetic" when it is consistent with the designer's brand.

Whether it's the decision of the designer, stylist or casting director, that decision to use basically all white models, reveals a trait that is unbecoming to modern society.

It can no longer be accepted, nor confused by the use of the Asian model.

As of last season, below are the fashion houses guilty of this racist act:

NEW YORK WOMEN

BCBG MAX AZRIA	MM6 Maison Martin	Rag & Bone
Belstaff	Margiela	Reed Krakoff
Calvin Klein Collection	Monique Lhuillier	Rodarte
Diesel Black Gold	Narciso Rodriguez	The Row
Donna Karan	Philosophy	Victoria Beckham
Marc by Marc Jacobs	Proenza Schouler	Victoria, Victoria Beckham

NEW YORK MEN

Duckie Brown	Robert Geller
N. Hollywood	Siki Im
Patrik Ervell	Tim Coppens
Perry Ellis by Duckie Brown	

From: The Coalition

February 4, 2014

Council of Fashion Designers of America
65 Bleecker, Floor 11
New York, NY 10012

Dear Diane Von Furstenberg/Steven Kolb:

Last season we addressed the international fashion industry for their lack of conduct in being racially diverse. There was a marked improvement on the runways and a positive response to the letters received by the major fashion councils and the designer brands they count as members. First we share the results:

NEW YORK

BCBG MAX Azria	**2** models of color from **1** pre-September 2013
Calvin Klein Collection	**5** models of color from **0** pre-September 2013
Diesel Black Gold	**2** models of color from **1** pre-September 2013
Donna Karan	**2** models of color from **1** pre-September 2013
Marc by Marc Jacobs	**2** models of color from **1** pre-September 2013
MM6 Maison Martin Margiela	**0** models of color from **0** pre-September 2013
Monique Lhuillier	**1** models of color from **0** pre-September 2013
Narciso Rodriguez	**4** models of color from **0** pre-September 2013
Philosophy	**2** models of color from **1** pre-September 2013
Proenza Schouler	**3** models of color from **0** pre-September 2013
Rag & Bone	**3** models of color from **1** pre-September 2013
Reed Krakoff	**3** models of color from **1** pre-September 2013
Rodarte	**3** models of color from **1** pre-September 2013
The Row	**4** models of color from **0** pre-September 2013
Victoria Beckham	**1** models of color from **0** pre-September 2013
Victoria, Victoria Beckham	**NO SHOW** **0** pre-September 2013

It is important to say that there are design houses serviced by casting directors and stylists who are latent, as they seem comfortable with stereotypical images. Although progress was made last season within certain houses, the objective is to continue this improvement across the entire industry. We look for consistency and not because of advocacy or a season lending to darker skin.

So we will continue to watch and reveal season to season.

Diversifying is not difficult. The resistance to do so is intriguing.

-Balance Diversity

Figure 1.3: Data Structure

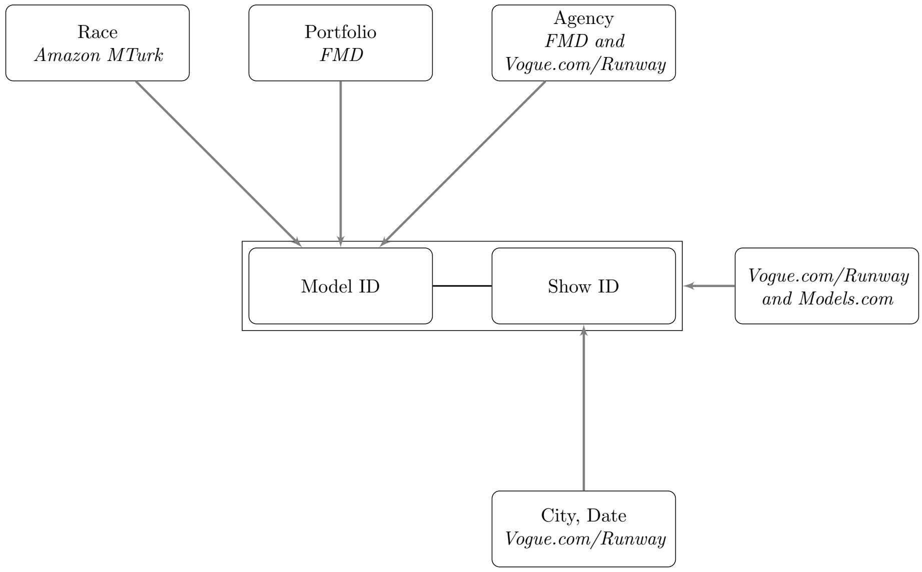
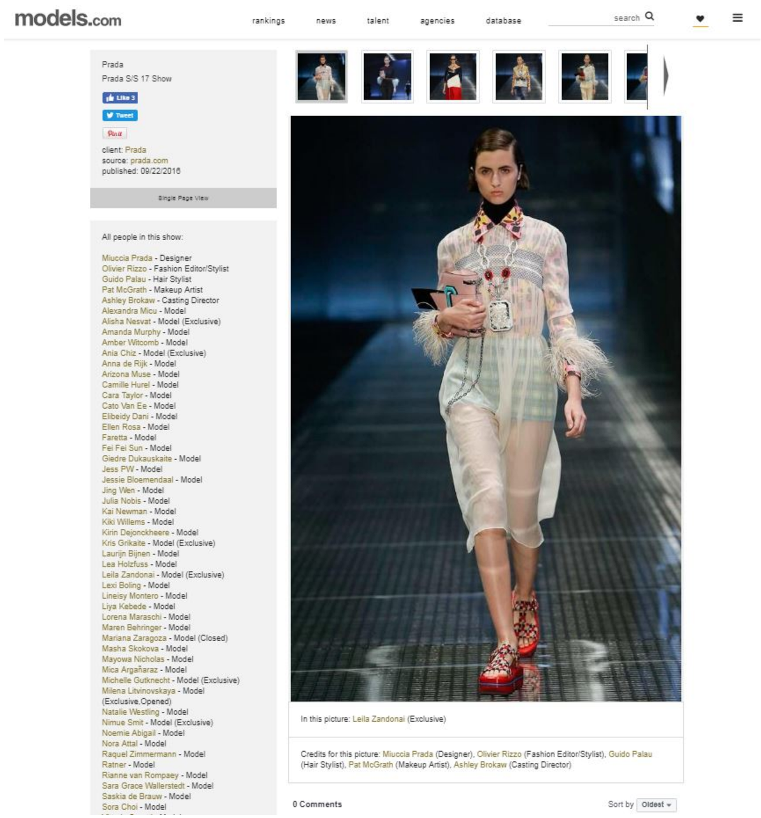


Figure 1.4: Data screenshots from Model.com (on the left) and Vogue.com



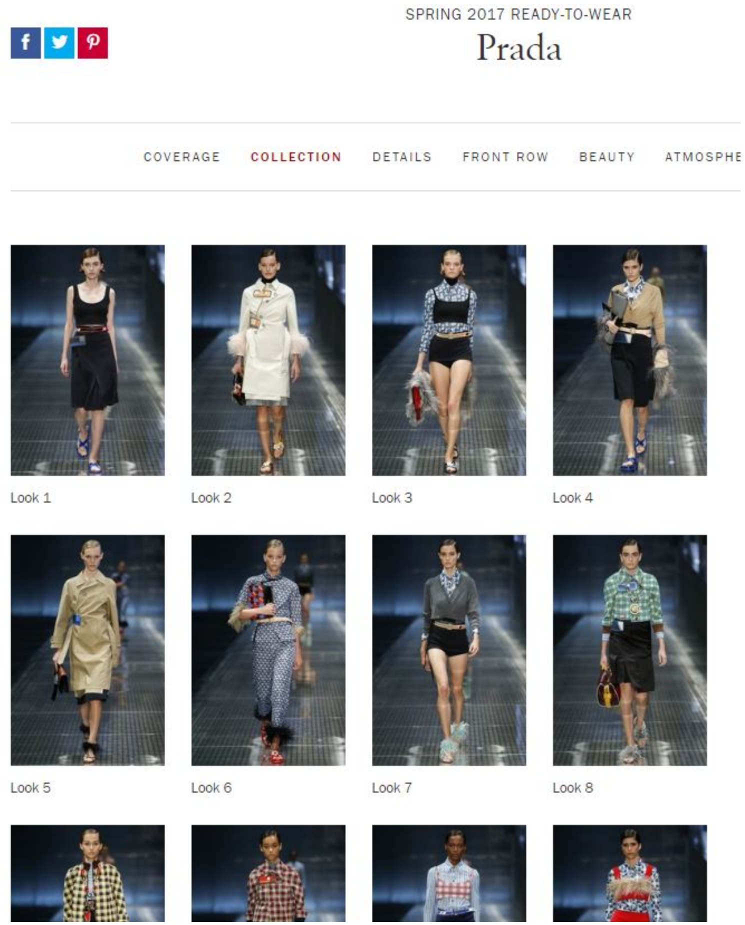
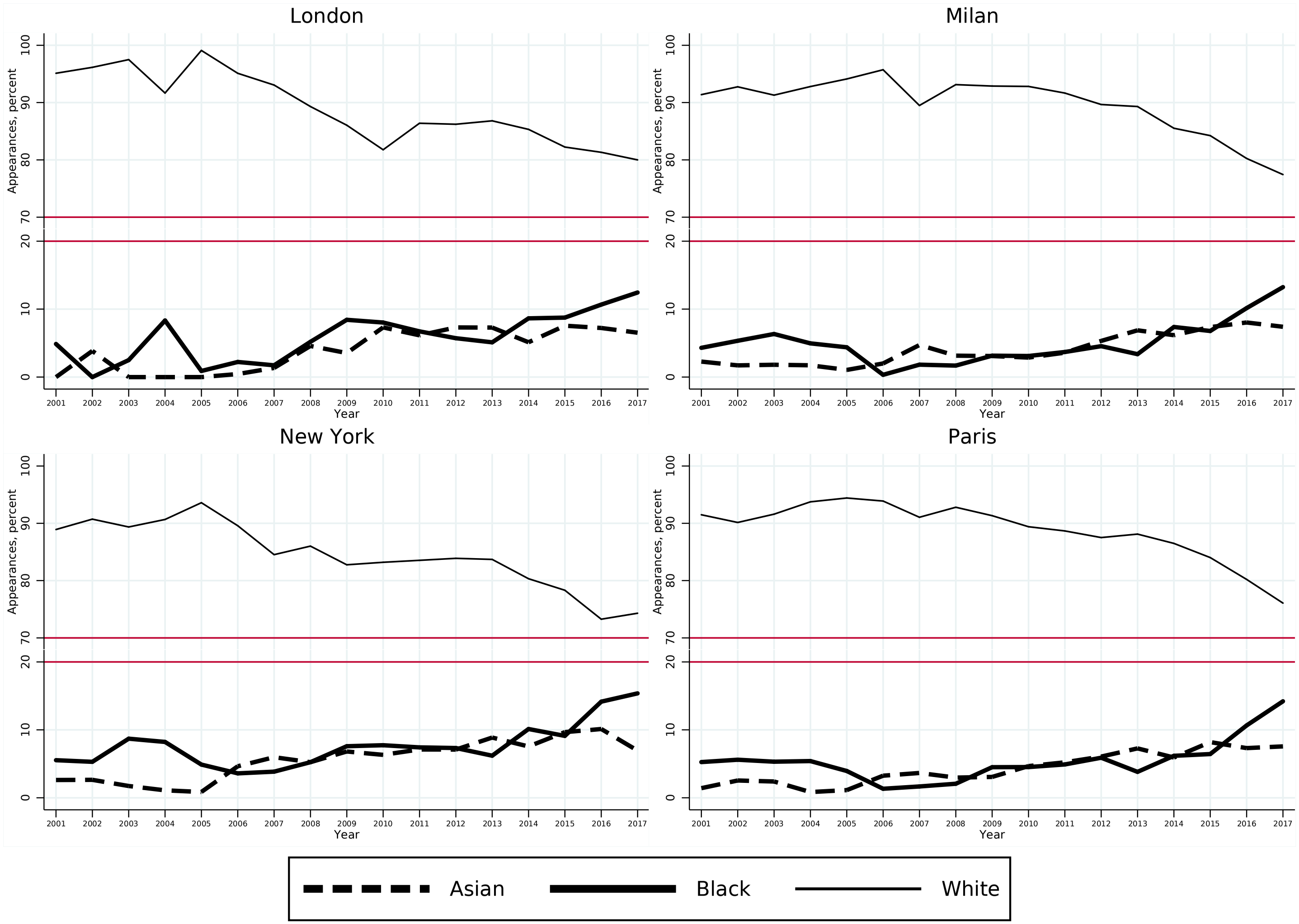


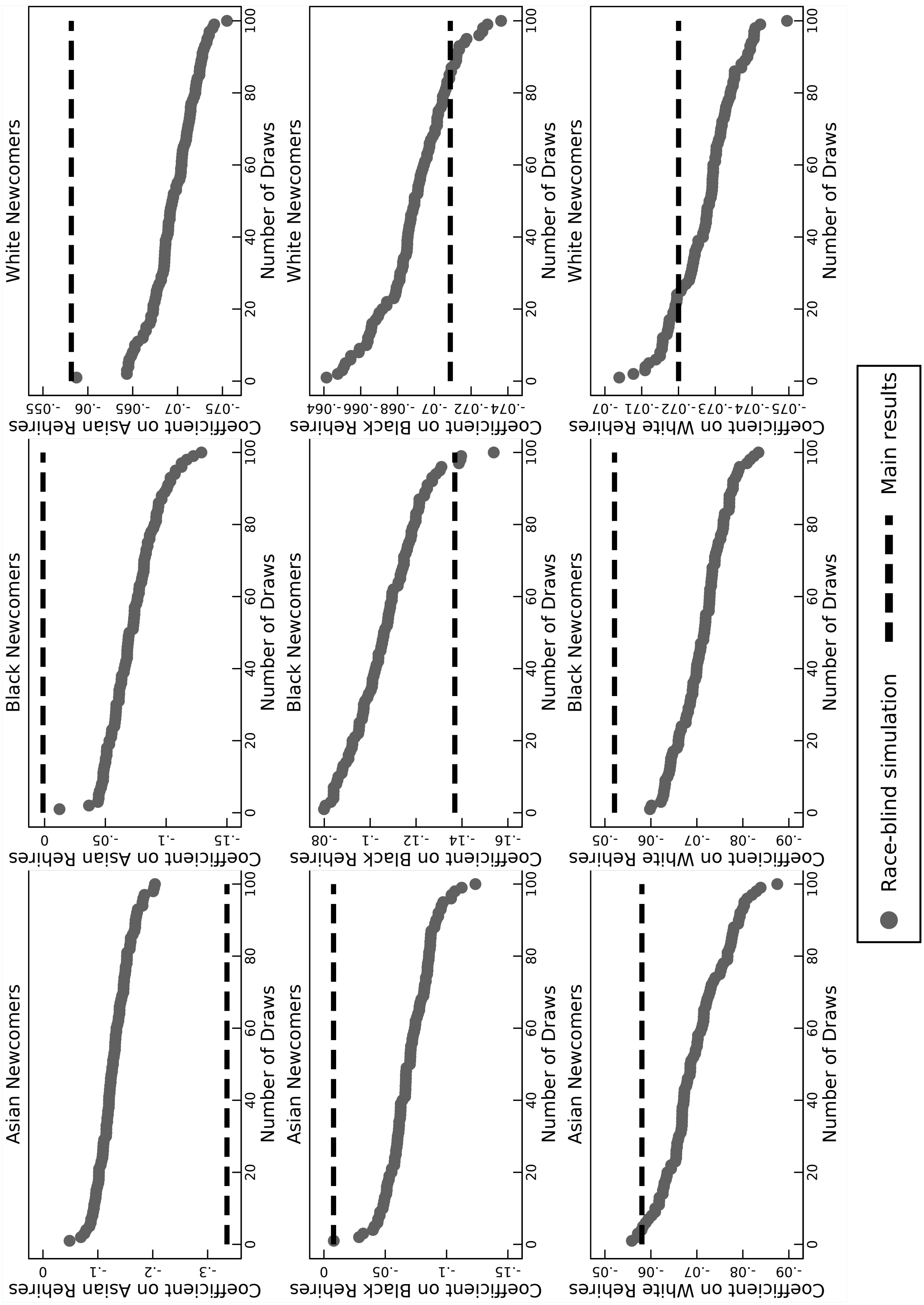
Figure 1.5: Models walk in the Carolina Herreira Spring 2017 Runway Show

Figure 1.6: Model appearances during Fashion Weeks in 2001-2017 by race



Notes: This figure presents race statistics, separately for each location of the Big Four Fashion Weeks, held twice a year, from 2001 to 2017. The unit of analysis is a model appearance in a fashion show. The sample is split by race and ethnicity. Each point on the graph represents the percentage of all appearances by models of a given race in a given year. I combine the data for Spring/Summer and Fall/Winter seasons in a given year. The values for Asian, Black, and White models are close but do not sum up to 100%. The models classified as Hispanic or Other are not shown. Asian models include East Asians, South-East Asians, and South Asians. White Hispanics are included in the White group. Black Hispanics are included in the Black group. Non-White and non-Black Hispanics are a separate group, Hispanic. Other includes visually non-White models (e.g., Middle Eastern, Native American, Pacific Islander) who are not classified elsewhere.

Figure 1.7: Simulation Results



Notes: This figure shows that hiring patterns in the data are inconsistent with race-blind hiring. The gray dots are estimated from the baseline regressions on the simulated data and sorted by size. They are plotted against the regression coefficients from the main regression in Table 1.6 shown as a dotted line. The outcome variable is newcomers of a given race, as a percentage of all hires. The new hires are defined as models who have never walked a runway for the brand before. The explanatory variables are the rehired models of different races and ethnicities, as a percentage of all hires in a show. They are scaled by the mean of dependent variable. The rehired models have walked in a runway for the brand before. Asian models include East Asians, South-East Asians, and South Asians.

Table 1.1: Show-level City Statistics

	Count	Percent
London	517	0.12
Milan	886	0.20
New York	1581	0.37
Paris	1345	0.31
Observations	4329	

Notes: This table shows the summary statistics on the show location in the main data. The unit of analysis is a show.

Table 1.2: Model-level Race Statistics

	Count	Percent
Asian	185	0.06
Black	236	0.08
Hispanic	53	0.02
Other	16	0.01
White	2544	0.84
Observations	3034	

Notes: This table shows the summary statistics on the race of models in the main data. The unit of analysis is a model. Asian models include East Asians, South-East Asians, and South Asians. White Hispanics are included in the White group. Black Hispanics are included in the Black group. Non-White and non-Black Hispanics are a separate group, Hispanic. Other includes visually non-White models (e.g., Middle Eastern, Native American, Pacific Islander) who are not classified elsewhere.

Table 1.3: Basic comparison regressions

	Age	Lookbooks, season	Editorials, season	Covers, season	Catalogs, season	Ads, season	Shows, season	Top agency
Asian	0.236	0.00989	0.345**	0.0613	-0.00141	0.0206	-0.309	-0.0311
	(0.520)	(0.0103)	(0.149)	(0.0668)	(0.00797)	(0.0495)	(0.639)	(0.0306)
Black	0.370	-0.0120	-0.216**	-0.183***	-0.00173	-0.0440	-1.705***	-0.0633**
	(0.765)	(0.00785)	(0.109)	(0.0706)	(0.00750)	(0.0429)	(0.434)	(0.0287)
Hispanic	-0.941	0.0161	-0.374**	-0.262***	-0.00812	-0.0535	-1.579***	-0.0752*
	(0.890)	(0.0219)	(0.170)	(0.0747)	(0.0137)	(0.0702)	(0.551)	(0.0455)
Other	-1.860**	-0.0308***	0.189	-0.224**	-0.0325***	-0.111**	-0.404	-0.00942
	(0.753)	(0.00969)	(0.269)	(0.0920)	(0.00870)	(0.0564)	(0.864)	(0.108)
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Season FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Brand FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Mean DV	22.73	0.03	0.99	0.41	0.02	0.26	8.21	0.32
Observations	7166	12322	12322	12322	12322	12322	13872	8322
R-sq	0.33	0.06	0.30	0.13	0.05	0.19	0.45	0.03

Notes: This table compares fashion models of different races and ethnicities. The unit of observation is model-season in all regressions, except for the last column. Top agency regression has model-city as a unit of observation. Season represents the six months before a Fashion Week in all regressions, except for Shows regression. In Shows regression, season is four weeks of Big Four Fashion Weeks held twice a year. Asian models include East Asians, South-East Asians, and South Asians. White Hispanics are included in the White group. Black Hispanics are included in the Black group. Non-White and non-Black Hispanics are a separate group, Hispanic. Other includes visually non-White models (e.g., Middle Eastern, Native American, Pacific Islander) who are not classified elsewhere. All regressions are estimated using OLS.
Standard errors in parentheses are clustered by model ID. * p<0.1, ** p<0.05, *** p<0.01.

Table 1.4: Basic comparison regressions (cont'd)

	Dress size	Shoe size	Height, cm	Lookbooks, career	Editorials, career	Covers, career	Catalogs, career	Ads, career	Shows, career	Career length	Rehire rate
Asian	-0.181*	-0.186***	-0.0719	0.0503	2.589*	-0.592	0.000835	-0.0766	-3.814	-1.239***	-0.00188
	(0.110)	(0.0667)	(0.202)	(0.0748)	(1.464)	(0.520)	(0.0599)	(0.382)	(5.240)	(0.380)	(0.0182)
Black	-0.201*	0.303***	0.170	-0.0669	-0.160	-0.826	0.00260	-0.107	-8.128**	0.0930	-0.0230
	(0.116)	(0.0581)	(0.166)	(0.0572)	(0.972)	(0.745)	(0.0568)	(0.342)	(3.601)	(0.462)	(0.0152)
Hispanic	0.218	-0.240	-0.175	0.0476	-2.030	-1.536**	0.0171	-0.377	-8.296	1.809*	0.0228
	(0.237)	(0.156)	(0.303)	(0.161)	(1.431)	(0.607)	(0.0962)	(0.527)	(7.338)	(1.096)	(0.0356)
Other	0.0794	-0.0731	0.00595	-0.227***	2.152	-1.854***	-0.158***	-0.395	-14.32*	-1.669*	-0.0122
	(0.323)	(0.220)	(0.357)	(0.0290)	(3.252)	(0.468)	(0.0252)	(0.767)	(8.248)	(0.939)	(0.0427)
Constant	3.837***	8.656***	178.0***	0.227***	6.030***	3.036***	0.158***	1.577***	38.64***	5.606***	0.208***
	(0.0282)	(0.0216)	(0.0463)	(0.0290)	(0.281)	(0.238)	(0.0252)	(0.110)	(1.441)	(0.132)	(0.00451)
Mean DV	3.82	8.66	178.00	0.23	6.13	2.92	0.16	1.56	37.55	5.56	0.21
Observations	2433	2318	2435	2467	2467	2467	2467	2467	3034	3034	3034
R-sq	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Notes: This table compares fashion models of different races and ethnicities. The unit of observation is model. Asian models include East Asians, South-East Asians, and South Asians. White Hispanics are included in the White group. Black Hispanics are included in the Black group. Non-White and non-Black Hispanics are a separate group, Hispanic. Other includes visually non-White models (e.g., Middle Eastern, Native American, Pacific Islander) who are not classified elsewhere. All regressions are estimated using OLS. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 1.5: Show-level Race Statistics

	mean	sd	min	max
All hires, count				
Asian	1.49	1.47	0	11
Black	1.66	1.91	0	23
Hispanic	0.37	0.67	0	6
Other	0.09	0.30	0	3
White	22.71	9.04	0	80
All hires, percentage points				
Asian	0.06	0.05	0.00	0.36
Black	0.06	0.07	0.00	0.95
Hispanic	0.01	0.03	0.00	0.26
Other	0.00	0.01	0.00	0.15
White	0.86	0.10	0.00	1.00
New hires, count				
Asian	0.82	0.99	0	7
Black	0.99	1.32	0	17
Hispanic	0.21	0.49	0	4
Other	0.06	0.25	0	2
White	12.24	5.31	0	49
New hires, percentage points				
Asian	0.03	0.04	0.00	0.36
Black	0.04	0.05	0.00	0.89
Hispanic	0.01	0.02	0.00	0.21
Other	0.00	0.01	0.00	0.15
White	0.49	0.18	0.00	1.00
Rehired, count				
Asian	0.67	0.99	0	8
Black	0.67	1.05	0	10
Hispanic	0.16	0.43	0	5
Other	0.03	0.17	0	2
White	10.47	7.42	0	56
Rehired, percentage points				
Asian	0.02	0.03	0.00	0.21
Black	0.02	0.04	0.00	0.30
Hispanic	0.01	0.02	0.00	0.22
Other	0.00	0.01	0.00	0.10
White	0.38	0.19	0.00	0.93
Observations	4329			

Notes: This table presents race statistics for the full sample and separately for newly hired and rehired models. The unit of analysis is a fashion show. The sample includes all major runway shows from the Big Four (New York, Milan, Paris, and London) Fashion Weeks, held twice a year, from 2000 to 2017. The new hires are the models who have never walked a runway for the brand before. The rehired models have walked in a runway for the brand before. The sample is split by race and ethnicity. Asian models include East Asians, South-East Asians, and South Asians. White Hispanics are included as White. Black Hispanics are included as Black. Non-White and non-Black Hispanics are a separate group, Hispanic. Other includes visually non-White models (e.g., Middle Eastern, Native American, Pacific Islander) who are not classified elsewhere.

Table 1.6: Brand-level regressions

New hires	All Races	Asian	Black	Hispanic	Other	White
Rehired All Races	-0.0698***					
	(2.14e-10)					
Rehired Asian		-0.335***	0.00132	-0.0581	-0.0500	-0.0582***
		(0.0388)	(0.0275)	(0.0606)	(0.101)	(0.00342)
Rehired Black		-0.00783	-0.137***	-0.0395	0.287**	-0.0709***
		(0.0248)	(0.0407)	(0.0539)	(0.124)	(0.00404)
Rehired Hispanic		-0.0320	0.00179	-0.290*	0.266	-0.0760***
		(0.0466)	(0.0449)	(0.158)	(0.167)	(0.00545)
Rehired Other		0.185	0.0715	0.157	-2.043***	-0.0923***
		(0.154)	(0.135)	(0.195)	(0.569)	(0.0160)
Rehired White		-0.0580***	-0.0521***	-0.0670***	-0.0891***	-0.0720***
		(0.00533)	(0.00613)	(0.0133)	(0.0199)	(0.000711)
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Season FE	Yes	Yes	Yes	Yes	Yes	Yes
Brand FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4329	4329	4329	4329	4329	4329
R-sq	1.00	0.40	0.41	0.26	0.24	0.91

Notes: This table shows the relationship between the number of rehired models and the number of new hires, by race. For example, the first coefficient in the second column estimates the change in Asian newcomers in response to the rehiring of an additional Asian model. The unit of observation is a fashion show. The sample includes all major runway shows from the Big Four (New York, Milan, Paris, and London) Fashion Weeks, held twice a year, from 2000 to 2017.

The outcome variable is newcomers of a given race, as a percentage of all hires. The new hires are defined as models who have never walked a runway for the brand before. The explanatory variables are the rehired models of different races and ethnicities, as a percentage of all hires in a show. They are scaled by the mean of the dependent variable. The rehired models have walked in a runway for the brand before. Asian models include East Asians, South-East Asians, and South Asians. White Hispanics are included in the White group. Black Hispanics are included in the Black group. Non-White and non-Black Hispanics are a separate group, Hispanic. Other includes visually non-White models (e.g., Middle Eastern, Native American, Pacific Islander) who are not classified elsewhere.

All regressions include city, season, and brand fixed effects. For example, Chanel Spring/Summer 2001 runway show in Paris switches on indicators for the city of Paris, the season of Spring/Summer 2001, and the Chanel brand.

All regressions are estimated using OLS. Standard errors, clustered on brand, are reported in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 1.7: Comparison of coefficients tests (p-values)

New hires (% of all hires)	Asian	Black	Hispanic	Other	White
	$\beta_a =$	$\beta_b =$	$\beta_h =$	$\beta_o =$	$\beta_w =$
β_a		0.00	0.17	0.00	0.00
β_b	0.00		0.15	0.00	0.79
β_h	0.00	0.03		0.00	0.46
β_o	0.00	0.16	0.07		0.21
β_w	0.00	0.05	0.15	0.00	

Notes: This table shows the p-values from the tests that compare coefficients from the regressions in Columns 2-6 in Table 1.6. The comparison is within regression. The empty cell indicates the coefficient of interest, which all other coefficients in regression are compared to. For example, the p-value in the second row of the first column ("Asian") tests the difference between coefficients on Rehired Asian and Rehired Black from Column 2 ("Asian") in Table 1.6.

Table 1.8: Brand-level regressions, by city

	London		Milan		New York		Paris	
New hires	Asian	Black	Asian	Black	Asian	Black	Asian	Black
Rehired Asian	-0.114	0.0613	-0.453***	-0.0460	-0.405***	-0.0248	-0.276***	0.0245
	(0.132)	(0.0960)	(0.0611)	(0.119)	(0.0608)	(0.0394)	(0.0583)	(0.0491)
Rehired Black	-0.0979	0.00153	0.0561	-0.0579	-0.0147	-0.212***	0.0140	-0.0748
	(0.0990)	(0.199)	(0.0611)	(0.115)	(0.0365)	(0.0580)	(0.0418)	(0.0525)
Rehired Hispanic	0.0313	-0.196	-0.193	0.140	0.0169	-0.0177	-0.110	0.0441
	(0.154)	(0.158)	(0.117)	(0.107)	(0.0577)	(0.0560)	(0.0926)	(0.103)
Rehired Other	0.446	0.171	0.174	-0.0771	-0.213	0.0684	0.0826	-0.245
	(0.302)	(0.281)	(0.540)	(0.268)	(0.265)	(0.204)	(0.251)	(0.158)
Rehired White	-0.0552**	-0.0366*	-0.0750***	-0.0597***	-0.0518***	-0.0543***	-0.0525***	-0.0529***
	(0.0236)	(0.0193)	(0.0112)	(0.0165)	(0.00932)	(0.0115)	(0.00885)	(0.0122)
Season FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Brand FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	517	517	886	886	1581	1581	1345	1345
R-sq	0.37	0.43	0.42	0.32	0.43	0.45	0.44	0.42

Notes: This table shows the relationship between the number of rehired models and the number of new hires, by city and race. For example, the first coefficient in the first column estimates the change in Asian newcomers in response to the rehiring of an additional Asian model during London Fashion Week. The unit of observation is a fashion show. The sample includes all major runway shows from the Big Four (New York, Milan, Paris, and London) Fashion Weeks, held twice a year, from 2000 to 2017.

The outcome variable is newcomers of a given race, as the percentage of all hires. The new hires are defined as models who have never walked a runway for the brand before. The explanatory variables are the rehired models of different races and ethnicities, as the percentage of all hires in a show. They are scaled by the mean of the dependent variable. The rehired models have walked in a runway for the brand before. Asian models include East Asians, South-East Asians, and South Asians. White Hispanics are included in the White group. Black Hispanics are included in the Black group. Non-White and non-Black Hispanics are a separate group, Hispanic. Other includes visually non-White models (e.g., Middle Eastern, Native American, Pacific Islander) who are not classified elsewhere.

All regressions include season and brand fixed effects. For example, Chanel Spring/Summer 2001 runway show switches on indicators for the season of Spring/Summer 2001, and the Chanel brand.

All regressions are estimated using OLS. Standard errors, clustered on brand, are reported in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 1.9: Brand-level regressions, by brand's number of Asian/Black models

	First Quartile		Second Quartile		Third Quartile		Fourth Quartile	
New hires	Asian	Black	Asian	Black	Asian	Black	Asian	Black
Rehired Asian	-0.280	0.127	-0.286***	0.00634	-0.345***	-0.0331	-0.253***	0.00166
	(0.211)	(0.125)	(0.0902)	(0.0729)	(0.0467)	(0.0482)	(0.0483)	(0.0318)
Rehired Black	-0.00419	-0.448***	-0.00636	-0.240***	-0.00584	-0.224***	-0.00324	-0.0504
	(0.0865)	(0.154)	(0.0600)	(0.0876)	(0.0388)	(0.0556)	(0.0358)	(0.0397)
Rehired Hispanic	-0.0880	-0.0989	0.00232	0.0951	-0.0798	-0.0371	0.00668	-0.0345
	(0.156)	(0.160)	(0.115)	(0.0801)	(0.0938)	(0.0960)	(0.0759)	(0.0645)
Rehired Other	-0.261	-0.795***	0.485	0.276	0.156	0.365*	0.109	-0.0376
	(0.864)	(0.213)	(0.318)	(0.412)	(0.251)	(0.204)	(0.183)	(0.140)
Rehired White	-0.0752***	-0.0533***	-0.0548***	-0.0543***	-0.0612***	-0.0385***	-0.0548***	-0.0582***
	(0.0160)	(0.0164)	(0.0107)	(0.0117)	(0.0105)	(0.0104)	(0.00992)	(0.0106)
Season FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Brand FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean Asian Models per show	0.62		1.14		1.64		2.58	
Mean Black Models per show		0.54		1.22		1.80		3.11
Observations	1103	1090	1065	1103	1094	1060	1067	1076
R-sq	0.25	0.28	0.30	0.34	0.35	0.39	0.51	0.43

Notes: This table shows the relationship between the number of rehired models and the number of new hires, by designer preferences, e.g. the designers who hire the most Asian (Black) models enter the top quartile. The first coefficient in the first column estimates the change in Asian newcomers in response to the rehiring of an additional Asian model by a group of designers who hire the smallest number of Asian models for their shows. The unit of observation is a fashion show. The sample includes all major runway shows from the Big Four (New York, Milan, Paris, and London) Fashion Weeks, held twice a year, from 2000 to 2017.

The outcome variable is newcomers of a given race, as the percentage of all hires. The new hires are defined as models who have never walked a runway for the brand before. The explanatory variables are the rehired models of different races and ethnicities, as the percentage of all hires in a show. They are scaled by the mean of the dependent variable. The rehired models have walked in a runway for the brand before. Asian models include East Asians, South-East Asians, and South Asians. White Hispanics are included in the White group. Black Hispanics are included in the Black group. Non-White and non-Black Hispanics are a separate group, Hispanic. Other includes visually non-White models (e.g., Middle Eastern, Native American, Pacific Islander) who are not classified elsewhere.

All regressions include season and brand fixed effects. For example, Chanel Spring/Summer 2001 runway show switches on indicators for the season of Spring/Summer 2001, and the Chanel brand.

All regressions are estimated using OLS. Standard errors, clustered on brand, are reported in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 1.10: Race-blind simulation regressions

New hires (% of all hires)	Asian	Black	White
Rehired Asian (% of all hires)	-0.1298	-0.0728	-0.0695
Rehired Black (% of all hires)	-0.0701	-0.1069	-0.0689
Rehired White (% of all hires)	-0.0698	-0.0712	-0.0727
City FE	Yes	Yes	Yes
Season FE	Yes	Yes	Yes
Brand FE	Yes	Yes	Yes
New hires, average count (by race at the top)	0.6148	1.2612	11.9266
Observations per draw	4352	4352	4352

Notes: This table shows the relationship between the number of rehired models and the number of new hires, by race. The coefficients in the table are the averages of the coefficients from the baseline regressions on a hundred random samples constructed under the assumption of race-blind hiring. The first coefficient in the first column estimates the change in Asian newcomers in response to the rehiring of an additional Asian model. The unit of observation is a fashion show. The outcome variable is newcomers of a given race, as the percentage of all hires. The new hires are defined as models who have never walked a runway for the brand before. The explanatory variables are the rehired models of different races and ethnicities, as the percentage of all hires in a show. They are scaled by the mean of the dependent variable. All regressions include city, season, and brand fixed effects and are estimated using OLS.

Table 1.11: Newcomers' portfolio comparison regressions

New Asian Hires	Lookbooks, season	Editorials, season	Covers, season	Catalogs, season	Ads, season	Lookbooks, career	Editorials, career	Covers, career	Catalogs, career	Ads, career
Rehired Asian	0.204	9.099***	0.627	-0.0369	1.570***	1.475**	24.51***	3.016	0.0329	6.315***
	(0.218)	(2.388)	(0.591)	(0.0572)	(0.594)	(0.576)	(8.299)	(1.833)	(0.190)	(1.519)
Mean DV	0.10	2.32	0.62	0.02	0.51	0.37	7.19	1.87	0.09	1.70
Observations	2789	2789	2789	2789	2789	2789	2789	2789	2789	2789
R-sq	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01

Notes. The table shows the relationship between the portfolio quality of Asian newcomers and the percentage of Asian rehires in a show. The unit of observation is model-show pair. The sample consists of all model-show pairs where model is an Asian newcomer. The shows without Asian newcomers are excluded from the sample. The outcome variable is the number of editorials, advertising campaigns or other portfolio items of new hires who are Asian. The new hires are defined as models who have never walked a runway for the brand before. The explanatory variable is the rehired models who are Asian, as the percentage of all hires in a show. All regressions are estimated using OLS. Standard errors, clustered on show, are reported in parentheses.

* p<0.1, ** p<0.05, *** p<0.01.

Table 1.12: Newcomers' portfolio comparison regressions

New Black Hires	Lookbooks, season	Editorials, season	Covers, season	Catalogs, season	Ads, season	Lookbooks, career	Editorials, career	Covers, career	Catalogs, career	Ads, career
Rehired Black	0.657***	5.169***	0.212	0.0206	0.721	1.316***	16.87***	2.888	0.463**	4.686***
	(0.215)	(1.441)	(0.243)	(0.0765)	(0.510)	(0.382)	(3.608)	(1.883)	(0.219)	(1.294)
Mean DV	0.08	1.42	0.30	0.05	0.51	0.20	3.09	1.01	0.15	1.30
Observations	3700	3700	3700	3700	3700	3700	3700	3700	3700	3700
R-sq	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.01

Notes. The table shows the relationship between the portfolio quality of Black newcomers and the percentage of Black rehires in a show. The unit of observation is model-show pair. The sample consists of all model-show pairs where model is a Black newcomer. The shows without Black newcomers are excluded from the sample. The outcome variable is the number of editorials, advertising campaigns or other portfolio items of new hires who are Black. The new hires are defined as models who have never walked a runway for the brand before. The explanatory variable is the rehired models who are Black, as the percentage of all hires in a show. All regressions are estimated using OLS. Standard errors, clustered on show, are reported in parentheses.

* p<0.1, ** p<0.05, *** p<0.01.

Table 1.13: Newcomers' portfolio comparison regressions

New White Hires	Lookbooks, season	Editorials, season	Covers, season	Catalogs, season	Ads, season	Lookbooks, career	Editorials, career	Covers, career	Catalogs, career	Ads, career
Rehired White	-0.0253**	-0.614***	0.254***	-0.00152	0.354***	-0.189***	-2.414***	0.681***	-0.0385*	0.374***
	(0.0101)	(0.110)	(0.0291)	(0.00487)	(0.0423)	(0.0359)	(0.311)	(0.152)	(0.0214)	(0.118)
Mean DV	0.07	1.16	0.37	0.02	0.42	0.22	3.00	1.31	0.09	1.22
Observations	48304	48304	48304	48304	48304	48304	48304	48304	48304	48304
R-sq	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Notes. The table shows the relationship between portfolio quality of White newcomers and the percentage of White rehires in a show. The unit of observation is model-show pair. The sample consists of all model-show pairs where model is a White newcomer. The shows without White newcomers are excluded from the sample. The outcome variable is the number of editorials, advertising campaigns or other portfolio items of new hires who are White. The new hires are defined as models who have never walked a runway for the brand before. The explanatory variable is the rehired models who are White, as the percentage of all hires in a show. All regressions are estimated using OLS. Standard errors, clustered on show, are reported in parentheses.

* p<0.1, ** p<0.05, *** p<0.01.

CHAPTER II

Local Fiscal Multiplier on R&D and Science Spending: Evidence from the American Recovery and Reinvestment Act

2.1 Introduction

What is the effect of federal R&D and science[29] spending on local employment? This paper analyzes the 2009 ARRA stimulus spending on R&D and science to estimate the effect of fiscal spending on employment [Keynes, 1936] and the effect of R&D and science on economic growth [Schumpeter, 1942]. We provide new insights to these old questions by finding large employment effect of R&D and science spending.

Science is generally perceived as a long-term endeavor, a foundation for applied research, valuable to the nation in the long run but hardly relevant for a short-term economic development. The main contribution of R&D and science to the economy lies in the areas of innovation, technological growth, and entrepreneurship. Investments in science take time to come to fruition and the outcomes take on various forms of codified knowledge (scientific publications, patents, algorithms, methods), new products and services services, as well as highly trained individuals. In addition to the long-term scientific contributions to the economy, R&D and science also affect short-term economic development through job creation. This paper focuses on this short-term effect, and approaches the scientific process as daily productive work, not too different from routine office work [Weinberg et al., 2014].

[29] Hereafter "R&D and science" and "research" are used interchangeably.

The existing economic literature does not provide a jobs multiplier specific to R&D and science spending. If the earnings and consumption of researchers are similar to those of workers in other industries, why would the jobs multiplier be different? In a standard Keynesian model, it does not matter where new income enters the economy. If we abstract away from the long-term benefits of new knowledge, digging ditches and building rockets should have similar short-term effects on job creation. If anything, there has been a presumption that "brick and mortar" infrastructure spending has greater employment effects, and that science is likely to give rise to creative destruction of jobs.[30]

Presuming that R&D spending exerts little or no short term stimulus effect can lead to under-investment in these important activities. Recent evidence suggests that the composition of government spending may matter for the size of the multiplier. Federal spending on non-durable goods, including services, has been found to generate a larger GDP multiplier than spending on durable goods [Boehm, 2016]. Feyrer and Sacerdote [2011] find considerable variation in the size of the multiplier for different types of spending, ranging from negative multipliers for education and public security to positive multipliers on low-income support, transportation, and energy. Chodorow-Reich et al. [2012] estimate the multiplier specifically for Medicaid outlays. Leduc and Wilson [2013] examine the multiplier on highway spending. We contribute to this literature by providing an estimate of the multiplier on R&D and science spending. The capital-to-labor ratio, earnings level, employee consumption patterns, complementarity with other sources of funding, uncertainty, and flexible capacity (i.e., lower adjustment costs) in R&D and science may all contribute to the differences between the multiplier on R&D and science and an aggregate multiplier on all government spending.

We treat the measurement of a fiscal multiplier on R&D and science spending as an empirical question. While discrepancies among multiplier estimates in the literature may reflect differences in macroeconomic conditions, the form of the stimulus (e.g., tax cuts, direct spending, or transfers), data sources, or the estimation approach, we investigate the possibility that differences in the spending purpose directly affect the multiplier.

[30] The larger multiplier on "brick and mortar" spending is also a reflection of presumptions about differences in the average earnings and marginal propensity to consume of different employees.

We abstract away from the general equilibrium effects of fiscal stimulus and estimate a local multiplier. Note that a local multiplier is a different indicator than the national multiplier in a Keynesian model. The national multiplier represents the value (in dollars of GDP or number of jobs) created after the government adds a dollar of stimulus to a closed economy. The national multiplier is higher when interest rates are low [Woodford, 2011], higher during recessions [Auerbach and Gorodnichenko, 2012], and lower for temporary increases in government spending [Baxter and King, 1993]. The local multiplier measures the change in the output or employment after adding a dollar of government spending to a smaller open economy, such as state or county, *relative* to other open economies within a fiscal union [Nakamura and Steinsson, 2014]. Unlike the national multiplier, the local multiplier estimated here is not sensitive to macroeconomic changes, like changes in the interest rate, that are common to all economies in a fiscal union. It is still affected by labor underutilization during recessions, as that varies across counties.

The variation used to estimate local multipliers comes from county- or state-level differences. Local fiscal multipliers, especially at the county level, may be estimated more precisely than a national multiplier because of a larger sample size. They, however, are sensitive to attenuation bias from measurement error in the location of spending. This paper uses new, more carefully constructed measures of the location of ARRA spending in order to reduce attenuation bias. Local multipliers are also sensitive to spillover effects from cross-county or cross-state mobility: they are increased by labor mobility if spending attracts in-migration to counties that receive more spending and they are decreased by mobility in consumption if employees spend their additional income in neighboring counties or states.

We examine changes in employment in response to federal spending on R&D and science in the context of the American Recovery and Reinvestment Act (ARRA). The ARRA was signed into law in February 2009. Its goal was to provide a large federal stimulus to reduce the toll of recession on the American economy. The large size and the speed of disbursement were two important aspects of ARRA. It generated a quantitatively significant, largely unanticipated shock to government spending.[31] As we discuss further below, R&D spending was particularly unanticipated, making it especially approapriate for estimating causal effects.

While most ARRA spending targeted areas hit hard by the recession, the geographic allocation of R&D and science spending was intended to be exogenous to economic conditions. The allocation was based on a peer review process or the availability of resources to carry out research projects. For example, when the stimulus bill was passed, National Science Foundation (NSF) program officers funded deserving proposals they had not previously had the resources to fund. Approximately 80% of stimulus-backed awards went to projects submitted prior to ARRA. The National Institutes of Health (NIH), on the other hand, issued a call for new proposals to be funded under ARRA [Harmon, 2010]. Even there, the recipients had to be institutions that were prepared to submit a credible NIH proposal under a very tight deadline, so the geographic allocation primarily reflected local scientific capability.

Even though these grants were allocated based on scientific merit or capability, they were not assigned randomly. This presents a challenge in measuring the effect of ARRA research spending on local employment. The counties receiving ARRA research awards may be different from other counties in ways that influence employment trends. In addition to controlling for factors affecting employment and modeling a change in employment trend for each county, we employ a two-stage strategy to reduce endogeneity problems. First, we estimate the probability of selection into receiving ARRA research funds and construct an inverse Mills ratio term to capture it. A county's research intensity is the main predictor of selection into receiving ARRA research award and award's size. We use two dummy variables as excluded covariates of research intensity: whether a county has a research university and whether there is at least one person employed in R&D and science services in the county prior to the recession. We estimate a cross-county IV regression on a subsample of counties with R&D and science awards to estimate the effect of ARRA research awards on the change in employment. We use the inverse Mills ratio from the

[31] Hourihan [2015] provides an overview of ARRA research spending in the context of total federal spending on R&D and science.

selection equation and two new instruments: the natural logarithm of doctoral degrees issued in a county in 2010 and the number of individuals employed in R&D and science per capita in 2007, in the cross-county IV regressions.

We find that during ARRA disbursement period, which lasted from 2009 to 2013, 27 jobs were added in response to one million USD in ARRA stimulus on R&D and science. Traditionally, the multiplier is presented in a form of a job-year cost. Converting our baseline result into job-year cost is not straightforward because the disbursement of ARRA funds took place over five years, the average length of a project exceeded two years, and the data on yearly payments are not available at the county-level. Taking all these into account leads to an estimated \$15,000 USD per job-year, one of the lowest estimates in the recent literature on fiscal multipliers.

By providing an estimate of the fiscal multiplier on R&D and science spending, we contribute to the economics literature in two ways. First, we contribute to the literature on fiscal multipliers. We provide the first estimate of the multiplier on R&D and science spending. We also contribute to the literature on the effect of R&D and science on local economy. Hausman [2012] measures the effect of university innovation on long-term economic growth of local economies. She finds that an additional \$10 million of the Department of Defense (DOD) funding or \$7 million of NIH funding before 1980 generated an additional job per county-industry after 1980. Dinerstein et al. [2014] find no evidence of employment growth in counties around universities in response to federal research funds. We study federal spending on R&D and science in general, not just spending on universities or in university counties, and focus on the short-term effects on local employment during and after the Great Recession, and find much larger effects.

2.2 Empirical Model

Our goal is to evaluate the effect of ARRA R&D and science spending on the change in local employment at the county level. We start with a simple model which estimates the average number of jobs created in a US county over ARRA disbursement period in response to a million USD in research spending. In our baseline specification, we use the change in employment from 2009 to 2013, a time period over which ARRA funds were disbursed to recipients in full, to capture all the spending shocks that accrued due to ARRA. The equation below captures this framework:

$$\frac{Emp_{c,2013} - Emp_{c,2009}}{\frac{1}{5}\Sigma_{n=2009}^{2013} Pop_{c,n}} = \alpha_s + \beta \frac{ARRA\_Res_c}{\frac{1}{5}\Sigma_{n=2009}^{2013} Pop_{c,n}} + \boldsymbol{X}_c \Gamma + \varepsilon_c,$$

where $Emp_{c,t}$ is employment in county c in year t, $Pop_{c,t}$ is population in county c in year t, $ARRA\_Res_c$ is total ARRA spending on research in county c in 2009-2013, $\boldsymbol{X_c}$ is a vector of control variables, α_s is a state-level shock, and ε_c is an error term.

This specification follows estimation strategies in earlier literature on cross-sectional fiscal multipliers [Chodorow-Reich et al., 2012; Wilson, 2012]. The minor difference is that we are estimating county-level multiplier. State-level multipliers are more common in the earlier studies.

The counties which received research awards under ARRA are, on average, larger than all other counties.[32] To account for that, we scale the outcome variable and all ARRA variables by the population averaged over disbursement period following the standard practice in the literature. We scale employment-based control variables by the population averaged over respective time periods.

We estimate the effect of ARRA research spending on employment at the county level. Because US counties are open economies, we have to account for spillover effects from different sources: worker spending outside a county, cross-county mobility for job opportunities, and mis-measurement of money flows from primary contractors to subcontractors and vendors. There is an important difference between attenuation due to open economy and due to mis-measurement or misreporting of the geography of spending. The spillovers from cross-county spending and labor mobility are inherent in the level of analysis. We test whether factor mobility is driving the estimates, but it doesn't necessarily make the estimates wrong. The location of spending, on the other hand, is a measurement issue, and we address that by providing a precise match between ARRA

[32] Table 2.9 shows summary statistics for counties which received ARRA spending on R&D and Science in comparison to all other counties.

spending and geographic location of all recipients, including vendors and subcontractors.

The spillover effects of the worker spending in other counties and mis-measurement of money flows cause the attenuation of estimates. The worker cross-county mobility leads to the overestimation of the effect. We control for spillover effects in two ways: track subcontractor and vendor transactions to their zipcodes[33] and control for ARRA research spending in adjacent counties. It is also possible that some counties received more ARRA stimulus in general due to seniority and political weight of their representatives. We proxy this "clout" with ARRA spending on all other issues.

We control for a number of factors which are correlated with the pace of economic recovery. We include the change in employment from 2007 to 2009 because counties that lost fewer jobs during recession might have less room to add new jobs during recovery. The post-recession employment changes might differ between urban and rural counties or depend on the fraction of manufacturing jobs in a county. For this reason, we add an indicator for metropolitan county and per capita count of individuals employed in manufacturing before recession. We also include state fixed effects to account for differences in post-recession recovery across states. Even after controlling for different characteristics of a county that are relevant to the changes in employment, we cannot exclude the possibility of county-specific trends. To alleviate this issue, we use prior data to model county-specific trends in employment changes over five-year periods on a rolling basis. We then extrapolate out of sample to predict the change in employment from 2009 to 2013 for each county.

We rely on a number of ARRA features to reduce concerns about strategic hiring and reallocation of funds. The stimulus was largely unanticipated by the final recipients. This is especially true for research spending. It does not fall under the areas traditionally subsidized during recessions and was not expected to be a part of ARRA until the last moment. This situation is somewhat unique as anti-recessionary spending is often anticipated.

[33] While it is possible that some of the recipients are importers, we are not able to track ARRA funds to foreign countries. We take export-import structure of a county as given and estimate the effect of ARRA on the number of jobs created locally. We do not include foreign jobs created as a result of ARRA into our multiplier.

We address the possibility of substitution between ARRA research awards and other sources of R&D and science funding. There is no evidence of substitution between on-going federal spending on R&D and science and ARRA research awards. The federal government gave out ARRA stimulus in addition to the federal R&D and science awards. The latter were trending flat over the past decade. Figure 1 from Hourihan [2015], shows ARRA R&D spending as a "bump" on top of a flat trend in federal stimulus for R&D. While somewhat less likely, it is possible that some recipients, mostly large universities, are not budget constrained and there is a substitution between ARRA research funds and institutions' own resources. However, universities' own institutional spending on R&D continued to increase over this time period according to the Higher Education Research and Development (HERD) survey. Recent economics literature provides evidence that federal aid caused universities to increase their investment in research and human capital [Dinerstein et al., 2014]. The same study acknowledges a slight reduction in endowment spending for private universities and state appropriations for public universities. These findings can be interpreted as a substitution effect but it is not sizable.

Yet, there are challenges in measuring the effect of ARRA research spending on local employment. The awards are not assigned randomly. There is a possibility that the counties receiving ARRA research awards are different from other counties. We use a Heckman-type correction to account for non-random selection of counties into receiving ARRA research awards. We estimate the probability of a county receiving an ARRA research award using all control variables and two good predictors of a county receiving ARRA research stimulus. They are a dummy for a research university in a county and a dummy for having any people employed in R&D and science services in a county before the recession. We estimate the following selection equation using probit:

$$S_c = \mathbb{1}[ARRA\_Res_c > 0]$$

$$S_c = \delta_s + \delta_1 Res\_Uni_c + \delta_2 Res\_County_{c,2007} + \boldsymbol{X}_c \Psi + \upsilon_c,$$

where Res_Uni_c is the dummy variable for a county with research university, $Res\_County_{c,2007}$

is the dummy for a county with employment in R&D and science services in 2007, δ_s is a state-level shock, and υ_c is an error term.

We construct an inverse Mills ratio, $\hat{\lambda}_c$, using predicted values from the estimated probit model. The inverse Mills ratio corrects the bias from non-random selection of counties into receiving ARRA research stimulus in a sample of counties with non-zero ARRA research awards.

We need to account for the endogeneity of ARRA research stimulus. This endogeneity is, in principle, less serious than for other types of federal spending, such as spending on unemployment or housing. Research awards are not assigned based on the socio-economic conditions of a county. Recipients received ARRA funds based on peer review, innovative potential, or existing infrastructure for scientific and technological discovery. Even though research awards are not based on socio-economic conditions of a county, they cannot be considered independent of them. The data show that counties with large research awards are more populous, urban, affluent, and have more complex industrial structure. One possibility is that counties with large research awards grow faster than counties with small awards. Another possibility is that counties with large research awards can smooth recessions better and have little room to add new jobs during the recovery.

We employ an instrumental variable strategy to account for resulting endogeneity. We use two different predictors of ARRA research stimulus, a natural logarithm of doctoral degrees awarded at the universities in a county in 2010 and the number of people employed in R&D and science services per capita before recession, as instruments. We also include the inverse Mills ratio, $\hat{\lambda}_c$, in the first stage to correct for the selection bias in ARRA counties with research awards:

$$\frac{ARRA\_Res_c}{\frac{1}{5}\Sigma_{n=2009}^{2013} Pop_{c,n}} = \phi_s + \phi_1 \ln(Doc\_Degrees_{c,2010}) + \phi_2 \frac{Res\_Emp_{c,2007}}{Pop_{c,2007}} + \phi_3 \hat{\lambda}_c + \boldsymbol{X}_c \Omega + \xi_c,$$

where ϕ_s is state-level shock, $Res\_Emp_{c,2007}$ is the number of individuals employed in R&D and scientific services in county c in 2007, $Doc\_Degrees_{c,2010}$ is the number of doctoral degrees awarded in county c in 2010, and ξ_c is the error term.

Our baseline estimate comes from the following cross-county instrumental variable regression on a sub-sample of counties with ARRA research awards:

$$\frac{Emp_{c,2013} - Emp_{c,2009}}{\frac{1}{5}\Sigma_{n=2009}^{2013} Pop_{c,n}} = \kappa_s + \beta_{IV} \frac{ARRA\_Res_c}{\frac{1}{5}\Sigma_{n=2009}^{2013} Pop_{c,n}} + \boldsymbol{X}_c \Theta + \eta_c,$$

where κ_s is state-level shock, and η_c is the error term.

2.3 Data

Total ARRA spending was an estimated \$831 billion[34], including contract, grant, and loan awards, expansion of entitlement programs, such as food stamps and unemployment insurance, direct grants to states, Medicaid match program, tax benefits and federal government consumption and investment. We will focus on ARRA transfers to individuals, businesses, and local institutions. The total amount of ARRA contract, grant, and loan awards reported by recipients is approximately \$278 billion, or about one-third of all ARRA spending.[35] This number does not include other components of ARRA.

Our main data source is the Cumulative National Summary of ARRA recipient reports.[36] The data contains reports from local governments, private entities, and individuals on the amount of stimulus received under ARRA.

The data on the Cumulative National Summary of Recipient Level Reports has a multi-level structure. Figure 2.1 provides an example to illustrate the hierarchy of ARRA award recipients in the data.

We observe the flow of funds from primary contractor to subcontractors and vendors. Most recipients disclose a place of performance (POP) with a five-digit zip code. We supplement zip codes from vendors' DUNS using Dun & Bradstreet Unique Partner Identification Key. We then map zip codes to US counties using US Census Zip Code Tabulation Area, look-up feature of Melissa Data, and HUD USPS zip code crosswalk

[34] Original estimate was \$787 billion. It was later adjusted up, to \$831 billion [Congressional Budget Office, 2012].

[35] Author calculations based on the recipient reports in the Cumulative National Summary.

[36] We downloaded the last version of these reports from the Federal Procurement Data System on October 26, 2015. The data collection at recipient level was finalized on September 30, 2015. There will be no further updates to this data set [Clark, 2015]. Before September 2015, this data was hosted on Recovery.gov.

files. Zip codes that cover two or more counties were assigned weights based on the county population share in the zip code using HUD USPS zip code crosswalk file. Table 2.1 presents the results of matching recipients' POPs to US counties. We match $270,334 million of ARRA awards from the Cumulative National Summary to a POP in a US county. Because ARRA reports allow us to trace both disbursement to a primary recipient as well as subcontracts we are able to trace spending with a high degree of local specificity. In other words, these data provide new and heretofore unused information about the national geography of R&D stimulus spending and thus support the local, county-level, models we estimate.

County level spending data are matched to public information about employment. We use data on total county employment and county employment in the private sector from the Bureau of Labor Statistics (BLS) Quarterly Census of Employment and Wages (QCEW).[37] We use annual average employment because we do not know specific dates of ARRA spending shocks to county employment.[38] We use data on total employment in a county to account for both direct and indirect job creation following ARRA stimulus and avoid the issue of overestimating the number of jobs if employees are switching jobs to fill in the positions funded by ARRA in the same county [Jones and Rothschild, 2011].

We use the QCEW for a control variable, the number of manufacturing jobs per capita in a county before the recession. For the number of people employed in R&D and science, we use data on county-level employment in scientific research and development services (NAICS (2012) code 5417) from the US Census County Business Patterns.

The literature in economics does not provide an established definition of research spending. We thus calculate two alternate measures of R&D spending that rely on the

[37] The Cumulative National Summary of ARRA recipient reports contains information on the number of jobs reported by ARRA recipients. We do not use this data for three reasons. First, they do not account for the jobs created indirectly, by recipients spending their income in the local economy. Second, the jobs are reported by the primary contractor and should represent the sum of all jobs created or retained by primary contractor as well as its vendors, subcontractors, and their sub-vendors. We can not track these jobs to the geographic location of subcontractors, vendors, and sub-vendors. Third, these job numbers are approximations resulting from primary contractors' uncertain knowledge about the linkage of ARRA spending to a specific job, as well as recipients' lack of incentive to calculate them precisely.

[38] The Federal Procurement Data System provides the data on spending disbursed over five-year period from 2009 to 2013 and spending disbursed in Q4 2013. We, therefore, can not construct a panel as the timing of disbursement at a quarter-level is not available.

missions of federal agencies that made grants and the purposes reported specifically for ARRA spending. Calculating those measures relies on two data sources, the Catalog of Federal Domestic Assistance (CFDA) and "Where Does the Money Go?" map from the Recovery.gov website, to define ARRA funds which contribute to R&D and science.

For every US county, we track research spending under ARRA to the zipcodes within its borders and also to the adjacent counties. We determine adjacent counties using US Census data on county adjacency from 2010.

All employment and ARRA spending variables are scaled by county population from the US Census Annual Estimates of the Resident Population by County.

To control for the type of county, we identify metro counties using Rural-urban Continuum Codes (2013) published by the US Department of Agriculture and counties with research universities using the Carnegie Classification (2010) and the Post-secondary University Survey (2013) published by the US Department of Education. We obtain the number of doctorate degrees awarded in a county from the Carnegie Classification (2010).

The variables and their data sources are summarized in Table 2.2.

2.4 Definition of Research Spending

We employ two definitions of R&D and science spending. Our main definition is based on the CFDA numbers: we identified 24 numbers describing scientific and research activities funded under ARRA. The CFDA numbers which mention the following are classified as research funding: research, science, census, statistics, policy evaluation, data, surveys, studies, laboratories, analysis, university building capacity. The CFDA numbers that mention library, education, including higher education, conservation, museums, training are not classified as research funding. Complete list of research CFDA numbers and allocated funds are in Table 2.3.

The secondary definition comes from "Where does the money go?" section of the Recovery.gov website. The section depicted a map with locations of ARRA recipients. The users could search an award based on different characteristics, including its purpose. All funds labeled "for R&D and Science" enter our secondary definition.

The data on the type of funds released on the Recovery.gov website were matched to the outlays data from the Cumulative National Summary.[39]

We use two definitions to recover missing values. We classify transactions with missing CFDA numbers based on the Recovery.gov data. Similarly, the transactions that were not matched to the data from "Where Does the Money Go?" feature are classified as R&D and science based on their CFDA number. The comparison of two definitions is in Table 2.4. Because our coding of CFDA numbers allows us to identify not only spending by traditional science agencies (e.g. the National Institute of Health and National Science Foundation) as well as programs within other agencies that are focused on R&D, the estimate of research spending using that definition is higher than that reported by the Recovery.gov.

Descriptive Analysis

Tables 2.5 and 2.6 show the breakdown of ARRA spending by year. Table 2.5 breaks ARRA awards by the year in which they were assigned to recipient, while Table 2.6 shows the year of the final report filed by recipients indicating work completion. We can see that the majority of awards were assigned in 2009 and completed in 2013. The pattern is similar for research and non-research awards with one caveat. While the percentage of awards assigned early, in 2009, is similar - 59% of research awards and 62% of non-research awards, research projects took longer to complete. Sixty percent of recipients of research awards filed final report in the last year (2013) as compared to 44% of recipients of non-research awards. The average time between receiving an award and filing a final report is 772 days for non-research projects and 890 days for research projects.[40] This finding is consistent with the notion that R&D and science projects are more long-term

[39] The data on the purpose of funds is only available for primary contractors and subcontractors. We assume the same purpose for their vendors. The data on the purpose of funds does not have id variable *award_key*. We used *award_number* and *order_number* as id variables. Some of the values are either missing or inconsistent with corresponding variables in the main data. We matched 99.02% of transactions, which represent 87.08% of outlay amounts, based on three variables: *award_type*, *award_number*, and *order_number*.

[40] Author calculations. We do not observe the date of the final report, only its year and quarter. We assign the date as the 45^{th} day of respective quarter.

and uncertain, but may also have to do with standard university reporting practices for grant awards.

Tables 2.7 and 2.8 contain the data on the counties with the largest amounts in research awards under ARRA. Table 2.7 shows that the counties with the largest aggregate amounts are located on the coasts or in large metropolitan areas. Among the top 25 counties, five are in California, each grossing at least 400 million USD in total ARRA funds on research, two are in Massachusetts, two in New York, two in Maryland, and Washington DC. Cook County in Illinois, Harris and Dallas Counties in Texas, Wayne County in Michigan, Philadelphia and Allegheny Counties in Pennsylvania, Wake County in North Carolina, Milwaukee County in Wisconsin, Maricopa County in Arizona, Miami-Dade County in Florida, Hamilton County in Ohio, Kong County in Washington are homes to large cities with complex industrial structures. Washtenaw County in Michigan is the location of the University of Michigan, the largest public university performer of research.

It is important to remember that counties with the largest aggregate amounts of research awards are not necessarily the most research-intensive. After scaling ARRA awards by population the coastal areas and counties with large cities are no longer prominent.

Figure 2.2 displays the largest recipients of research spending under ARRA in aggregate and per capita terms at a county level. A little over half of all the counties received some research funds between 2009 and 2013. However, the distribution of awards, even when examined per capita, is skewed. About one-third of all counties received more than 5 USD per capita over five years. 382 counties received more than 50 USD per capita, 244 more than 100 USD per capita, and 34 counties more than 500 USD per capita.

The list of counties with the largest amounts in research awards is different if compiled on per capita basis. Suffolk county in Massachusetts and Washtenaw County in Michigan are the only counties from Table 2.7 to appear in Table 2.8, which displays the top 25 counties on per capita basis.

Some of the counties enter Table 2.8 due to low population numbers. If the denominator in the per capita definition of research awards is small, even 10 million USD in federal awards over five years will place a county at the top of the list, as evidenced by Esmeralda County in Nevada. Three Nevada counties as well as three remote boroughs in Alaska, Morrow County in Oregon, and Pontotoc County in Mississippi are low population counties which received federal funds for Renewable Energy Research and Development. Orange County in North Carolina, Washtenaw County in Michigan, Tompkins County in New York, and Suffolk County in Massachusetts are homes to large research universities. Delphi Corp, a recipient of a large grant in Conservation R&D, has one of its main offices in Kokomo, Howard County. Los Alamos County is on the list due to the grants of the Los Alamos National Laboratory from the National Science Foundation (NSF) and the Department of Energy (DOE). Anderson County in Tennessee is a home of the Oak Ridge National Laboratory. Marinette County in Wisconsin is on the list mainly because of the two large contracts of Marinette Marine Corp. with the NSF and the National Oceanic and Atmospheric Administration (NOAA).

Table 2.9 compares counties with research spending and counties with spending on all other purposes. Between 2009 and 2013, every county received some ARRA awards for purposes other than research. About half of all counties received ARRA stimulus for R&D and science in the same time period. The counties with research awards are more populous, urban, twice as likely to have a research university, and have more individuals employed in scientific research and development services. More doctoral degrees are awarded in the counties with ARRA research awards. They are similar to all other counties in the share of individuals employed in manufacturing and the increase in the unemployment rate during the Great Recession.

Tables 2.10 and 2.11 examine the geographical dispersion of ARRA awards. Table 2.10 includes all awards. Table 2.11 excludes awards with only one recipient.[41] We separate awards with many recipients because subcontractors and vendors change the geography of spending. They are often overlooked at a less granular analyses of fiscal spending, resulting in attenuated estimates. Both tables show the same pattern: primary contractors on research awards are more likely than primary contractors on non-research awards to have subcontractors and vendors outside their zip code, county, and state.

[41] By default, these awards have all transactions and 100% amount in the same zip code, county, and state. They constitute about 60% of all awards in the data.

However, the relative amounts they are sending to other zip codes and counties are smaller than corresponding amounts sent by primary contractors on all other awards. Primary contractors on research send, on average, 12% of the total award amount outside their zip code, including 10% of the total amount going outside the county. Corresponding percentages for primary contractors on non-research awards are 17 and 13. The pattern reverses at the state level. Primary contractors on research awards send, on average, 8% of total award amounts outside their state, while primary contractors on non-research awards do so with only 3% of award amounts. This pattern is inconclusive about the "stickiness" of research awards in comparison to all other awards. However, it is evident that research contractors have more remote subcontractors and vendors. This finding is expected as research contracts require specialized materials, processes, and services.

Table 2.12 shows summary statistics for the 3,102 counties in the sample. The top panel shows details on outcome variables, total employment and private sector employment, denoted as changes per capita. On average, total county employment per capita rose by about 0.8% between 2009 and 2013. During the same time period, average private sector employment increased by about 1%.

The next panel shows the details on ARRA spending variables. A little over half of all counties received some federal funds on research under ARRA. The average amount was $34 per capita over five years, from 2009 to 2013. In contrast, every US county received some ARRA funds for purposes other than research and the average amount is more than twenty times higher, $824 per capita, over the same period.

We use two instruments, an indicator that a county has a research university and an indicator that a county has at least one person employed in R&D and science in 2007, in the selection equation. We use three instruments, different from the first two, in the first stage of IV regression. They are the number of individuals employed in R&D per capita in 2007, the natural logarithm of doctoral degrees awarded in 2010, and the inverse Mills ratio. According to the Carnegie Classification (2010), 205, or 6.5% of, US counties have a research university. According to the County Business Patterns, only 11.3% of US counties have at least one individual employed in R&D in 2007. On average, a county had

304 individuals employed in R&D for every million residents. In contrast, there were, on average, 46,257 individuals employed in manufacturing per million residents in the same year. The numbers for the doctoral degrees are similarly skewed. On average, there were 19 doctoral degrees awarded in a US county in 2010, but the 90^{th} percentile is zero.

The bottom panel contains information on employment change during the recession. On average, between 2007 and 2009, total employment in a county fell by 1.5%. The corresponding number for private sector employment is 1.6%.

2.5 Baseline Results

2.5.1 Heckman Correction

Table 2.13 presents the estimates from the probit regression[42] used to correct for selection of counties into receiving ARRA research awards. The outcome variable is a dummy for a county with non-zero ARRA spending on research. Column 1 shows the results of the baseline regression. Column 2 presents the same regression as in Column 1 using the secondary definition of research spending. Column 3 gives results for the same regression as in Column 1 substituting total employment with private sector employment.

The counties with ARRA research awards are different from the rest of the country. They are more urban, more likely to have a research university and individuals employed in R&D and science, and are more likely to be surrounded by the counties that also received ARRA research stimulus. According to the baseline specification, at least one R&D job in a county in 2007 increases the probability of receiving ARRA research awards by 33%. A research university in a county increases the probability of receiving ARRA research awards by 55%. Urban counties are, on average, 24% more likely to receive ARRA research awards. One thousand USD in ARRA research awards per capita received by adjacent counties increases the probability of the focal county getting ARRA research funds by 39%, suggesting that this spending, like other R&D measures, is subject to agglomeration effects. One standard deviation increase in employment in manufacturing per capita increases the probability of receiving ARRA research funds by 4%.

[42] We present marginal effects for easier interpretation of coefficients.

Negative coefficients on the change in employment during recession suggest that counties that were later awarded ARRA stimulus on research fared better during the recession. The consequences of that are not clear. The counties with ARRA research awards may have slower growth rates because there is less room for new jobs during the recovery, or higher post-recession growth rates because they, in general, have better socio-economic profiles, or a mix of both.

We use the estimates from the Heckman correction regression to construct the inverse Mills ratio. We include it in the first stage regression on a subsample of counties with ARRA research awards. It corrects for the selection of counties into receiving ARRA stimulus for R&D and science by controlling the part of the error term for which selection into getting funded affects the funding amount.

2.5.2 First Stage

Table 2.14 provides first-stage results of the IV regression. In all specifications, the instruments are good predictors of the endogeneous variable, ARRA spending on research. The baseline specification shows that conditional on receiving ARRA stimulus, one extra person employed in R&D and science before the recession, increases research spending under ARRA by 15,300 USD. Likewise, a one percent increase in awarded doctorates in a county increases ARRA research spending by 19 USD per capita. These magnitudes are large. From descriptive statistics, we know that less than one-tenth of all counties received more than 100 USD in ARRA spending on research per capita. However, the number of counties where any doctorate degrees were awarded in 2010 is small - 312. The number of counties with more than a thousand people employed in R&D and science services in 2007 is even smaller: 89.

The inverse Mills ratio from the selection model is positively signed but not statistically significant in all three specifications. It suggests that unobserved factors that make counties more likely to get ARRA stimulus on research are also associated with larger amounts in stimulus per capita but the relationship is not statistically significant.

The robust first-stage F-statistic in the baseline specification is 17. The same statistic in the regression which uses our second definition of R&D and science spending is 21. The F-statistic in private employment regression is 28.

2.5.3 Main Results

Table 2.15 presents an endogeneous OLS regression in Column 1 and second-stage results of the IV regressions in Columns 2-4. The outcome variable in each regression is the change in employment in a county from 2009 to 2013 per capita. The OLS estimate of the jobs multiplier is 14 with a p-value below 0.01. The coefficient on the IV estimate is much higher, 27, with a p-value below 0.01. Between 2009 and 2013, on average, a county added 27 new jobs in response to one million USD in ARRA research spending. The estimate is roughly the same if we use the secondary definition of ARRA research spending. The majority of jobs, 23 out of 27, were in the private sector.

The OLS estimate is much smaller than the IV estimate suggesting a negative correlation between the change in employment and unobserved characteristics of counties with high research spending. This finding is consistent with the notion that these counties had less room to add new jobs during the recovery.

The effect of an increase in research spending per capita in adjacent counties is positive and statistically significant in IV specifications. We scale the coefficient by the average of the ratio of the population in adjacent counties to the population of the focal county to make the coefficient comparable to the coefficient in the change in employment in the focal county. The coefficients range from 0.9 to 1.2 and are statistically significant. The OLS coefficient equals 0.3 and is not statistically significant. This result rules out negative spillover effects of research spending. The IV specifications suggest that there are positive spillover effects - a county adds jobs in response to research spending in adjacent counties. The coefficients in IV regressions imply one additional job added in a county in response to one million USD in research spending in adjacent counties. Overall, our baseline estimate suggests that one million USD in R&D stimulus spending generated 27 jobs in the county that received the award and one job in counties adjacent to it.

The coefficients on all other ARRA spending are negative and significant. They display a well-known selection effect: federal stimulus goes disproportionately to the counties with worse economic conditions. The initial economic conditions in these counties mask improvements from the federal stimulus. This variable is included as a control and we are not instrumenting for it.

Among other control variables, a dummy for urban counties is consistently positive and significant across all specifications, the number employed in manufacturing before recession is positive and statistically significant in IV specifications, and the change in employment due to recession changes sign based on specification and is not statistically significant. The county-specific trend in employment change is positive but not statistically significant across all specifications.

2.6 Robustness Checks

In Table 2.16, we build the baseline model one variable at a time. The unconditional effect of instrumented ARRA research spending on a sample of counties with ARRA research awards is 36 additional jobs. All control variables impact the coefficient. The largest reduction in the magnitude of the coefficient comes from adding an indicator for Metro County to the model suggesting strong positive correlation between the size of ARRA research spending per capita and a level of urbanization. The changes in the coefficient of interest from the addition of control variables to the model indicate positive correlation between the size of ARRA research spending and the size of ARRA research spending in adjacent counties and negative correlation between the size of ARRA research spending and the size of all other ARRA spending, in per capita terms. Across all specifications in Table 2.16, the coefficient on ARRA research spending varies from 22 to 42.

The three instrumental variable regressions from Table 2.15, including the baseline specification, are presented in Table 2.17 with omitted selection stage and inverted Mills ratio. The coefficients are larger than respective main results, ranging from 32 to 38. This result suggest large selection bias in ARRA award assignments stemming from underlying capabilities of a county to carry out R&D and science activities.

In Table 2.18, we evaluate the robustness of baseline results by splitting the sample. One possible concern is that the baseline results are driven by coastal counties with developed research infrastructure, such as California and Massachusetts. In the first column, we report baseline estimates after dropping counties in Massachusetts and California. The estimate changes very little, from 27 to 25 jobs in response to one million USD in ARRA research spending over the five year period. It remains statistically significant with p-value less than 0.01.

The second column shows results for counties without research universities.[43] This time the change is more pronounced. The coefficient reduces by one third, from 27 to 18 jobs, but is still significant. This finding suggests that the results are disproportionately driven by counties with research universities. It is hardly surprising as ARRA research stimulus is complimentary to the infrastructure for R&D and science.

The third column checks if results are affected by remote, low populated counties and boroughs in Nevada and Alaska. They received high per capita research awards which are low in aggregate terms. The baseline specification omitting Alaska and Nevada shows a coefficient almost identical to the coefficient obtained from the baseline suggesting that counties in Nevada and Alaska were on average very similar in their return to ARRA research spending to the average county in the sample.

Finally, the fourth column checks if large metropolitan areas with top 25 aggregate ARRA research award drive the results. The change is rather large with coefficient dropping to 21 but not as large as omitting counties with research universities.

We also check for the possibility that our main result may be driven by spurious correlation. We estimate the baseline regression omitting the main variable of interest, ARRA research spending, and the county-specific change in employment trend[44] using OLS:

$$\frac{Emp_{c,t} - Emp_{c,2009}}{\frac{1}{5}\Sigma_{n=2009}^{t} Pop_{c,n}} = \tilde{\kappa_s} + \boldsymbol{X^*}_c \tilde{\Theta} + \tilde{\eta_c},$$

[43] We cannot use dummy for research university in the selection equation because by construction none of the counties in this specification have research university.

[44] We use the data on the change in employment before 2009 to construct county-specific change in employment trend. The same data are used to construct the outcome variable in robustness regressions.

where $\boldsymbol{X^*}_c$ is a vector of control variables without the change in employment trend, $\tilde{\kappa_s}$ is state-level shock, and $\tilde{\eta_c}$ is an error term.

We predict the change in employment using control variables for several time periods before the recession[45] (2009-2002, 2009-2003, 2009-2004 etc.) and for several time periods after the recession (2009-2010, 2009-2011, etc). In the absence of spurious correlation, the effect of ARRA research spending is included in the residuals of post-recession regressions but should not be present in the residuals of pre-recession regressions. We construct predicted ARRA research spending using a linear combination of instruments from the first-stage regression (the number of persons employed in R&D per capita in 2007, the natural logarithm of doctoral degrees awarded in a county in 2010, and the inverse Mills ratio from the selection equation). We regress the residuals from the employment regressions on predicted ARRA spending on research. The results are in Figure 2.3. The estimated coefficient is around zero for pre-recession periods indicating that conditional on control variables there is no spurious correlation between the change in employment and instrumental variables. Positive and significant coefficients in the post-recession period are picking up the correlation between the effect of ARRA spending included in the residual and instrumental variables.

2.7 Discussion

2.7.1 Job-Years

We assess the magnitude of our main result by calculating the yearly cost of a job. In the data, we do not observe yearly payments to recipients but we can examine the relationship between ARRA awards and the change in employment over different lengths of time. Table 2.19 shows the coefficients from the regression of the change in employment on ARRA research spending in focal and adjacent counties. One million USD in ARRA stimulus on research added 4 jobs from 2009 to 2010, 16 jobs from 2009 to 2011, 19 jobs from 2009 to 2012, and 27 jobs from 2009 to 2013 in focal counties. Additionally,

[45] We omit 2009-2007 and 2009-2008 because the change in employment between 2007 and 2009 is included in the baseline specification as a control variable.

one job was added from 2009 to 2012 and one job was added from 2009 to 2013 in adjacent counties, with no jobs added or destroyed in earlier period. If we assume that all new jobs lasted for the whole year and no new jobs were created in response to ARRA after 2013, then the total number of job-years created in response to one million USD in ARRA research spending is the sum of the coefficients from four regressions in Table 2.19. We get a total of 66 job-years with a standard error of 13.[46] We interpret this result with caution because an overidentification test fails for regressions on the change in employment in 2009-2010 and 2009-2011. A similar result using OLS is 30 job-years with a standard error of 2 (Table 2.20). These estimates convert to approximately 15,000 USD per job-year using our baseline specification and 33,000 USD per job-year using endogenous OLS regression suggesting large selection bias. Regardless of any longer term economic benefits that accrue to R&D supported by ARRA spending, these investments resulted in significant employment stimulus effects. The cost is further reduced if we add jobs created in adjacent counties to our calculation.

2.7.2 State-level Results

In order to compare our findings to other published models, we attempt to estimate baseline results at the state level. We have to make changes to the main specification to obtain state-level estimates. First, we are no longer applying the Heckman correction because there is no selection into receiving ARRA spending on research at the state level. Every state was awarded some R&D and science funding between 2009 and 2013. We can no longer include control for ARRA research spending in adjacent counties. Instead of a dummy variable controlling for metro counties, we add a state-level measure of urbanization, defined as the number of metro counties in a state.

The instrumental variable strategy has to be modified as well. The number of doctoral students aggregated at a state level is no longer a good predictor of ARRA R&D and science spending. Therefore, we no longer include it as an instrument. Without the

[46] We calculate the standard error on the job-year cost estimates using Delta method. We assume asymptotic normality and independence for the coefficients on ARRA research spending in Tables 2.19 and 2.20.

inverse Mills ratio from the selection model and the number of doctoral students as instruments, we are left with Employed in R&D and science per capita (2007) as the only instrument in our baseline specification. Due to small sample size, we use conventional standard errors.

We present the state level results in Table 2.21. We are not able to obtain a precise estimate of the effect of ARRA R&D and science spending. The coefficient on the main variable of interest is large, larger than comparable magnitudes at the county level, but so are standard errors. The robust first stage F-statistic for the main regression is 12.82.

Table 2.21 provides no evidence of negative spillover effects within a state. If research-intensive counties were to "steal" already employed people from other counties in a state, the coefficient estimate in the state-level regression should be smaller than the respective estimate at the county-level.

The magnitude of the coefficient suggests that the variation in the effect of ARRA spending across states within a region is larger than the variation across counties within a state.

2.7.3 Comparison to the Estimates in the Literature

We compare our baseline estimate to the results in recent studies using similar methodology.[47] We find that our multiplier is large in comparison to the results in other studies.[48]

Wilson [2012] applies cross-state IV methodology to analyze the effect of ARRA grants on total non-farm employment. The estimated effect suggests a cost of $125,000 per job-year in the first year of disbursement. It is higher than our estimate of $15,000 per job-year, or our state-level results, which are not measured precisely. Shorter time-horizon, the analysis at the county level, the possibility of jobs created by ARRA lasting past the first year, as well as the possibility of the R&D and science spending having a higher multiplier can explain the differences between our estimate and a smaller estimate in

[47] We are not comparing our estimates to the results from macroeconomic models, such as Blinder and Zandi [2010] and reports by the CEA and CBO. The direct comparison of local and national multipliers requires a number of assumptions which do not hold in our case.

[48] Chodorow-Reich [2017] contains a review of the recent literature on geographic cross-sectional fiscal spending multipliers.

Wilson [2012].

Chodorow-Reich et al. [2012] apply similar methodology to estimate the effect of ARRA Medicaid reimbursements at a state level. They use past Medicaid spending per capita as an instrument for ARRA stimulus and find a cost of $26,000 per job-year. Conley and Dupor [2013] find a much smaller multiplier, $202,000 per job-year, using ARRA obligations. Leduc and Wilson [2013] estimate employment regression[49] as part of their study of the flypaper effect. They estimate a cost of job-year as $62,500 for highway grants over a three year period (2008-2011).

These papers estimate effects within the first two years of ARRA, at the state level and using the data on the federal grants to states with a breakdown by agency. Our baseline estimate comes from county-level data from the direct recipients of ARRA contracts and grants. The data are broken down by CFDA numbers, a more granular measure than the breakdown by funding department, over five years of full ARRA disbursement.

Feyrer and Sacerdote [2011] use ARRA recipient-level reports on all types of spending at the county level in the first twenty months of ARRA disbursement and estimate a cost of $400,000 per job year. They locate about $85 billion in spending at the county level, while we locate about $270 billion. They also estimate state-level effect of approximately $111,000 per job year.[50] Similarly to Feyrer and Sacerdote [2011], our state-level estimates are higher than the county-level estimates. They attribute it to the positive spillover effects on employment. A non-ARRA paper estimating multiplier at the county level is Serrato and Wingender [2016]. They document a cost of $30,000 per job year.

A direct comparison of our study to recent literature is complicated by the differences in methodology, data sources for employment and ARRA spending, and the time period during which the stimulus was disbursed to recipients. It is, however, evident that our estimate of the R&D and science multiplier is larger than the estimates of the common multiplier or multipliers on other types of spending in the recent literature. In other words, our findings at the county level imply both that R&D spending has significant

[49] Employment regressions are in Table 8 of 2014 working version of their 2017 paper.

[50] We take the estimate in Column 1 of Table 3 because the specification is closest available comparison to our specification.

stimulus effects and that those effects are larger than those that have been reported for many other types of federal stimulus.

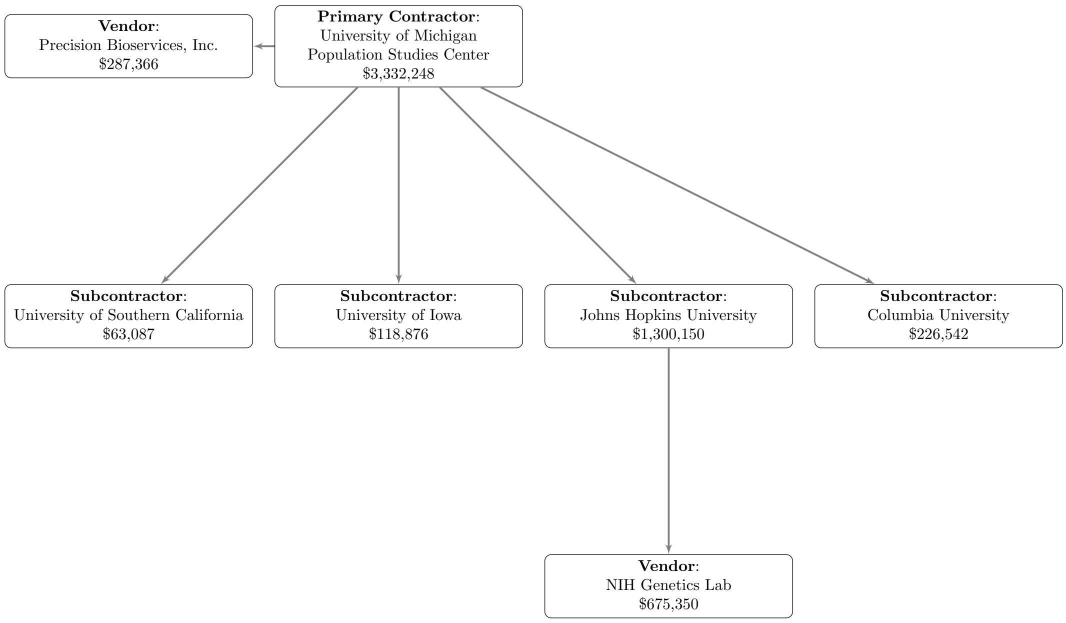
2.8 Conclusion

We examine the impact of ARRA R&D and science spending on local employment. Cross-county IV regressions indicate that ARRA spending on R&D and science has substantively large, positive, statistically significant effects on employment at the county level. We find that between 2009 and 2013, the full ARRA disbursement period, 27 jobs were added in response to one million in spending on research. The majority of jobs, 23 out of 27, were in private sector. Additional analysis provides an estimate of the cost per job-year. We find that 66 job-years were created from one million USD in ARRA research funds which converts to the cost of about $15,000 per job-year. Split sample regressions suggest that the effect is larger for counties with research universities.

Overall, the effect of ARRA spending on R&D and science estimated in our paper is larger than comparable results for federal stimulus in general as well as federal stimulus on health or infrastructure. In addition to any longer term returns realized to discovery and training conducted in the course of ARRA funding R&D, there are substantial short term employment returns to public investments in science and research.

Figures and Tables

Figure 2.1: An example of ARRA award flows

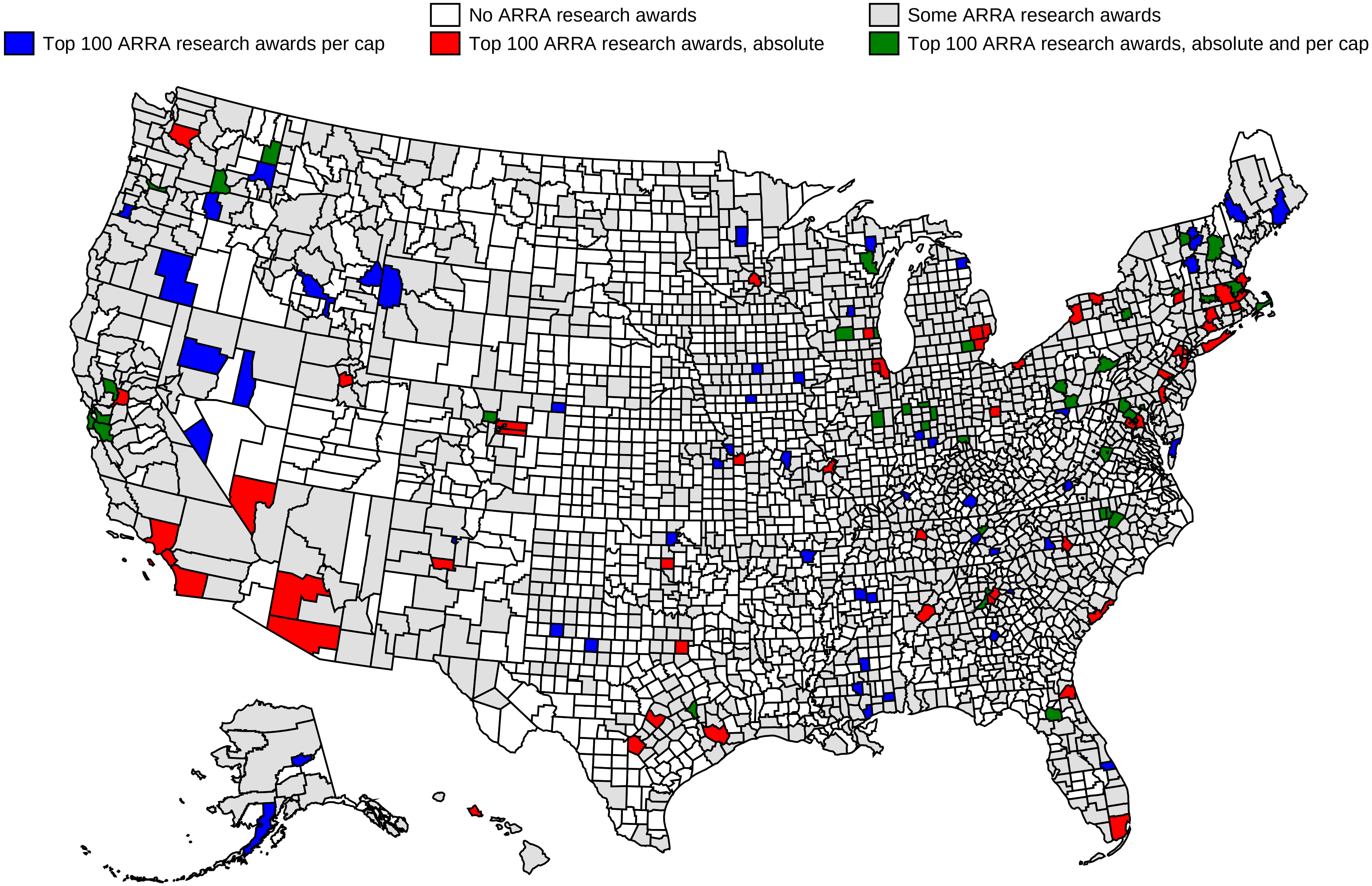


Grant name: Expanding a National Resource for Genetic Research in Behavioral & Health Sciences.

Grant description: This project will utilize high-throughput genetic technologies in a major longitudinal behavioral study and renew the biomedical research community by building scientific partnerships for the integration of behavioral and genetic science. The 7,000 individual participants to be genotyped will be added to a database of 13,000 being constructed under an earlier ARRA project.

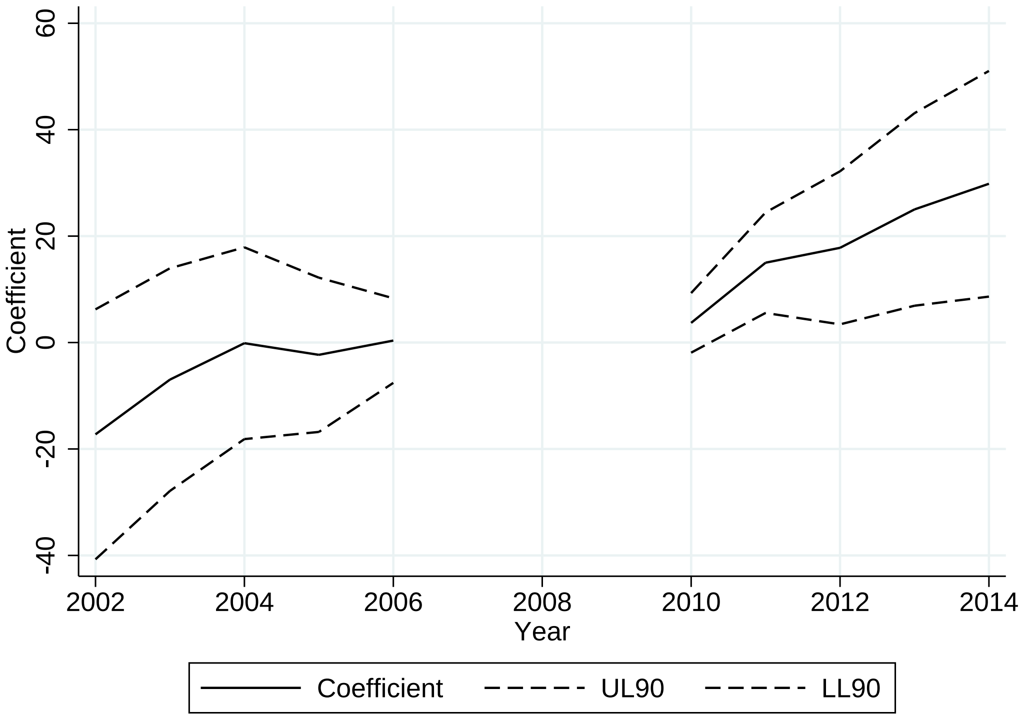
Funding agency: NIH.
Award number: 1 RC4 AG 039029-01.
Total amount: $6,003,620.

Figure 2.2: County-level Science and R&D Spening per capita under ARRA (2009-2013)



Notes: The map shows county-level spending on research per capita under the ARRA in 2009-2013.
The data comes from the recipient reports in the Cumulative National Summary. The ARRA spending on research is defined by the CFDA numbers associated with R&D and science.

Figure 2.3: Robustness Check, Spurious Correlation



Notes: The graph examines the possibility of spurious correlation in the main result. We estimate the regression of the residuals from the baseline specification after omitting ARRA research spending on the predicted values of ARRA research spending. The predicted ARRA research spending is constructed using the instruments from the first stage. The year on the horizontal axis indicates the t in the outcome variable $\frac{Emp_{c,t}-Emp_{c,2009}}{\frac{1}{5}\Sigma_{n=2009}^{t}Pop_{c,n}}$ in the post-recession baseline regressions and $\frac{Emp_{c,2009}-Emp_{c,t}}{\frac{1}{5}\Sigma_{n=2009}^{t}Pop_{c,n}}$ in the pre-recession baseline regressions used to construct residuals. The vertical axes contains the coefficients from the regression of residuals on the predicted ARRA research spending for each time period. We omit 2009-2007 and 2009-2008 because the change in employment between 2007 and 2009 is included in the baseline specification as a control variable.

Table 2.1: Sample Construction.

Data	Obs
Raw Data	615,226
Awards with Primary Contractor	615,189
Total Award Amount is not missing	615,188
Transactions without duplicate Primary Contractor	615,171
Transactions without duplicate Subcontractor	615,162
Transactions without duplicate Subcontractor's Vendor	615,150
Transactions without negative local amount	613,224
Non-zero local amounts	564,588
Non-missing local amounts	557,003
Country is US, PR, VI, or missing	556,535
Zipcode is not missing and can be matched to a county	552,384

Notes: The table contains information on sample construction. All observations are at transaction level. Transaction level is the most granular level and includes information on amounts received by primary contractors, subcontractors, and vendors separately. Transactions are linked to places of performance and their five-digit zip codes.

Table 2.2: Variable Description and Data Sources

Variable	*Variable Description*	*Source*
County ID		US Census 2010 ZIP Code Tabulation Area (ZCTA) to County Data. D&B Unique Partner Identification Key (UPIK). Melissa Data. Zipcode Look-up. HUD USPS ZIP Code Crosswalk Files. 2010 Q1.
State ID		Same sources as County ID
CFDA number	Assigned by the federal government to indicate the purpose of spending	Catalog of Federal Domestic Assistance (CFDA)
Purpose of Spending	Assigned by the staff of Recovery.gov to indicate the purpose of spending, used to construct an alternative definition of R&D and science spending	Recovery.gov, Where did the money go? (downloaded on July 6, 2015)
Change in Employment, 2009-2013		Bureau of Labor Statistics. Quarterly Census of Employment and Wages. County High-Level Employment, 2000-2014.
Change in Private Sector Employment, 2009-2013		Same source as Change in Employment
Employment in Manufacturing, 2007	Pre-recession employment in manufacturing	Same source as Change in Employment
Change in Employment, 2007-2009	The severity of recession in a county	Same source as Change in Employment
Change in Employment Trend	The trend is a linear extrapolation of the predicted values from the regression of the change in employment on time. All regressions are calculated separately for each county from 2000 to 2009 in rolling five-year intervals.	Same source as Change in Employment
Employment in R&D, 2007	Pre-recession employment in R&D	US Census. North American Industry Classification System (2012). US Census. County Business Patterns, 2007-2014.
ARRA Spending on Research, 2009-2013		Federal Procurement Data System, Recovery Data, Recipient-reported. Cumulative National Summary Feb 17, 2009 - Dec 31, 2013 (downloaded on October 26, 2015).
All Other ARRA Spending, 2009-2013		Same source as ARRA Spending on Research
ARRA Spending on Research in Adjacent Counties, 2009-2013		Same source as ARRA Spending on Research US Census. County Adjacency, 2010.
Population		US Census. Annual Estimates of the Resident Population by County, 2010-2014. US Census. Annual Resident Population Estimates for States and Counties, 2000-2009. US Census. Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico, 2010-2014. US Census. Annual Estimates of the Population for the United States, Regions, States, and Puerto Rico, 2000-2009.
Population in Adjacent Counties		Same sources as Population. US Census. County Adjacency, 2010.
Doctoral Degrees	The number of doctoral degrees awarded in a county	The 2010 Edition of the Carnegie Classification (v. May 6, 2016)
Research University	Research-intensive universities as defined in the Carnegie Classification	Same source as Doctoral Degrees US Department of Education. Post-secondary University Survey, 2013 (inventory.data.gov)
Metropolitan County	Counties in metropolitan areas (population exceeding 100,000)	US Department of Agriculture. Rural-Urban Continuum Codes, 2013 (v. May 10, 2013)

Table 2.3: CFDA numbers for ARRA Research Spending.

CFDA Code	*Program Title*	*Agency*	*Amount, mln USD*
11.463	Habitat Conservation	Department of Commerce / National Oceanic and Atmospheric Administration (NOAA)	159
11.609	Measurement and Engineering Research and Standards	Department of Commerce / National Institute of Standards and Technology (NIST)	56
12.431	Basic Scientific Research	Department of Defense / U.S. Army Materiel Command	6
12.630	Basic, Applied, and Advanced Research in Science and Engineering	Department of Defense / Office of the Secretary of Defense	5
12.800	Air Force Defense Research Sciences Program	Department of Defense / Department of the Air Force, Materiel Command	6
15.807	Earthquake Hazards Research Grants	Department of the Interior / U.S. Geological Survey	8
15.817	National Geospatial Program: Building The National Map	Department of the Interior / U.S. Geological Survey	6
15.818	Volcano Hazards Program Research and Monitoring	Department of the Interior / U.S. Geological Survey	7
43.005	Exploration, Recovery Act	National Aeronautics and Space Administration	50
43.006	Science, Recovery Act	National Aeronautics and Space Administration	6
47.082	Trans-NSF Recovery Act Reasearch Support	National Science Foundation	2,913
81.049	Office of Science Financial Assistance Program	Department of Energy	481
81.086	Conservation Research and Development	Department of Energy	2,619
81.087	Renewable Energy Research and Development	Department of Energy	1,457
81.089	Fossil Energy Research and Development	Department of Energy	119
81.122	Electricity Delivery and Energy Reliability, Research, Development and Analysis	Department of Energy	4,889
81.135	Advanced Research Projects Agency - Energy	Department of Energy	351
93.420	ARRA - Community Health Applied Research Network	Department of Health and Human Services / Health Resources and Services Administration	10
93.701	Trans-NIH Recovery Act Research Support	Department of Health and Human Services / National Institutes of Health	8,282
93.702	National Center for Research Resources, Recovery Act Construction Support	Department of Health and Human Services / National Institutes of Health	814
93.715	Recovery Act Comparative Effectiveness Research - AHRQ	Department of Health and Human Services / Agency for Healthcare Research and Quality	308
93.726	ARRA Accelerating Adoption of Comparative Effectiveness Research (CER)	Department of Health and Human Services / National Institutes of Health	18
93.728	ARRA - Strategic Health IT Advanced Research Projects (SHARP)	Department of Health and Human Services / Office of the Secretary	76
93.730	ARRA Prevention Research Centers Comparative Effectiveness Research Program	Department of Health and Human Services / Centers for Disease Control and Prevention	10
Missing			3,729
Total			26,383

Notes. Our primary definition of research spending is based on the CFDA numbers. 89.4% of transactions linked to a US county in the Cumulative National Summary have a CFDA number. A total of 262 unique CFDA numbers are assigned to awards under ARRA. We classify 24 CFDA numbers in the table above as spending on research. We include transactions with missing CFDA numbers in our main definition if they were classified as research in the "Where Does the Money Go?" feature of Recovery.gov website.

Table 2.4: Comparison Table for Definitions of Research Spending

Agency	*CFDA*	*Recovery.gov*
Department of Energy	11,097.60	3,479.20
National Institutes of Health	9,827.35	9,793.66
National Science Foundation	2,966.76	2,731.93
National Aeronautics and Space Administration	867.47	867.47
Department of Health and Human Services (other than NIH)	733.22	645.81
Department of Commerce	525.32	524.37
Department of Defense	342.33	341.53
Department of Interior	21.92	0.04
Department of Homeland Security	0.48	0.00
Department of Transportation	0.30	66.49
General Services Administration	0.03	0.03
Department of Education	0.00	0.49
Total	26,382.77	18,451.01

Notes. The table contains comparison between the two definitions of research spending. The main definition uses CFDA numbers. The secondary definition comes from "Where does the money go?" section of the Recovery.gov website.

Department of Energy: The discrepancy between two definitions arises from including CFDA numbers 81.086 (Conservation R&D), 81.087 (Renewable Energy R&D), 81.122 (Electricity Delivery and Energy Reliability R&D and Analysis), and 81.135 (Advanced Research Projects Agency - Energy) in the main definition. These transactions are classified as "Energy" in the secondary definition. Additionally, seven transactions from CFDA number 81.049 (Office of Science Financial Assistance Program) are classified as "Energy" and "Other" in the secondary definition. Six awards under the CFDA number 81.126 (Federal Loan Guarantees for Innovative Energy Technologies) which is not part of the main definition, are included in the secondary definition.

National Institutes of Health: The discrepancy between two definitions arises from 30 NIH grants classified as "Other", "Health, and "Unemployment" (not "R&D and Science"), in the secondary definition.

National Science Foundation: The discrepancy between two definitions arises from 173 grants classified as "Infrastructure", "Transportation", "Education", or "Other" in the secondary definition.

Department of Health and Human Services: The discrepancy between two definitions arises from including CFDA numbers 93.420 (ARRA - Community Health Applied Research Network) and 93.728 (ARRA - Strategic Health IT Advanced Research Projects, SHARP) in the main definition. These transactions are classified as "Health" in the secondary definition. Additionally, one grant from CFDA number 93.726 (ARRA Accelerating Adoption of Comparative Effectiveness Research (CER)) is classified as "Other" in the secondary definition.

Department of Commerce: One grant from CFDA number 11.609 (Measurement and Engineering Research and Standards) is classified as "Infrastructure" in the secondary definition.

Department of Defense: The discrepancy between two definitions arises from three transactions under CFDA number 12.431 (Basic Scientific Research) which are classified as "Infrastructure" and "Other" in the secondary definition.

Department of Interior: The discrepancy between two definitions arises from including CFDA numbers 15.807 (Earthquake Hazard Research Grants), 15.817 (National Geospacial Program), and 15.818 (Volcano Hazards Program Research and Monitoring) in the main definition. These transactions are classified as "Energy" in the secondary definition.

Department of Homeland Security: The discrepancy between two definitions arises from including CFDA number 81.087 (Renewable Energy Research and Development) in the main definition. These transactions are classified as "Energy" in the secondary definition.

Department of Transportation: The discrepancy between two definitions arises from including 32 grants under CFDA number 20.205 (Highway Planning and Construction) in the secondary definition as "R&D and Science" funds. They are classified as non-R&D and Science spending in the main definition.

Department of Education: The discrepancy between two definitions arises from including two grants under 84.033 (Federal Work-Study Program) in the secondary definition as "R&D and Science" funds. They are classified as non-R&D and Science spending in the main definition.

Table 2.5: ARRA Awards Assigned by Year

	2009	2010	2011	2012	2013
ARRA R&D and science spending (mill USD)	14,438	9,107	610	275	221
All other ARRA spending (mill USD)	146,262	68,472	19,487	2,174	1,281

Notes: The data in the table comes from the Cumulative National Summary of ARRA Recipient Reports. The total amount constitutes about one-third of the total ARRA package. The data in the table is split by the year in which ARRA award was assigned. ARRA spending on research is defined using selected CFDA numbers. The discrepancy between Tables 2.5 and 2.6 comes from missing values in *award_date* variable.

Table 2.6: ARRA Awards Completion by Year

	2009	2010	2011	2012	2013
ARRA R&D and science spending (mill USD)	78	1,048	3,648	5,687	15,922
All other ARRA spending (mill USD)	3,874	20,660	59,855	54,535	110,398

Notes: The data in the table comes from the Cumulative National Summary of ARRA Recipient Reports. The total amount constitutes about one-third of the total ARRA package. The data in the table is split by the year in which recipients filed a final report upon work completion. ARRA spending on research is defined using selected CFDA numbers. The discrepancy between Tables 2.5 and 2.6 comes from missing values in *award_date* varible.

Table 2.7: Counties with the Largest ARRA Research Spending in 2009-2013.

County	State	Amount, USD	Percent of Total Amount
Los Angeles County	CA	867,029,440	3.35
Suffolk County	MA	790,944,576	3.06
Cook County	IL	770,771,648	2.98
New York County	NY	653,069,760	2.52
Middlesex County	MA	626,809,920	2.42
San Diego County	CA	546,829,824	2.11
Santa Clara County	CA	500,317,344	1.93
Alameda County	CA	467,040,160	1.81
Harris County	TX	445,125,856	1.72
Dallas County	TX	439,479,392	1.70
Wayne County	MI	419,185,472	1.62
Philadelphia County	PA	417,763,904	1.61
San Mateo County	CA	402,788,544	1.56
Wake County	NC	379,635,648	1.47
Milwaukee County	WI	349,739,072	1.35
Suffolk County	NY	348,517,472	1.35
Montgomery County	MD	342,431,232	1.32
Baltimore city	MD	333,228,384	1.29
Washtenaw County	MI	313,387,072	1.21
Hamilton County	OH	313,201,184	1.21
Maricopa County	AZ	300,493,120	1.16
Allegheny County	PA	295,211,648	1.14
Miami-Dade County	FL	290,971,552	1.12
King County	WA	283,775,200	1.10
District of Columbia	DC	258,183,168	1.00
Top 25	.	11,155,930,112	43.12
.	.		
.	.		
.	.		
Top 100	.	19,557,523,456	75.59

Notes. The table lists 25 counties with the largest awards from ARRA spending on research between 2009 and 2013. ARRA spending on research is defined using selected CFDA numbers.

Table 2.8: Counties with the Largest ARRA Research Spending per capita in 2009-2013.

County	State	Amount, USD per cap	Percent of Total Amount
Esmeralda County	NV	12,675	0.04
Bristol Bay Borough	AK	10,221	0.04
Marinette County	WI	3,258	0.53
Morrow County	OR	2,213	0.10
Anderson County	TN	1,831	0.52
Los Alamos County	NM	1,770	0.12
Pontotoc County	MS	1,734	0.19
Orange County	NC	1,345	0.66
Howard County	IN	1,276	0.41
Eureka County	NV	1,181	0.01
Lake and Peninsula Borough	AK	1,094	0.01
Suffolk County	MA	1,086	3.06
Tompkins County	NY	989	0.39
Washtenaw County	MI	912	1.21
Rutland County	VT	865	0.21
Grafton County	NH	843	0.28
Boulder County	CO	820	0.93
Durham County	NC	812	0.81
Graham County	NC	794	0.03
Centre County	PA	713	0.40
Noble County	OK	707	0.03
Albemarle County	VA	679	0.25
Schenectady County	NY	660	0.38
Fairbanks North Star Borough	AK	611	0.22
Pershing County	NV	608	0.02
Top 25	.		10.84
.	.		
.	.		
.	.		
Top 100	.		43.44

Notes. The table lists 25 counties with the largest per capita awards from ARRA spending on research between 2009 and 2013. ARRA spending on research is defined using selected CFDA numbers.

Table 2.9: Summary Statistics for Counties with ARRA Research Awards vs. All Counties

	ARRA, R&D and Science mean	All Counties mean
Employment (2013)	74,257	41,058
Population (2013)	177,643	100,782
Change in Employment per cap (2007-2009)	-0.0181	-0.0151
County with Research University	0.13	0.07
County with Employed in R&D (2007)	0.21	0.11
Metro County	0.55	0.37
Employed in Manufacturing per cap (2007)	0.051	0.046
Employed in R&D per cap (2007)	0.001	0.000
Doctoral Degrees in a County (2010)	37	19
Observations	1584	3102

Notes: The table contains summary statistics separately for counties with non-zero ARRA spending on research in comparison to all other counties. The unit of analysis is a county. ARRA spending on research is defined using selected CFDA numbers.
County with Research University is an indicator which equals one if a county has R1, R2, or R3 University by the definition in Carnegie Classification (2010). Metro County is an indicator variable which takes the value of one if a county is a Metropolitan County by the definition in 2013 Rural-urban Continuum Codes. Employed in R&D and Employed in Manufacturing in 2007 are scaled by the county population in 2007.

Table 2.10: ARRA Spending Dispersion (All Awards)

	All other, mean	R&D and Science, mean	diff	t-stat
Transactions (%)				
Different Zipcode	0.19	0.26	-0.07	-20.65
Different County	0.16	0.24	-0.08	-25.38
Different State	0.10	0.23	-0.12	-43.01
Amount(%)				
Different Zipcode	0.17	0.12	0.05	27.37
Different County	0.13	0.10	0.03	16.59
Different State	0.03	0.08	-0.05	-34.06

Notes. The table shows the dispersion of ARRA research awards in comparison to all other ARRA awards. The data are summarized using all awards, including awards with one recipient (primary contractor). The top panel shows the percent of transactions with place of performance (POP) outside primary contractor's zip code, county, and state. The bottom panel shows the percent of award amount received by the subcontractors and vendors registered outside primary contractor's zip code, county, and state. ARRA spending on research is defined using selected CFDA numbers.

Table 2.11: ARRA Spending Dispersion (Awards with at least one vendor or subcontractor)

	All other, mean	R&D and Science, mean	diff	t-stat
Transactions (%)				
Different Zipcode	0.57	0.85	-0.28	-53.43
Different County	0.49	0.80	-0.31	-55.63
Different State	0.30	0.75	-0.45	-76.36
Amount(%)				
Different Zipcode	0.50	0.38	0.12	27.71
Different County	0.39	0.34	0.06	12.70
Different State	0.10	0.27	-0.17	-45.35

Notes. The table shows the dispersion of ARRA research awards in comparison to all other ARRA awards. The data excludes awards with one recipient (primary contractor). The top panel shows the percent of transactions with place of performance (POP) outside primary contractor's zip code, county, and state. The bottom panel shows the percent of award amount received by the subcontractors and vendors registered outside primary contractor's zip code, county, and state. ARRA spending on research is defined using selected CFDA numbers.

Table 2.12: County Summary Statistics (2009-2013)

	count	mean	sd	min	max
Outcome Variables					
Annual Change in Total Employment per cap, 2009-2013	3102	0.00763	0.0340	-0.320	0.576
Annual Change in Private Sector Employment per cap, 2009-2013	3102	0.01023	0.0336	-0.302	0.570
ARRA Spending Variables					
ARRA Research Spending (mill per cap)	3102	0.00003	0.0003	0.000	0.013
ARRA Research Spending in Adjacent Counties (mill per cap)	3102	0.00005	0.0001	0.000	0.002
All Other ARRA Spending (mill per cap)	3102	0.00082	0.0017	0.000	0.047
Instrumental Variable					
Employed in R&D per cap (2007)	3102	0.00030	0.0022	0.000	0.082
County with Employed in R&D (2007)	3102	0.11348	0.3172	0.000	1.000
Doctorate Degrees in a County (2010)	3102	19	109	0.000	2361
County with Research University	3102	0.06544	0.2473	0.000	1.000
Inverse Mills Ratio	1584	0.56940	0.4283	0.000	1.846
Control Variables					
Metro County	3102	0.37105	0.4832	0.000	1.000
Change in Total Employment per cap, 2007-2009	3102	-0.01511	0.0216	-0.180	0.220
Change in Private Sector Employment per cap, 2007-2009	3102	-0.01599	0.0230	-0.279	0.241
Employed in Manufacturing per cap (2007)	3102	0.04626	0.0436	0.000	0.479

Notes: The table contains summary statistics for baseline regression. The unit of analysis is a county. The sample size is all US counties for all variables, except inverse Mills ratio. The sample size for inverse Mills ratio is all counties with non-zero ARRA research awards. The time period for outcome variables and ARRA spending variables is 2009-2013. All outcome variables and ARRA spending variables are scaled by county population averaged over the same time period. All ARRA spending variables are in millions of USD. ARRA spending on research is defined using selected CFDA numbers.

County with Research University is an indicator which equals one if a county has R1, R2, or R3 University by the definition in Carnegie Classification (2010). The inverse Mills ratio is constructed using predicted values from the probit regression of the probability of getting ARRA research awards on control variables and two instruments: County with Research University and County with Employed in R&D.

Metro County is an indicator variable which takes the value of one if a county is a Metropolitan County by the definition in 2013 Rural-urban Continuum Codes.

The change in the total number of employed in a county between 2007 and 2009 and the change in the number of employed in the private sector in a county between 2007 and 2009 are scaled by the average population in a county during this period.

Employed in R&D and Employed in Manufacturing in 2007 are scaled by the county population in 2007.

Table 2.13: Heckman Correction Results

County with ARRA Research Spending in 2009-2013	Baseline dy/dx	Sec Definition dy/dx	Private Sector dy/dx
County with Employed in R&D and Science (2007)	0.334***	0.286***	0.331***
	(0.0441)	(0.0328)	(0.0441)
County with Research University	0.547***	0.543***	0.546***
	(0.132)	(0.0857)	(0.132)
Metro County	0.240***	0.241***	0.236***
	(0.0159)	(0.0138)	(0.0160)
ARRA Research Spending in Adjacent Counties (mln per cap, 2009-2013)	391.8**	257.0**	402.9**
	(159.9)	(119.3)	(164.4)
All Other ARRA Spending (mln per cap, 2009-2013)	4.185	3.398	3.143
	(4.154)	(3.159)	(4.063)
Employed in Manufacturing per cap (2007)	0.937***	0.706***	0.842***
	(0.207)	(0.180)	(0.208)
Change in Employment per cap (2007-2009)	-1.083***	-0.882**	-1.562***
	(0.419)	(0.372)	(0.446)
County-specific Change in Employment Trend	0.359***	0.293**	0.507***
	(0.127)	(0.114)	(0.141)
State FE	Yes	Yes	Yes
Observations	3102	3102	3056

Notes: The table contains Heckman correction for the selection of counties into receiving ARRA research awards. The first column shows baseline regression. ARRA spending on research is defined using selected CFDA numbers. The second column shows the same regression using secondary definition of research spending based on “Where does the money go?” section of the Recovery.gov website. The third column shows regression using private sector employment, instead of total employment, in the county-specific change in employment trend and the change in employment per cap (2007-2009).
In all regressions, a county is the unit of analysis. The sample includes all US counties. The outcome variable is a dummy variable for a county with non-zero ARRA spending on research in 2009-2013.
ARRA spending on research in adjacent counties and all other ARRA spending are included as controls. The variables are in millions of USD over 2009-2013, divided by the population in a given county averaged over the same period of time.
County with Research University is an indicator which equals one if a county has R1, R2, or R3 University by the definition in Carnegie Classification (2010). Metro County is an indicator variable which takes the value of one if a county is a Metropolitan County by the definition in 2013 Rural-urban Continuum Codes. We control for the number of people employed in manufacturing per capita in a county in 2007. We also include the change in the total number of employed workers in a county between 2007 and 2009 divided by the average population in a county during this period.
Each regression includes state fixed effects and a county-specific change in employment trend. The trend is a linear extrapolation of the predicted values from the regression of the change in employment on time. All regressions are calculated separately for each county from 2000 to 2009 in rolling five-year intervals.
Robust standard errors in parentheses.
* $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 2.14: First-Stage Results

ARRA Research Spending (mln per cap, 2009-2013)	Baseline	Sec Definition	Private Sector
Employed in R&D and science per cap (2007)	0.0153***	0.0147***	0.0159***
	(0.00311)	(0.00350)	(0.00308)
Doctoral Degrees in a county (2010, ln)	0.0000185***	0.0000161***	0.0000204***
	(0.00000447)	(0.00000372)	(0.00000359)
Inverse Mills Ratio	0.0000350	0.0000214	0.0000340
	(0.0000271)	(0.0000146)	(0.0000213)
Metro County	-0.00000813	-0.00000587	0.00000215
	(0.0000194)	(0.0000139)	(0.0000142)
ARRA Research Spending in Adjacent Counties (mln per cap, 2009-2013)	0.0181	-0.0148	0.0872
	(0.0899)	(0.0774)	(0.0652)
All Other ARRA Spending (mln per cap, 2009-2013)	0.0153	0.0116	0.00887
	(0.0100)	(0.00815)	(0.00543)
Employed in Manufacturing per cap (2007)	0.000202	0.0000705	0.000218
	(0.000232)	(0.000175)	(0.000224)
Change in Employment per cap (2007-2009)	-0.000206	-0.000538	-0.000741
	(0.000718)	(0.000587)	(0.000700)
County-specific Change in Employment Trend	0.000645**	0.000427	0.000432
	(0.000320)	(0.000288)	(0.000295)
Constant	-0.0000622	-0.0000689	-0.000108
	(0.0000535)	(0.0000449)	(0.0000663)
State FE	Yes	Yes	Yes
Observations	1584	1204	1570
R-sq	0.13	0.20	0.21

Notes: The table contains the first stage of IV regressions. The first column shows baseline regression. ARRA spending on research is defined using selected CFDA numbers. The second column shows the same regression using secondary definition of research spending based on "Where does the money go?" section of the Recovery.gov website. The third column shows regression using private sector employment, instead of total employment, in the county-specific change in employment trend and the change in employment per cap (2007-2009). In all regressions, a county is the unit of analysis. The sample includes all US counties receiving ARRA Research stimulus in 2009-2013. The outcome variable is ARRA spending on research from 2009 to 2013. ARRA spending on research in adjacent counties and all other ARRA spending are included as controls. All three variables are in millions of USD over 2009-2013, divided by the population in a given county averaged over the same period of time.

ARRA spending on research is an endogeneous variable. It is instrumented by a natural logarithm of the number of doctoral degrees awarded at the universities in a county, the number of individuals employed in R&D and scientific services per capita in 2007, and a Heckman correction term. The Heckman correction term is an inverse Mills ratio of predicted values from the probit regression of the probability of receiving ARRA research stimulus on all control variables and two other instruments: dummies for a county with research university and a county with employed in R&D and science in 2007.

Metro County is an indicator variable which takes the value of one if a county is a Metropolitan County by the definition in 2013 Rural-urban Continuum Codes. We also control for the number of people employed in manufacturing per capita in a county in 2007 and the change in the total number of employed workers in a county between 2007 and 2009 divided by the average population in a county during this period.

Each regression includes state fixed effects and a county-specific change in employment trend. The trend is a linear extrapolation of the predicted values from the regression of the change in employment on time. All regressions are calculated separately for each county from 2000 to 2009 in rolling five-year intervals.

Robust standard errors in parentheses.

* $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 2.15: Total Employment Baseline Results

Change in Employment per cap, 2009-2013	OLS	IV Baseline	IV Sec def	IV Priv sec
ARRA Research Spending (mill per cap, 2009-2013)	14.14***	26.75***	27.34***	22.95***
	(1.289)	(7.949)	(9.747)	(6.615)
ARRA Research Spending in Adjacent Counties (mill per cap, 2009-2013)	0.319	0.984**	1.202**	0.915**
	(0.408)	(0.390)	(0.561)	(0.396)
All Other ARRA Spending (mill per cap, 2009-2013)	-0.522	-1.484**	-1.589**	-1.219**
	(0.358)	(0.578)	(0.672)	(0.474)
Change in Employment per cap (2007-2009)	-0.0785	-0.0124	0.0403	-0.0469
	(0.0908)	(0.0652)	(0.0724)	(0.0690)
Employed in Manufacturing per cap (2007)	0.0268	0.0780**	0.0981***	0.0665**
	(0.0286)	(0.0327)	(0.0360)	(0.0302)
Metro County	0.00669***	0.00673***	0.00814***	0.00562***
	(0.00139)	(0.00186)	(0.00171)	(0.00180)
County-specific change in employment trend	0.0215	0.0278	0.0187	0.0217
	(0.0214)	(0.0329)	(0.0318)	(0.0348)
Constant	0.0157***			
	(0.00476)			
State FE	Yes	Absorbed	Absorbed	Absorbed
Observations	3102	1584	1204	1570
Robust First-Stage F		17.39	20.64	27.81
Overidentification test		0.99	0.21	0.88

Notes: The first column shows endogeneous OLS regression. The second column shows the second stage of the baseline IV regression. ARRA spending on research is defined using selected CFDA numbers. The third column shows the same regression as the one in the second column using secondary definition of research spending based on "Where does the money go?" section of the Recovery.gov website. The fourth column shows regression with private sector employment, instead of total employment, in the outcome variable as well as the county-specific change in employment trend and the change in employment per cap (2007-2009).
In all regressions, a county is the unit of analysis. The sample in the first column includes all US counties. The sample in columns 2-4 includes all US counties receiving ARRA research stimulus in 2009-2013. The outcome variable is the change in employment from 2009 to 2013 divided by the population averaged over the same period of time.
ARRA spending on research is an exogeneous variable in the first column and an endogeneous variable in columns 2-4. It is instrumented by a natural logarithm of the number of doctoral degrees awarded at the universities in a county, the number of individuals employed in R&D and scientific services per capita in 2007, and a Heckman correction term. The Heckman correction term is an inverse Mills ratio of the predicted values from the probit regression of the probability of receiving ARRA research stimulus on all control variables and two other instruments: dummies for a county with research university and a county with employed in R&D and science in 2007. ARRA spending on research in adjacent counties and all other ARRA spending are included as controls. All three variables are in millions of USD paid in 2009-2013, divided by the population averaged over the same period of time. Additionally, ARRA spending on research in adjacent counties is scaled by the average of the ratio of the population in adjacent counties to the population of the focal county to simplify coefficient interpretation.
County with research university is an indicator which equals one if a county has R1, R2, or R3 University by the definition in Carnegie Classification (2010). Metro County is an indicator variable which takes the value of one if a county is a Metropolitan County by the definition in 2013 Rural-urban Continuum Codes. We also control for the number of people employed in manufacturing per capita in a county in 2007 and the change in the total number of employed workers in a county between 2007 and 2009, divided by the average population in a county during this period.
OLS regression in Column 1 includes state fixed effects. Instrumental variable regressions in Columns 2-4 have state fixed effects absorbed to allow for standard errors to be clustered at the state level.
Each regression includes a change in employment trend. The trend is a linear extrapolation of the predicted values from the regression of the change in employment on time. All regressions are calculated separately for each county from 2000 to 2009 in rolling five-year intervals.
Robust standard errors clustered on state are in parentheses.
* $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 2.16: Total Employment Results, build the model

Change in Employment per cap, 2009-2013	1	2	3	4	5	6	7	Baseline
ARRA Research	35.88***	33.96***	38.03***	37.97***	41.54***	23.30***	21.56**	26.75***
Spending (mill per cap, 2009-2013)	(11.54)	(10.93)	(10.46)	(10.71)	(12.69)	(8.582)	(8.439)	(7.949)
ARRA Research		0.640	0.594	0.582	0.600	0.422	0.437	0.984**
Spending in Adjacent Counties (mill per cap, 2009-2013)		(0.533)	(0.540)	(0.560)	(0.578)	(0.479)	(0.476)	(0.390)
All Other ARRA			-1.412**	-1.452**	-1.486**	-0.945**	-0.968*	-1.484**
Spending (mill per cap, 2009-2013)			(0.638)	(0.634)	(0.631)	(0.477)	(0.519)	(0.578)
Change in Employment				0.0675	0.160	0.152	0.119	-0.0124
per cap (2007-2009)				(0.103)	(0.102)	(0.0941)	(0.0893)	(0.0652)
Employed in					0.0924***	0.0955***	0.0954***	0.0780**
Manufacturing per cap (2007)					(0.0326)	(0.0302)	(0.0302)	(0.0327)
Metro County						0.00644***	0.00619***	0.00673***
						(0.00162)	(0.00162)	(0.00186)
County-specific							0.0286	0.0278
change in employment trend							(0.0318)	(0.0329)
State FE	No	No	No	No	No	No	No	Absorbed
Observations	1584	1584	1584	1584	1584	1584	1584	1584
Robust First-Stage F	19.53	20.75	17.85	17.53	14.92	39.39	37.26	17.39
Overidentification test	0.49	0.45	0.53	0.44	0.30	0.36	0.40	0.99

Notes: The table contains the second stage of the baseline IV regressions with omitted controlled variables gradually building the full model in the last column. The unit of analysis is a county. ARRA spending on research is defined using selected CFDA numbers.
Robust standard errors clustered on state are in parentheses.
p<0.1, ** p<0.05, *** p<0.01.

Table 2.17: Total Employment Results without Selection Stage

Change in Employment per cap, 2009-2013	IV Baseline	IV Sec def	IV Priv sec
ARRA Research Spending (mill per cap, 2009-2013)	35.65***	38.30***	31.78***
	(9.473)	(12.36)	(7.729)
ARRA Research Spending in Adjacent Counties (mill per cap, 2009-2013)	0.265	0.666	0.204
	(0.393)	(0.506)	(0.371)
All Other ARRA Spending (mill per cap, 2009-2013)	-0.622	-0.388	-0.555
	(0.388)	(0.490)	(0.366)
Change in Employment per cap (2007-2009)	-0.0705	-0.0711	-0.0999
	(0.0904)	(0.0906)	(0.103)
Employed in Manufacturing per cap (2007)	0.0249	0.0268	0.0157
	(0.0279)	(0.0286)	(0.0291)
Metro County	0.00624***	0.00624***	0.00503***
	(0.00138)	(0.00139)	(0.00133)
County-specific change in employment trend	0.0150	0.0175	0.00544
	(0.0225)	(0.0231)	(0.0266)
State FE	Absorbed	Absorbed	Absorbed
Observations	3102	3102	3056
Robust First-Stage F	47.62	46.76	55.07
Overidentification test	0.57	0.41	0.86

Notes: The table contains the second stage of the IV regressions in Table 2.15 on a sample of all US counties with omitted selection stage. The unit of analysis is a county. ARRA spending on research is defined using selected CFDA numbers.
Robust standard errors clustered on state are in parentheses.
p<0.1, ** p<0.05, *** p<0.01.

Table 2.18: Split Sample Results

Change in Employment per cap, 2009-2013	No California, Massachusetts	Counties without Research Universities	No Nevada, Alaska	No Top 25 ARRA Research Counties
ARRA Research Spending (mill per cap, 2009-2013)	25.24*** (7.796)	17.95*** (6.754)	26.78*** (7.413)	21.39*** (6.965)
ARRA Research Spending in Adjacent Counties (mill per cap, 2009-2013)	0.867** (0.383)	0.710** (0.356)	0.852** (0.374)	0.915** (0.388)
All Other ARRA Spending (mill per cap, 2009-2013)	-1.459*** (0.567)	-1.372*** (0.521)	-1.555*** (0.593)	-1.391*** (0.524)
Change in Employment per cap (2007-2009)	-0.0108 (0.0643)	-0.0133 (0.0703)	-0.0181 (0.0700)	-0.00435 (0.0634)
Employed in Manufacturing per cap (2007)	0.0765** (0.0329)	0.0785** (0.0341)	0.0791** (0.0366)	0.0809** (0.0323)
Metro County	0.00660*** (0.00191)	0.00689*** (0.00179)	0.00687*** (0.00190)	0.00657*** (0.00186)
County-specific change in employment trend	0.0242 (0.0323)	0.0293 (0.0325)	0.0318 (0.0337)	0.0291 (0.0316)
State FE	Absorbed	Absorbed	Absorbed	Absorbed
Observations	1523	1382	1560	1559
Robust First-Stage F	14.56	13.69	28.35	11.63
Overidentification test	0.96	0.86	0.99	1.00

Notes: The table shows the second stage of the baseline IV regression on split samples. The sample in the first column excludes counties in Massachusetts and California. The sample in the second column excludes all counties with research universities. The sample in the third column excludes counties in Nevada and Alaska. Finally, the sample in the fourth column excludes the top 25 counties with the highest ARRA research spending.
In all regressions, a county is the unit of analysis. The sample includes the counties with non-zero ARRA research awards. ARRA spending on research is defined using selected CFDA numbers. The outcome variable is the change in employment from 2009 to 2013 divided by the population averaged over the same period of time.
ARRA spending on research is an endogeneous variable. It is instrumented by the natural logarithm of the number of doctoral degrees awarded at the universities located in a county, the number of individuals employed in R&D and scientific services per capita in 2007, and a Heckman correction term. The Heckman correction term is an inverse Mills ratio of predicted values from the probit regression of the probability of receiving ARRA research stimulus on all control variables and two other instruments: dummies for a county with research university and a county with employed in R&D and science in 2007. The dummy for a county with research university is omitted from the construction of Heckman correction term because the sample excludes counties with research universities. County with Research University is an indicator which equals one if a county has R1, R2, or R3 University by the definition in Carnegie Classification (2010).
ARRA spending on research in adjacent counties and all other ARRA spending are included as controls. All three variables are in millions of USD paid in 2009-2013, divided by the population averaged over the same period of time. Additionally, ARRA spending on research in adjacent counties is scaled by the average of the ratio of the population in adjacent counties to the population of the focal county to simplify coefficient interpretation.
The state fixed effects are absorbed to allow for standard errors to be clustered at the state level.
Metro County is an indicator variable which takes the value of one if a county is a Metropolitan County by the definition in 2013 Rural-urban Continuum Codes. We also control for the number of people employed in manufacturing per capita in a county in 2007 and the change in the total number of employed workers in a county between 2007 and 2009, divided by the average population in a county during this period.
Each regression includes a change in employment trend. The trend is a linear extrapolation of the predicted values from the regression of the change in employment on time. All regressions are calculated separately for each county from 2000 to 2009 in rolling five-year intervals.
Robust standard errors clustered on state are in parentheses.
* p<0.1, ** p<0.05, *** p<0.01.

Table 2.19: Change in Total Employment from 2009

Change in Employment per cap	2009-2010	2009-2011	2009-2012	2009-2013
ARRA Research Spending	4.219	16.37***	18.87***	26.75***
(mill per cap)	(4.836)	(4.806)	(5.455)	(7.949)
ARRA Research Spending in	0.0546	0.326	0.796***	0.984**
Adjacent Counties (mill per cap)	(0.158)	(0.254)	(0.284)	(0.390)
Observations	1584	1584	1584	1584
Robust First-Stage F	21.71	17.16	18.01	17.39
Overidentification test	0.02	0.05	0.73	0.99

Notes: The table contains the regression coefficients used in calculating the cost of a job-year. Each column is the second stage of the baseline IV regression on a different outcome variable. The outcome variable is the change in employment over the time period at the top of each column divided by the population in a given county averaged over the same period of time.
The unit of analysis is a county. ARRA spending on research is defined using selected CFDA numbers. The sample includes all US counties with non-zero ARRA research awards.
Robust standard errors clustered on state are in parentheses.
p<0.1, ** p<0.05, *** p<0.01.

Table 2.20: Change in Total Employment from 2009 (Endogeneous Regression)

Change in Employment per cap	2009-2010	2009-2011	2009-2012	2009-2013
ARRA Research Spending	2.072***	4.870***	8.643***	14.14***
(mill per cap)	(0.491)	(0.585)	(0.714)	(1.289)
ARRA Research Spending in	-0.00389	0.0121	0.277	0.319
Adjacent Counties (mill per cap)	(0.151)	(0.286)	(0.353)	(0.408)
Observations	3102	3102	3102	3102
R-sq	0.06	0.09	0.11	0.12

Notes: The table contains the regression coefficients used in calculating the cost of a job-year. Each column is an endogeneous regression as in Column 1 of Table 2.15 on a different outcome variable. The outcome variable is the change in employment over the time period at the top of each column divided by the population in a given county averaged over the same period of time.
The unit of analysis is a county. ARRA spending on research is defined using selected CFDA numbers. The sample includes all US counties.
Robust standard errors in parentheses.
* p<0.1, ** p<0.05, *** p<0.01.

Table 2.21: State-Level Results

Change in Employment per cap, 2009-2013	OLS	IV Baseline	IV Sec def	IV Priv sec
ARRA Research Spending (mill per cap, 2009-2013)	17.68	51.65	41.09	33.76
	(58.31)	(96.06)	(86.38)	(90.62)
All Other ARRA Spending (mill per cap, 2009-2013)	6.928	4.832	6.450	4.481
	(4.992)	(6.608)	(4.410)	(6.096)
Change in Employment per cap (2007-2009)	-0.0843	-0.0533	-0.0809	-0.0942
	(0.397)	(0.343)	(0.340)	(0.352)
Employed in Manufacturing per cap (2007)	-0.222	-0.233	-0.194	-0.235
	(0.229)	(0.195)	(0.202)	(0.184)
Number of Metro Counties in a State	0.0000284	0.0000203	0.0000393	0.0000433
	(0.000168)	(0.000143)	(0.000144)	(0.000136)
State-specific change in employment trend	0.213	0.201	0.200	0.201
	(0.150)	(0.130)	(0.135)	(0.127)
Constant	0.0117	0.0116	0.00956	0.0153
	(0.0152)	(0.0121)	(0.0123)	(0.0117)
Region FE	Yes	Yes	Yes	Yes
Observations	51	51	51	51
Robust First-Stage F		12.82	21.15	13.06

Notes: The first column shows endogeneous OLS regression. The second column shows the second stage of the baseline IV regression. ARRA spending on research is defined using selected CFDA numbers. The third column shows the same regression using secondary definition of research spending based on "Where does the money go?" section of the Recovery.gov website. The fourth column shows regression with private sector employment, instead of total employment, in the outcome variable as well as the state-specific change in employment trend and the change in employment per cap (2007-2009).

In all regressions, a state is the unit of analysis. The outcome variable is the change in employment from 2009 to 2013 divided by the population averaged over the same period of time.

ARRA spending on research is an exogeneous variable in the first column and an endogeneous variable in columns 2-4. It is instrumented by the number of doctoral degrees awarded at the universities in a state and the number of individuals employed in R&D and scientific services per capita in 2007.

All other ARRA spending is included as controls. The two spending variables are in millions of USD paid in 2009-2013, divided by the population averaged over the same period of time.

The Number of Metro Counties in a State is a count variable capturing the number of counties corresponding to the definition of Metropolitan County in 2013 Rural-urban Continuum Codes. We also control for the number of people employed in manufacturing per capita in a state in 2007 and the change in the total number of employed workers in a state between 2007 and 2009, divided by the average population in a state during this period.

Each regression includes region fixed effects and a change in employment trend. The trend is a linear extrapolation of the predicted values from the regression of the change in employment on time. All regressions are calculated separately for each state from 2000 to 2009 in rolling five-year intervals.

Conventional standard errors in parentheses.

* $p<0.1$, ** $p<0.05$, *** $p<0.01$.

BIBLIOGRAPHY

Acemoglu, D. and J. D. Angrist (2001). Consequences of employment protection? the case of the americans with disabilities act. *Journal of Political Economy 109*(5), 915–957.

Arcidiacono, P. (2018). Expert report for students for fair admissions, inc. v. harvard. Technical report.

Auerbach, A. J. and Y. Gorodnichenko (2012). Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy 4*(2), 1–27.

Baxter, M. and R. G. King (1993). Fiscal policy in general equilibrium. *The American Economic Review*, 315–334.

Blinder, A. S. and M. M. Zandi (2010). *How the great recession was brought to an end.*

Boehm, C. E. (2016). Government spending and durable goods. *Working paper*.

Card, D. (2017). Expert report for students for fair admissions, inc. v. harvard. Technical report.

Chodorow-Reich, G. (2017). Geographic cross-sectional fiscal spending multipliers: What have we learned? *National Bureau of Economic Research*.

Chodorow-Reich, G., L. Feiveson, Z. Liscow, and W. G. Woolston (2012). Does state fiscal relief during recessions increase employment? Evidence from the American Recovery and Reinvestment Act. *American Economic Journal: Economic Policy 4*(3), 118–45.

Clark, C. S. (2015, Sep). Historic effort to track stimulus spending wraps up. *Government Executive*.

Congressional Budget Office (2012, February). *Estimated impact of the American Recovery and Reinvestment Act on employment and economic output from October 2011 through December 2011.* Congressional Budget Office.

Conley, T. G. and B. Dupor (2013). The American Recovery and Reinvestment Act: Solely a government jobs program? *Journal of monetary Economics 60*(5), 535–549.

D'Angelo, L. (2016). 2016 top list of black models. *Vogue Italia*. Accessed: 2017-10-30.

DeLeire, T. (2000). The wage and employment effects of the americans with disabilities act. *Journal of Human Resources*, 693–715.

Dezső, C. L., D. G. Ross, and J. Uribe (2016). Is there an implicit quota on women in top management? a large-sample statistical analysis. *Strategic Management Journal 37*(1), 98–115.

Dinerstein, M. F., C. M. Hoxby, J. Meer, and P. Villanueva (2014). Did the fiscal stimulus work for universities? In *How the Financial Crisis and Great Recession Affected Higher Education*, pp. 263–320. University of Chicago Press.

Dustmann, C., A. Glitz, U. Schönberg, and H. Brücker (2015). Referral-based job search networks. *The Review of Economic Studies 83*(2), 514–546.

Feyrer, J. and B. Sacerdote (2011). Did the stimulus stimulate? Real time estimates of the effects of the American Recovery and Reinvestment Act. *National Bureau of Economic Research*.

Fryer, Jr, R. G. (2009). Implicit quotas. *The Journal of Legal Studies 38*(1), 1–20.

Giuliano, L., D. I. Levine, and J. Leonard (2009). Manager race and the race of new hires. *Journal of Labor Economics 27*(4), 589–631.

Godart, F. C. and A. Mears (2009). How do cultural producers make creative decisions? lessons from the catwalk. *Social Forces 88*(2), 671–692.

Harmon, K. (2010, Feb). Is the Recovery Act stimulating science and the economy? *Scientific American*.

Hausman, N. (2012). University innovation, local economic growth, and entrepreneurship.

Hourihan, M. (2015). Historical trends in federal R&D. *Washington DC: The American Association for the Advancement of Science.(AAAS Report XXXIX, Research and Development FY 2014)*, 23–28.

Jones, G. and D. M. Rothschild (2011). Did stimulus dollars hire the unemployed? Answers to questions about the American Recovery and Reinvestment Act. *Working paper*.

Kanter, R. M. (1977a). *Men and women of the corporation.* Basic Books.

Kanter, R. M. (1977b). Some effects of proportions on group life: Skewed sex ratios and responses to token women. *American Journal of Sociology 82*(5), 965–990.

Keynes, J. M. (1936). *General theory of employment, interest and money.* Macmillan.

Lang, K. (1986). A language theory of discrimination. *The Quarterly Journal of Economics 101*(2), 363–382.

Leduc, S. and D. Wilson (2013). Are state governments roadblocks to federal stimulus? Evidence from highway grants in the 2009 Recovery Act. *Federal Reserve Bank of San Francisco Working Paper 16*.

Leonard, J. S. (1990). The impact of affirmative action regulation and equal employment law on black employment. *The Journal of Economic Perspectives 4*(4), 47–63.

Mau, D. (2014). How much it costs to show at new york fashion week? *Fashionista*. Accessed: 2017-10-30.

Mears, A. (2008). Discipline of the catwalk: Gender, power and uncertainty in fashion modeling. *Ethnography 9*(4), 429–456.

Mears, A. (2010). Size zero high-end ethnic: Cultural production and the reproduction of culture in fashion modeling. *Poetics 38*, 21–46.

Mears, A. (2011a). *Pricing beauty: The making of a fashion model.* Univ of California Press.

Mears, A. (2011b). Pricing looks: Circuits of value in fashion modeling markets. In *The worth of goods: Valuation and pricing in the economy.*

Moore-Karim, A. (2016). How "black models matter" became more than a viral street style message. *Fashionista*. Accessed: 2017-10-30.

Nakamura, E. and J. Steinsson (2014). Fiscal stimulus in a monetary union: Evidence from US regions. *American Economic Review 104*(3), 753–92.

Nisita, L. (2013). What models really get paid for fashion week. *Refinery29*. Accessed: 2017-11-13.

Odell, A. (2013). Fashion models finally earn money after pleas for cash. *BuzzFeed*. Accessed: 2017-10-30.

Oyer, P. and S. Schaefer (2002). Sorting, quotas, and the civil rights act of 1991: who hires when it's hard to fire? *The Journal of Law and Economics 45*(1), 41–68.

Oyer, P. and S. Schaefer (2010). Personnel economics: hiring and incentives. Technical report, National Bureau of Economic Research.

Oyer, P. and S. Schaefer (2012). Firm/employee matching: An industry study of american lawyers. Technical report, National Bureau of Economic Research.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The American Economic Review 62*(4), 659–661.

Pilkington, E. (2007). Supermodels launch anti-racism protest. *The Guardian*. Accessed: 2017-10-30.

Reich, M., D. M. Gordon, and R. C. Edwards (1973). A theory of labor market segmentation. *The American Economic Review 63*(2), 359–365.

Safronova, V., J. Nikas, and N. Osipova (2017). What it's truly like to be a fashion model. *The New York Times*. Accessed: 2017-10-30.

Sauers, J. (2013). Dior, chanel, prada, gucci, ysl: Who's the most diverse? Accessed: 2017-10-30.

Schumpeter, J. A. (1942). *Capitalism, socialism and democracy.* Harper and Brothers.

Serrato, J. C. S. and P. Wingender (2016). Estimating local fiscal multipliers. *National Bureau of Economic Research*.

Sørensen, J. B. (2004). The organizational demography of racial employment segregation. *American Journal of Sociology 110*(3), 626–671.

Tai, C. (2017). Diversity report: The fall 2017 ads were more inclusive than the runways for the first time ever. *The Fashion Spot*. Accessed: 2017-10-30.

Teeman, T. (2013). Chanel iman: modelling, racism, and me. Accessed: 2018-12-28.

von Furstenberg, D. and S. Kolb (2016). New york fashion week memo. *CFDA*. Accessed: 2017-10-30.

Weinberg, B. A., J. Owen-Smith, R. F. Rosen, L. Schwarz, B. M. Allen, R. E. Weiss, and J. Lane (2014). Science funding and short-term economic activity. *Science 344*(6179), 41–43.

Welch, F. (1976). Employment quotas for minorities. *Journal of Political Economy 84*(4, Part 2), S105–S141.

Williams, K. Y. and C. A. O'Reilly III (1998). Demography and diversity in organizations: A review of 40 years of research. *Research in Organizational Behavior 20*, 77–140.

Wilson, D. J. (2012). Fiscal spending jobs multipliers: Evidence from the 2009 American Recovery and Reinvestment Act. *American Economic Journal: Economic Policy 4*(3), 251–82.

Woodford, M. (2011). Simple analytics of the government expenditure multiplier. *American Economic Journal: Macroeconomics 3*(1), 1–35.