Superstar Earners and Market Size: Evidence from the Roll-Out of TV∗

Felix Koenig§
London School of Economics

JOB MARKET PAPER

December 2018
(Latest Version)

Abstract

This paper exploits a historic natural experiment to test a leading explanation for top income growth, the superstar theory. The superstar theory attributes rising top incomes to expanding market reach. The theory can be tested during a drastic expansion in market reach that occurred in entertainment during the roll-out of TV in the mid-twentieth century. At the time, TV stations predominantly broadcast shows that were filmed locally live. Exogenous variation in the local deployment of TV filming allows me to identify the effect on the entertainer wage distribution by comparing local labor markets differentially exposed to local TV show filming. I analyze newly collected data on local entertainer labor markets and find that markets where a TV station is launched experience a sharp increase in income concentration at the top of the distribution. Wages of stars grow nearly 20% in response to a fourfold increase in market reach. A distinctive pattern of wage changes allows me to distinguish the superstar model from alternative explanations. Growth of top pay occurs simultaneously with widening income differences at the top, a decline in middle-income jobs, an increase in low-paid jobs and a fall in total entertainer employment. Labor market institutions that might mediate top income growth have little impact on the magnitude of superstar effects. Only employer monopsony power dampens growth of top wages significantly.

Keywords: Superstar Effect, Inequality, Top Incomes, Market Concentration
JEL classification: J31, J23, J24, M52, O33, D31

∗I thank David Autor, Alan Manning and Steve Pischke for their guidance and advice. I also thank Daron Acemoglu, Josh Angrist, Jan Bakker, Oriana Bandiera, Daniel Chandler, Jeremiah Dittmar, Thomas Drechsel, Horst Entorf, George Fenton, Andy Ferrara, Torsten Figueiredo Walter, Xavier Gabaix, Killian Huber, Simon Jäger, Philipp Kircher, Camille Landais, Matt Lowe, Ben Moll, Nicas Moneke, Markus Nagler, Barbara Petrongolo, David Price, Arthur Seibold, Marco Tabellini, Catherine Thomas, Joonas Thukuri, Anna Valero and John van Reenen as well as seminar participants at LSE, MIT, Goethe University, ZEW Mannheim, RES Junior Symposium in Bristol, IZA summer school at Ammersee, EMCON in Chicago, Ski and Labor in Landeck, EEA Annual Congress in Cologne and EDP Jamboree at EUI. §London School of Economics and Political Science, Department of Economics, Houghton Street, London WC2A 2AE, U.K. E-Mail: f.koenig@lse.ac.uk; Web: www.felixkoenig.com.
1 Introduction

In recent decades top incomes have risen sharply, particularly in management, law, engineering and finance.\(^1\) Many economists link such growing top income trends to superstar effects, which arise when star workers can apply their talent on a bigger scale.\(^2\) As it gets easier to reach many consumers simultaneously, a greater share of consumers will flock to the most talented workers in the profession – the “superstars”. Such a shift in demand creates rising incomes at the top and simultaneously reduces incomes for less talented workers. The superstar theory has recently come to the forefront of public debate, due to the rise of digital and online technologies.\(^3\) A central concern is that new, scalable production technologies will lead to a surge in superstar effects. Despite the prominence of the theory, there is little causal evidence for the existence of such superstar effects.

This paper tests the superstar theory by exploiting a historic natural experiment in the entertainment sector which gives rise to exogenous variation in market reach. A dramatic shift in market reach occurred in the entertainment sector during the roll-out of TV in the US in the mid-20th century. Before the introduction of TV, a live performance could be watched by a few hundred individuals, while after the introduction of TV, the same performance could be watched by millions. Exogenous variation in the regional exposure to TV filming is used to test the main predictions of the superstar model in local entertainer labor markets. My results confirm the key predictions of the superstar model. I find that greater market reach leads to sharp growth of top incomes, while simultaneously eroding demand for mediocre workers. In labor markets where entertainers can reach a bigger audience, consumers shifts towards the stars of the profession. This shift in demand has pronounced consequences for labor market returns and moves entertainer labor markets closer to winner-takes-all markets. The pattern of wage changes closely aligns with the predictions of the superstar model but is at odds with leading alternative models.

The variation from the roll-out of TV enables me to address three key empirical challenges for a test of the superstar model. The first challenge is to distinguish superstar effects from other time trends that affect top income concentration, such as trends in deregulation and pay setting norms. The roll-out of TV stations proceeded in a staggered fashion and affected different parts of the US at different times. This within country variation allows me to test the effect of expanding market reach while holding aggregate time trends constant. To estimate the effect of expanding market reach, I use a difference in differences estimator that compares local entertainer labor markets with differential exposure to local TV filming. The comparison uses the fact that pioneering TV stations relied on local talent for TV show filming.\(^4\) Technical restrictions constrained the same show from being broadcast simultaneously from multiple stations and, as a result, TV was characterized by multiple local TV stations that independently broadcast content to the local population. TV networks,

\(^1\)For aggregate data see Alvaredo et al. (2018); for occupation specific US data see Kaplan and Rauh (2013); Bakija et al. (2012).

\(^2\)The superstar model is developed in Rosen (1981); Sattinger (1975); Tinbergen (1956). Applications of the model include Gabaix et al. (2015); Terviö (2008); Gabaix and Landier (2008); Garicano and Hubbard (2007); Cook and Frank (1995).

\(^3\)For a discussion of superstar effects in the digital economy see Bas et al. (2018); Guellec and Paunov (2017); OECD (2016); Acemoglu et al. (2014).

\(^4\)TV shows were effectively a non-tradable service. Recording was, in principle, possible in the form of “kinescopes.” However, the image quality of this technology was poor and such TV displays were unpopular. Shows produced elsewhere were a poor substitute for local productions.
which later harmonized programming across the US, initially had a limited influence over local programming. For a local entertainer, the construction of a TV station was therefore similar to the construction of a hypothetical giant theater that would hold an entire local population.

A second challenge for a test of the superstar model are changes in market reach that arise endogenously, for instance from technical change that responds to local wages. In the case of TV, the local technical change occurred through a government licensing process. Two features of the process generate variation in market reach independently of endogenous factors that usually drive technology adoption, such as the intensity of local competition or local factor costs. The first such feature are rules that determine how locations are prioritized for TV deployment. A second feature is a halt in licensing in 1948, which became necessary because of signal interference between stations. This interference was caused by an error in the regulator’s assignment model and interrupted the TV roll-out for nearly four years. I identify stations that were about to launch but narrowly missed out due to the license freeze. I exploit both of these regulatory constraints to generate local variation in access to TV.

The third challenge is that most technical innovations are multi-purpose and affect several occupations and industries simultaneously. Such multi-dimensional changes make it difficult to study the effect of market reach in isolation. Digital technologies, for instance, improve production scalability but also change the flow of information, the observability of worker performance and other aspects of production that may affect labor market outcomes. TV lends itself to study the effect of greater production scalability because the technology had limited additional applicability. Moreover, TV applied to a specific group of entertainer occupations, while production in the rest of the economy remained broadly unchanged. I built a novel dataset from archival records that allows me to track changes in production and consumption of entertainment. This data cover local entertainer labor markets across the mainland US throughout the 1940s and 1950s. On the production side, the data show where, when and why TV filming became feasible. Specifically, the data include information on the universe of broadcasting licenses of TV stations, their locations and audience sizes, as well as the historical capacity of over 3,000 performance venues. I combine this information with administrative records on the TV station licensing process, including information on how locations were prioritized. On the demand side, the data quantify the shift in labor demand. I digitize archival sources that report spending at roughly 4,000 local entertainment venues and contain information on prices and show revenues. With this data, I can trace demand shifts and associated changes in entertainers’ marginal revenue product. These records are linked to US Census micro-data that capture labor market outcomes in entertainment and beyond.

In line with the superstar model’s headline prediction, I find that top incomes grow sharply in response to growing market reach. A local TV station boosts pay at the 99th percentile by about 19% and expands the audience by roughly 300%. To put this wage growth into context, I look at the position of entertainers in the US wage distribution. Star entertainers rise markedly in the US wage distribution with the launch of a TV station. The share of local entertainers in the top 1% of the US wage distribution almost doubles. Locations that narrowly miss out on the launch of a TV station due to the regulatory shutdown see no growth in top entertainer pay. Similarly, I find no effect on other professions.

Evidence for such endogenous technical change is presented in Blundell et al. (1999), the theory in Acemoglu (1998).
The superstar effect is specific to the time periods, places and professions involved in local TV filming, reinforcing confidence that the effect is caused by TV.

Data on entertainment consumption allows me to test for labor demand shifts directly. The results show that local TV filming increases the audience for the biggest local shows more than fourfold while drastically reducing attendance at ordinary local live entertainment and, consequently, the revenues of local live performances.

To distinguish superstar effects from other mechanisms, I derive additional predictions that are specific to the superstar model. In cross-sectional data, superstar effects and the effect of canonical labor demand shifts are indistinguishable. However, the pattern of wage changes over time differentiates the superstar model from a wide range of alternative mechanisms. Specifically, in a superstar model technical change that allows for greater market reach moves labor markets closer to a winner-takes-all market. These trends are captured by four testable predictions: (i) disproportionate wage growth at the top, (ii) decreasing wages for mediocre workers, (iii) falling employment and (iv) growing wage dispersion at the top. The empirical results confirm these patterns. The impact of technical change on the wage distribution has a striking U-shaped pattern characteristic of the superstar effect. The gains at the top occur together with a decline in mid-income jobs and a growing low-pay sector. Moreover, I find that pay differences among top earners increase, as the wage gap between the 90th and 99th percentile widens significantly. I further document substantial employment losses. When TV signal becomes available in a local area, entertainment employment declines around 13%. These results show that demand is becoming concentrated on star workers, at the expense of mediocre workers.

Next, I quantify how much inequality superstar effects generate by estimating the elasticity of top pay to changes in market size. Data on audiences and prices allow me measure market size in terms of the number of customers and revenues. I use this data in an instrumental variable (IV) strategy, where the launch of a TV station is the instrument for market size. This IV estimator shows that doubling audience size increases wages at the 99th percentile of entertainer pay distribution by 17%. The superstar effect can explain about 70% of differences in top incomes across local entertainer labor markets. An equivalent IV estimate that quantifies demand concentration in terms of revenues, finds a similar magnitude of superstar effects. As revenues become concentrated on top shows, 22 cents of each dollar go to top earners.

A potential concern with the empirical strategy is that the launch of a TV stations is related to local trends that affect top pay in entertainment. I leverage the decline of local TV filming for a powerful parallel-trends test. The invention of videotape in 1956 made transporting and replicating shows attractive and led to modern production, in which shows are centrally produced and broadcast across the country. This resulted in the demise of local TV filming, and regional differences in the availability of production technologies therefore disappeared. As a consequence, local outcome differences ought to revert to their pre-treatment levels. This test goes beyond standard pre-trend checks, leveraging both pre- and post-treatment periods to verify common trends. The data confirms that the regulated TV roll-out is orthogonal to local trends.

The analysis next addresses how the results translate to other settings. First, I analyze how the effect of local demand shocks relates to the effect of aggregate shocks. Such effects

---

6The middle of the income distribution also hollows out in models of routinization where technology replaces mid-skilled workers (e.g. Goos et al., 2010). This differs from superstar models, where mid-skilled workers are replaced by star workers and technology acts as a vehicle for stars to project their talent.
differ if there are spillovers between local labor markets. Since live entertainment shows are by nature consumed locally, one of the main channels for spillovers is shut-down in this setting. There is no cross-labor market trade in output. The main potential link between local labor markets is entertainer mobility, which is observable, and the response of such mobility to TV launches can therefore be quantified. In practice, the TV roll-out creates limited incentives to move because the position of a local labor market in the ranking of market reach remains largely unchanged. In line with the stable ranking of locations, empirical tests show that entertainer mobility does not significantly change with the launch of a TV station. In a second extension, I explore how superstar effects change with imperfect competition in the labor market. Imperfect competition changes the predictions of the superstar model substantially, as gains from technical progress are no longer passed on to workers. The regulated entry of TV stations allows me to test empirically how competition affects the magnitude of superstar effects. Contrary to popular belief, it is not the lack of competition that raises top incomes, but rather more intense competition for talent.

The literature on superstar effects dates back over six decades. The model first appeared in the 1950s and was popularized in the following decades (Tinbergen, 1956; Sattinger, 1975; Rosen, 1981). Despite this long tradition, there is no common modeling framework to study superstar effects. In an attempt to structure the literature, I develop a unifying framework that nests many of the existing superstar models (including Costinot and Vogel, 2010; Terviö, 2008; Gabaix and Landier, 2008; Teulings, 1995; Rosen, 1981; Sattinger, 1979, 1975). The common core of the superstar model is a matching model where heterogenous workers are matched to differentiated tasks or markets. I use a benchmark version of the model to show how improving production scalability affects the demand for talent and, ultimately, wages. Previous empirical applications use the superstar model to explain the distribution of CEO pay (Edmans and Gabaix, 2016; Gabaix and Landier, 2008; Terviö, 2008). Such studies calibrate key model parameters to the correlation of pay at the top and market size. My study instead focuses on distinguishing the superstar model from leading alternative models. In a link to the previous literature, I additionally provide a comparison of OLS and IV estimates for the key model parameters.

Technical progress leads to changes in labor markets beyond superstar effects. Canonical models of technical change include models of efficiency units (Stigler, 1961), skill biased technical change (Acemoglu and Autor, 2011; Katz and Murphy, 1992) and routine bias technical change (Acemoglu and Restrepo, 2016; Autor and Dorn, 2013; Goos et al., 2010). I show how the superstar model differs from such models and derive testable predictions that allow me to distinguish the models in the data. The empirical tests of this study use variation in technology in line with previous work (e.g. Acemoglu and Restrepo, 2017; Michaels and Graetz, 2018; Akerman et al., 2013), but in contrast to those studies, I analyze a technical change that expands market reach and test for superstar effects.
Recent work has applied the superstar model beyond the labor market and showed that superstar effects can account for growing market concentration in product markets. When applied to firms, the superstar model rationalizes increasing dispersion in firm size and changing factor shares (Eeckhout and Kircher, 2018; Autor et al., 2017). There is growing concern that internet-based technologies lead to sharp increases in market concentration, which some observers link to rising mark-ups and rents (for evidence on rising mark-ups see De Loecker and Eeckhout, 2017). I show that market concentration and mark-ups need not go hand in hand. In the superstar model integrated markets reallocate resources to more talented workers and market concentration can arise in a fully competitive setting.

A concern for policymakers is that in the presence of superstar effects, gains from technical progress are distributed unevenly across the population (e.g. OECD, 2016). A range of policies have been suggested to address this. I analyze whether three popular interventions – changes in tax rates, workforce training and increased labor market competitiveness – reduce the inequality effects of growing market size (discussion of policies include Piketty et al., 2014; Scheuer and Werning, 2017; OECD, 2016; Bok, 1993). My test exploits regional variation in these policies and suggests that none of the interventions substantially reduces the rise of inequality.

The remainder of this paper is organized as follows. In Section 2 I derive the key predictions of a superstar model and contrast them with alternative models. Section 3 describes the data and archival sources. Section 4 reports results of the empirical tests of the superstar model. Section 5 estimates the magnitude of superstar effects. Section 6 discusses how imperfect competition and policies interact with superstar effects. Finally, Section 7 concludes.

2 The Superstar Model

The term “superstar effect” has been used to describe many different concepts, the aim of this section is to clarify the meaning and derive a definition from the superstar model. Part of the difficulty of pinning down the superstar effect is that different types of superstar models are being used and their connection is obscure. I derive a unifying superstar framework that nests the common models and illustrates the key properties that characterize a superstar market. I then illustrate the predictions of the superstar model using a simple benchmark version of the model. Next, I will show that many of the superstar models’ predictions can be replicated by conventional models of labor demand. Finally, I illustrate how technical progress generates predictions that differentiate the superstar model from a wide class of alternative models.

2.1 A Benchmark Superstar Model

A superstar model is an assignment model where heterogenous workers are matched with heterogenous tasks. In the context of entertainment we can think of workers as actors and of tasks as shows. Actors have different and unique talent \( t \) and the talent types can be ranked, actors are thus vertically differentiated. Shows differ in their innate productivity characteristics, think of these characteristics as the performance venue’s audience capacity, or size denoted by \( s \).

\[ \text{In the literature, differences in job characteristics are often referred to as “market size” or “firm value”. Important to the model is that the characteristics are innate and cannot be changed at the time of hiring.} \]
A general superstar model that nests the standard versions of superstar models in a unifying framework is presented in the Appendix C.1. Here I will illustrate the key mechanics of the model by developing a benchmark version, building on Sattinger (1979), that allows for closed form solutions.

**Labor Supply and Demand**

Since workers are differentiated, we need to characterize the labor supply of each worker type. Assume that each worker supplies one unit of labor inelastically, the labor supply then is the same as the distribution of worker types. In the same way labor demand is characterized by the distribution of venue sizes. The benchmark model makes the simplifying assumption that both actor abilities and show sizes follow a Pareto distribution (More general results are illustrated in Appendix C.2). Denote the probability that an actor's talent is above a threshold $t$ by $p^t$ and the equivalent for shows by $p^s$. I denote by $x_p$ the value of a variable $x$ at percentile $p$. The inverse CDFs of the two Pareto distributions are then given by:

$$p^t = t^{-\frac{1}{\alpha}}$$  \hfill (1)

$$p^s = s^{-\frac{1}{\beta}}$$  \hfill (2)

The distribution of talent and venue size is characterized by the shape parameter of the respective Pareto distribution. Here the shape parameter is the inverse of the exponents, respectively $\alpha$ and $\beta$ and a bigger value implies greater dispersion. Next, assume that workers and shows are matched one-to-one, each show hires exactly one actor and an actor performs in one show.\(^1\) One-to-one matching has been widely adopted in the superstar literature to keep the model simple (e.g. Gabaix and Landier (2008); Terviö (2008)), but extensions to one-to-many matching lead to similar results (as in Garicano (2000); Rosen (1981); Sattinger (1975)).\(^2\)

**Production**

A matched actor-show pair produces revenue $F(s, t)$. The key assumption of a superstar model is that more talented workers have a comparative advantage in larger markets, which in the entertainment context implies that adding an extra seat to a theater affects revenues more when a better actor is performing. In other words, the superstar model assumes that $F(s, t)$ is super-modular.\(^3\) A Cobb-Douglas production function guarantees this and allows for a simple closed form solution. I therefore assume that production is given by:

Further, note that these characteristics are not the same as the employer’s market value, which depends on both innate characteristics and endogenous factors such as talent employed and the price of talent. In the empirical section I will address how to distinguish a change in $S_i$ from the endogenous firm value $Y$.

\(^1\)One-to-one matching implies imperfect substitutability of talent. Since each show is matched to only one worker of quality $t$, this worker cannot be replaced by two workers with quality $\frac{t}{2}$, or with two workers of any type.

\(^2\)The one-to-many matching features equilibrium cut-offs that determine which share of jobs is performed by which type of workers. Highly talented actors’ comparative advantage in juggling many shows implies that they serve a greater share of the shows.

\(^3\)In some theoretical work a related assumption is used and $F(s, t)$ is assumed to be log-supermodular. This assumption is neither implied by nor does it imply super-modularity.
\[ F(s, t) = \pi s^\gamma t^\delta \]

where \( \pi \) is the price of a unit of output. This production function exhibits comparative advantage because \( \frac{\partial F(s, t)}{\partial s t} > 0 \).

**Equilibrium**

The equilibrium of this setting consists of an assignment function of actors to shows \( (s = \sigma(t)) \) and a wage schedule that ensures the assignment is incentive compatible. Moreover, markets clear at \( \pi: \int_F(s, s(t))dtds = D(\pi), \) where \( D(\pi) \) is the demand for entertainment. I will state the equilibrium conditions and leave the proof for the appendix. The first equilibrium condition is positive assortative matching (PAM): the best actor performs in the biggest show, the second in the second biggest and so forth. The second equilibrium condition is that the wage schedule guarantees incentive compatibility, no actor or show manager wants to be matched with a different type. The two equilibrium conditions are given by

\[ p^i = p^s(\sigma(\hat{t})) \iff \sigma(\hat{t}) = \hat{t}^{\beta} \]

\[ w'(\hat{t}) = F_t(\sigma(\hat{t}), \hat{t}) = \delta \pi t^{(1/\xi - 1)} \]

Equation 3 is a formal expression of PAM, it states that percentiles in the size and talent distributions are the same. We can use this equilibrium condition together with the inverse CDF functions 1 and 2 to solve for the matching function \( \sigma(t) \). The second equilibrium condition states that the wage increase for a marginally more talented worker equals the marginal product of the worker in the equilibrium assignment. The second equality uses the equilibrium assignment from equation 3 to eliminate \( s \) and write wages in as a function of equilibrium talent \( \hat{t} \). The exponent is defined as \( \xi \equiv \frac{\alpha}{\delta \alpha + \gamma \beta} \). The resulting equilibrium condition implies that wages are competitive despite the fact that each actor and show is a monopolist of its type. The competitive result holds because we assume types are distributed continuously. The actor’s outside option is thus only infinitesimally worse and the actor receives the full productivity gains of moving to a bigger show. Take the alternative case, where the distribution of show types has jumps; some theater venues are thus discretely bigger than their competition. Here the show producer would keep all the productivity gains. A lack of competition among employers therefore dampens wages.

To solve for wages, integrate equation 4. This pins down wages up to a constant and for simplicity I set that constant to zero. Wages are then given by:

\[ w(t) = \xi \delta \pi t^{1/\xi} \]

To solve for the wage distribution, eliminate \( t \) from equation 5 by using equation 1. Assortative matching and \( F_t > 0 \) ensure that the percentile of the wage distribution corresponds to the percentile of the talent distribution in equilibrium \( (p^w = p^s) \). We therefore arrive at the superstar wage distribution with \( \lambda = (\xi \delta \pi)^{\xi/\alpha} \).

\[ p^w = \lambda w_p^{\frac{\xi}{\alpha}} \]

Wages follow a Pareto distribution, with the shape parameter \( \frac{\alpha}{\xi} \). Recall that the shape parameter of the talent distribution is \( \alpha \). Comparing the two shape parameters, reveals that
wages are more dispersed than talent if $\xi < 1$. For small values of $\xi$ the superstar model therefore produces large wage differences, even if talent differences are small. I call this result the “talent amplifier effect”, which has been the focus of much early literature (discussions include Rosen (1981); Tinbergen (1956); Sattinger (1975)). The talent amplifier effect is a consequence of PAM. To see this, take two workers who have similar levels of talent and thus similar levels of productivity if they perform in the same venue. In equilibrium PAM implies that the more talented worker is assigned to a larger and more lucrative venue, which increases the productivity differences between the two workers. Wages are competitive and reflect these productivity differences and are therefore more unequal than the pure talent difference would suggest. This talent amplifier effect holds when $\xi < 1$, which occurs when large show venues are scarce enough to overcome potential opposing effects from decreasing returns to scale (aka if $\frac{\beta}{\alpha} > \frac{1-\delta}{\gamma}$). In what follows I assume that this restriction holds.\(^{14}\)

A test of the talent amplifier effect has proven difficult. Such a test requires knowledge of the talent distribution to distinguish the talent amplifier effect from an alternative model where a skewed income distribution is the result of a highly skewed distribution of talent. The lack of a cardinal metric for talent has made it difficult to test this implication of the superstar model.

Time series changes in the wage distribution generated by the superstar model are more distinct. In the model wage changes are driven by changes in market size. To see this, use the fact that $p^w = p^s = p$ and substitute equation 2 into 6 and take logs. Wages at percentile $p$ can than be expressed as:

$$\ln(w_p) = \ln(\xi \delta \pi) + \frac{\alpha}{\beta \xi} \ln(s_p)$$

Wages are a function of market size and wage growth is thus proportional to changes in market size. I call this relation the “superstar effect.” The related literature has used this result to generate two insights. First, dispersion in firm size does not grow quickly enough in a random growth model to generate transition dynamics that account for the sharp rise in income concentration in the US (Gabaix et al. (2015)). However, with a more nuanced growth process, the superstar model matches the data. Second, CEO pay can be explained by this relation when firm values are used as an empirical analogue to the size distribution (Gabaix and Landier (2008); Terviö (2008)). Although these results illustrate the model’s potential power, they do not preclude the possibility that other factors cause the relationship between wages and market size. A similar relation arises from alternative models; most notably, models of endogenous technical change link labor productivity and firm productivity (see Blundell et al. (1999) for an empirical illustration).

### 2.2 The Effect of Technical Change

The remainder of this section illustrates a pattern of changes in $w$ that allows to distinguish the superstar channel from other potential channels. In the empirical application the exogenous instrument will rule out spurious findings from range of mechanisms, however, such exogenous variation in technology does not rule out that technical change affects wages through other channels than superstar effects. To distinguish different models of technical

\(^{14}\)This additional restriction is not required in other versions of the superstar model. For example, Sattinger (1975) assumes log super-modularity in production and does not require additional assumptions on the spacing of the distributions.
change, I derive patterns of wage changes from the superstar model that are distinct from conventional models of technical change.

2.2.1 Superstar Effects and Technical Change

The superstar effect is the result of expanding markets reach. A tractable way of modeling such a change is allowing production to become more scalable and thus reducing the diseconomies to scale in the production function. Assume that this change takes the form of $\delta' = s \cdot \delta$ and $\gamma' = s \cdot \gamma$ with $s > 1$. The new wage distribution (call the new period $t + 1$) is therefore found by substitute the new values $\xi'$ and $\lambda'$ into equation 6:

$$p_{t+1}^w = \lambda' w_p \frac{\xi'}{\alpha}$$

(8)

Since we assumed that labor supply is inelastic we can solve for wage growth by dividing the new and old wage distributions evaluated at percentile $p$. Wages in $t + 1$ are given by 8 and period $t$ wages are given by 6. Wage growth at percentile $p$ is therefore:

$$g_{w_p} = \frac{w_{t+1}^p}{w_p^t} = \psi_p \frac{\alpha}{\xi} (s-1)$$

(9)

Where $\psi = \left( \frac{\lambda'}{\lambda} \right)^{\alpha(s-1)}$. These equations reveal, that the reduction of diseconomies to scale has differential effects at different parts of the distribution. The effect are summarized in Figure 1, the wage distribution shifts inward and pivots out. The intuition is that more productive workers are matched with bigger shows and therefore operate on a bigger scale, diseconomies to scale are more binding for this group. Such top workers therefore benefit most from better scalability of production. The effect can be seen in two changes in equation 8, the shape and scale parameter of the Pareto wage distribution change. Compared to equation 6 $\xi' = \xi < \xi$, which implies that wage differences between workers grow, the wage distribution pivots out and the distribution becomes more right skewed. Besides this top income growth, there is an additional level effect on the wage distribution operating through $\lambda'$. This is a level effect that reduces wages at all levels. The level effect is a consequence of expansion in the availability of entertainment, since more entertainment is being produced, the entertainment market clears at a lower price for talent $\pi$. As a result the Pareto scale parameter falls ($\lambda' < \lambda$) and the wage distribution shifts inward (see equation 8). This case illustrates one of the key features of a superstar model, the potential for cannibalization effects. The greater availability of stars, reduces demand for the rest of the profession and in the limit, a single superstar serves the entire market. In summary, the bottom of the distribution benefits little from better scalability but suffer from the fall in the price for talent units, while at the top of the distribution the bigger scalability over-compensates for the fall in $\pi$. Previously unattained income levels are reached at the top and bottom ends of the distribution, while mid-paid jobs simultaneously disappear.

---

15 An alternative but ultimately equivalent way of modeling this change is to allow the size distribution to change, for example by increasing the shape parameter $\alpha$ (see Appendix C.3).

16 If we maintain that the outside option is fixed at a level $b$, the lowest wages are fixed at $b$ and adjustment occurs through exit rather than falling wages. Wages at the bottom could decline if there is a cost to exiting, for example search costs, or if payoffs from the outside option also fall.

17 Notice that if $\pi$ is unchanged (ie if demand for entertainment is perfectly elastic), $\lambda$ would rise. I assume that demand is sufficiently inelastic to rule this case out.
For empirical tests, it will be useful to derive separate predictions for different parts of the distribution. I will illustrate the effect of technical change by deriving which types of jobs are created and which ones are being destroyed. Consider the number of jobs that pay wage \( w \), given by the density of the wage distribution \( f(w) \). To derive the density take the derivative of \( f(w) \) with respect to \( w \) and multiply by minus one:

\[
f(w) = \frac{\lambda \xi}{\alpha} w^{-\frac{\xi}{\alpha} - 1}
\]

The two effects of technical change are visible again here. Since \( \xi' = \frac{\xi}{\alpha} < \xi \) and \( \lambda' < \lambda \) the Pareto scale parameters falls, while the shape parameter \( \frac{\alpha}{\xi} \) increases. This again leads to a level decrease but an outward pivot of the distribution. The implications for the growth of high and low paid jobs can be computed by dividing the mass of jobs with wage \( w \) in period \( t + 1 \) with its mass in period \( t \). The growth in the share of actors with wage \( w \), denoted by \( g_e(w) \), is given by:

\[
g_e(w) = f_{t+1}(w) \frac{f(t)}{f(t+1)} = \frac{\lambda' \xi' w^{(s-1) \frac{\xi}{\alpha}} - 1}{\lambda \xi} \quad (10)
\]

This growth rate is illustrated for different wage bins in Panel B of Figure 1. While the magnitude of the changes depends on distributional assumptions, the pattern is independent of these assumption. Jobs that pay at the extremes of the distribution are becoming more common, while mid-income jobs are disappearing. The effect of technical change is therefore U-shaped across the wage distribution. To see this note that \( g_e(w) \) is increasing in \( w \) and will be positive for large \( w \). The fraction of top paid actors is therefore growing, with effects becoming more pronounced at higher \( w \). By contrast, for lower values of \( w \) the growth rate turns negative since \( \frac{\lambda' \xi'}{\lambda \xi} < 1 \). Also note that the two distributions do not have the same support. Incomes that were previously outside the range of the income distribution appear in both tails of the distribution through technical change. The growth rate of such previously non-existing job types is undefined, as we would divide by zero. However, the share of jobs increases unambiguously. Panel B of Figure 1 groups the wage tails into a final wage bin and report the growth rate for a bin that has support in both distributions. By using wage bins I can compute growth rates for wage bins that span the full wage distribution.

### 2.2.2 Alternative Models of Technical Change

Next, I compare the effect of technical change in a superstar model to it’s effect in conventional models. A key difference between superstar models and standard labor demand models is worker substitutability. In a superstar model all worker types are unique and imperfectly substitutable, while in standard labor demand models some worker groups are perfectly substitutable, and this difference has testable implications for the effect of technical change. A classic case is the canonical model of “Skill Biased Technical Change” (SBTC). This model features low- and high-skill groups and workers within each skill group are perfectly substitutable. This simple model is silent on top income dispersion, but can be extended to feature a continuum of worker types. To contrast this with superstar models, I maintain perfect substitutability within skill groups but allow two workers in the same skill group to have different skill quantities. The literature refers to such differences in skill quantity as “efficiency units”. Assume workers at percentile \( p \) have an amount of skill \( q_p \) with \( q_p \sim Q(p) \). Since the skill units are perfect substitutes, the model features a single market clearing
price for skill \( \pi \) (this type of model is developed in Stigler (1961)).\(^{18}\) Workers are paid in proportion to their skill \( w = pq \). With the right distribution of efficiency units, the SBTC model fits any wage distribution, therefore in the cross-section, the SBTC model is indistinguishable from the superstar model.

To examine the differences between the heterogenous workers who are perfect versus imperfect substitutes, I focus on a single group. Specifically, I will abstract away from the low-skill group and focus on income dispersion among the high-skilled.\(^{19}\) First consider the baseline case, where labor supply is perfectly inelastic and all workers with skill above \( \bar{p} \) are participating in the market.\(^{20}\) A skill-biased demand shift (SBD) increases the demand for talent \( D(\pi) \) to \( D'(\pi) > D(\pi) \).\(^{21}\) Market clearing implies that increase in demand for talent increases the price of talent \( \pi \) to \( \pi' \). A unit of talent becomes more valuable, and the more talent a worker has, the more she benefits from the growth in \( \pi \). After a SBD shock in period \( t + 1 \) wages at percentile \( p \) are given by:

\[
 w_{t+1}^p = \pi' \cdot q_p = w_t^p \frac{\pi'}{\pi}
\]

The effect of a skill-biased demand shift is proportional to the previous wage level. The wage growth at percentile \( p \) is given by:

\[
 g^w_p = \frac{\pi'}{\pi} \cdot q_p = g^w
\]

Notice that \( g^w \) does not carry a subscript for percentiles. All wage increases are proportional to talent and the growth rate is therefore constant across all percentiles. The intuition for this result is that workers are perfect substitutes. If a worker can be replaced by two workers with half the talent, wages are thus always proportional to the difference in talent. Wage growth is equal to growth in the skill premium \( (g^\pi = \frac{\pi'}{\pi} > 0) \), independent of \( p \). To compare the results to the superstar effect, assume as above that talent is Pareto distributed \( (p = q - \frac{1}{\alpha}) \).\(^{22}\) This allows us to solve for the wage distribution:

\[
 p_t = \left(\frac{w}{\pi}\right)^{-\frac{1}{\alpha}}
\]

The growth in the skill premium to \( \pi' \) leads to an outward shift in the wage distribution that is illustrated in logs in Figure 2. First, notice that the original wage distribution is identical to the result of the superstar model. With the right assumption on its parameters, the SBTC

\(^{18}\)Note that we can make the SBTC coincide with a model of unique talent. This would require that the number of skill groups goes to infinity; eventually each worker is her own skill group and thus is imperfectly substitutable. In that case, differences between the SBTC and superstar models are a result of the type of technical change. In the superstar model, star workers displace other workers, while in SBTC models workers of different types are q-complements.

\(^{19}\)The results for the fully fledged model are equivalent and presented in Appendix C.5.

\(^{20}\)This assumption is immaterial here but becomes relevant if one introduces matching (as in Eeckhout and Kircher (2018)).

\(^{21}\)In the conventional model, the demand shift is a result of a skill augmenting change in productivity (see Appendix C.5). Here we take the reduced form approach of modeling the skill-biased demand shift as a change in the demand for talent.

\(^{22}\)To cut through the debate on assumptions related to the talent distribution, I show which talent distribution is required for this model to match the 1939 wage distribution and what shift in the skill premium is needed to account for the growth in top earners between 1939 and 1969. The predicted wage change pattern for the rest of the distribution is shown in the Appendix Figure D2.
and superstar models yield the same result and makes the two models indistinguishable in cross-sectional data.

2.2.3 Testable Differences

Technical change, however, leads to a distinctive change, visible in Figures 2 and 1. The SBTC and superstar models have strikingly different effects: the former leads to an intercept shift, while the latter shifts and pivots the wage distribution. Cannibalization effects and fractal inequality distinguish the two models from one another. Fractal inequality refers to pay growth at the top that becomes more pronounced as one moves up the top tail of the pay distribution. Cannibalization indicates that top income growth is accompanied by negative effects for mediocre workers. This is visible at the middle and bottom parts of the wage distribution, where mid-income jobs disappear and low-pay jobs emerge. These effects are summarized by four testable propositions:

**Proposition 2.1.** Top pay growth: For two percentiles at the top of the wage distribution \( p' > p \) a superstar effect predicts that wage growth \( g^w \) meets: \( g^w_{p'} > g^w_p \), while a SBD shock has \( g^w_{p'} = g^w_p \).

A superstar model generates disproportionate gains at the top, while wages grow proportionally to the level of skill in a model of SBD shocks. The SBD model does not generate skewed income growth because of the law of one price. The shift in the price for talent will affect all talent units equally and therefore lead to wage growth that is proportional to a worker’s talent. As a result, the wage growth rates are the same across the distribution.\(^{23}\) The result follows immediately from equations 9 and 11.

**Proposition 2.2.** Mediocre worker pay: In a superstar model \( w_{p}^{t+1} < w_{p}^{t} \) is feasible, while in a SBD model \( w_{p}^{t+1} > w_{p}^{t} \) at all percentiles.

Mediocre workers lose out due to superstar effects, while wage growth is always positive in the SBD model. A SBD shock is a positive demand shift that increases wages across the board. The first part of the proposition follows straight from equation 11. For the second part, consider equation 8 and solve for the wage: \( w_{p}^{t+1} = \left( \frac{\lambda'}{\pi} \right)^{\frac{\alpha}{\xi}} \). Technical progress leads to negative wage growth if \( \lambda \) declines fast enough. To see this let \( p \to 1 \), wages at the bottom of the distribution converge to \( w_{p}^{t+1} \to (\lambda')^{\frac{\alpha}{\xi}} \). Falling wages occur if \( (\lambda')^{\frac{\alpha}{\xi}} < (\lambda)^{\frac{\alpha}{\xi}} \), or if demand is sufficiently elastic. The intuition is that in the superstar model, technical progress allows stars to steal some of the business of lesser stars, while in the SBD model q-complementarity of worker types guarantees that wages grow if any type becomes more productive.

**Proposition 2.3.** Employment: In a superstar model \( \bar{p}_{t+1} > \bar{p}_{t} \), while in a SBD model \( \bar{p}_{t+1} < \bar{p}_{t} \).

With entry and exit, the participation threshold \( \bar{p} \) determines which worker are active in the market. Employment declines in a superstar model as stars’ growing reach pushes other workers out of the market. With positive demand shocks, by contrast, quantity and wages move in the same direction. The price for talent \( \pi \) rises in response to SBD shocks but falls.

\(^{23}\)With additional skill groups this holds approximately for the highly talented individuals within a skill group.
with superstar effects. The proposition then follows from the market clearing condition in equation 25, meaning that a higher price for talent leads to more participation.

**Proposition 2.4.** Dispersion at the top: Income differences within the top tail increase with superstar effects but not with SBD shocks. This implies:

i) For a percentile ratio \( R_{p', p} \) of percentiles \( p' > p \): \( R_{p'}^{t+1} > R_p^t \) in a superstar model and \( R_{p'}^{t+1} = R_p^t \) in a SBD model,

ii) For relative top income shares \( s_{p} \): \( s_{1\%}^{t+1} / s_{10\%}^{t+1} > s_{1\%}^t / s_{10\%}^t \) in a superstar model and \( s_{1\%}^{t+1} / s_{10\%}^{t+1} = s_{1\%}^t / s_{10\%}^t \) in a SBD model.

This proposition highlights distinctive results on the income dispersion within the top tail. A superstar model exhibits a fractal inequality, so that moving up a rank in the talent distribution becomes more valuable. The same does not hold with the SBTC model, where the relative pay differences remain stable. The proposition is derived in Appendix C.6.

A natural question is whether extensions to the SBTC model allow it to replicate these results. The key distinction highlighted so far is that a SBTC model features groups of perfectly substitutable workers, whereas workers are imperfect substitutes in the superstar model. However, there are additional differences between superstar and SBTC models. To see this, consider the case where there is a continuum of skill groups in the SBTC model. Workers in two different skill groups are imperfect substitutes, with a continuum of skill groups each worker is his own skill group and hence all workers are imperfect substitutes. This extended SBTC model can generate fractal wage inequality, it requires technical change that increases productivity in an escalating fashion towards the top. However, the model will not feature cannibalization effects. A positive demand shock translates into gains across the range of the distribution, which is proved in Appendix C.7. Even a SBTC model where all workers are imperfectly substitutable will therefore not feature cannibalization effects and will not replicate propositions 2.2 and 2.3.

In summary, the superstar effect leads to four testable predictions:

1. disproportional wage growth at the top,
2. decreasing wages for mediocre workers,
3. falling employment and
4. growing dispersion of wages at the top.

Effects one and four reflect the fractal inequality effect, while effects two and three capture the cannibalization effect.

### 3 Data

For the empirical analysis I collect novel data from archival sources. The data cover information on both the production and consumption of entertainment in the middle of the 20th century. Data on consumption covers local-level consumer spending and attendance at entertainment venues, while the production side includes inputs and technology that are locally available. These data are linked to entertainers’ labor market records.
3.1 Production Technology

TV Data For each labor market I compute two measures of TV: exposure to television filming and exposure to television broadcasting. The first captures the change in the production technology and records where TV shows are produced. The second measures where local entertainers face competition from television.

Television Filming Data on television facilities come from the “Annual Television Factbooks.” The Factbooks record the address, technical equipment, launch date, assigned channel and call letter for each TV station. I geocode the studio locations and match them to the local labor market. This allows me to track the roll-out of TV across local labor markets. The location of TV studios is one of the main sources of variation in the analysis. The year 1949 is of particular interest, as labor market data from the Census is available for that year. Figure 3 shows where broadcasting took place in 1949. The treatment variable for a local area sums up the number of stations active in the local labor market at the beginning of a year and therefore assumes that all stations were filming locally at that time. There are a handful of exceptions, as a few stations operated a local network. These interconnected stations could relay local shows to nearby stations through upgraded phone lines (run by AT&T) or microwave relay technology (run by Bell). Interconnection was rarely feasible because the technical infrastructure was still in its infancy. In my main specifications I code all members of such networks as treated. This approach avoids potential endogenous selection of filming locations within the network.

TV Licensing FCC data on the priority of locations capture which places are about to receive a TV license. During the license freeze period this data identifies locations that narrowly missed out on the launch of a TV station. The freeze in licenses began in 1948 and continued until 1952. The data are based on the weekly bulletins from the Federal Communication Commission (FCC), which are summarized annually in the “Television Factbook.” Additionally, I collect information on the rule used to prioritize locations. This rule was published in a few years and reveals that the priority ranking of the TV roll-out was based on fixed location characteristics. This lends credibility to the assumption that the TV timing did not respond to local demand shocks.

Videotape Filming Local filming was ultimately superseded by centralized productions. With the invention of the videotape in 1956 television production shifted away from local stations, towards places where conditions for filming are most favorable. Centrally produced shows could then be broadcast across the country at low cost and at the time appropriate for the local time zone. The decline of local TV filming serves as a check on the validity of the estimates, since I can test if local superstar effects disappear. To do so, I will control

---

24 Previous data collection efforts have used derived products of this data. Gentzkow (2006) uses information published by the Annual Television Factbook to build a dataset of signal coverage throughout the US.

25 Robustness checks explore alternative treatments. As expected, within those networks effects on top incomes appear in the labor market where filming is mostly located.

26 The invention of the videotape is another example of a technology that expanded market reach. However, this variation is less suitable for a test of the superstar model, the location of filming for videotape is not exogenously imposed. Instead markets with the lowest production cost were selected for filming. Production cost are determined by endogenous factors such as local wages and tax rates as well as fixed location characteristics (e.g. sunshine hours, availability of equipment and expertise, local scenery).
for places where filming centralizes, which raises the potential challenge of an endogenous control variable. To avoid this endogeneity issue, I proxy locations of centralized filming with locations where movie filming took place in the 1920s. This measure is pre-determined and picks up location incentives that come from permanent regional characteristics. The data on the location of film shoots come from the “Internet and Movie Database” (ImDB). ImDB is a widely used platform (self-proclaimed number one worldwide) for information on movies and holds metadata on over 4 million movies. In 1920, around 200 movies were produced in the US. For each labor market I compute the share of movies produced in this market in 1920.

3.2 Demand Data

**Television Broadcasting** I use information on TV signal to calculate TV shows’ audience size. These data on TV viewership allow me to quantify the shift in market reach experienced by local entertainers. Information on TV stations’ signal reach comes from Fenton and Koenig (2018), who re-construct historic catchment areas. For details on the data construction see Fenton and Koenig (2018). Based on the signal data, I compute a TV station’s potential audience size. TV airwaves travel beyond the local labor markets where shows are filmed. I combine the information on each TV station’s catchment area with Census data on household location and TV ownership. The match is performed by mapping TV signal information onto geographic units available in the Census. The geographic match uses the boundary shapefiles provided by NHGIS (Manson et al. (2017)). I then identify how many TV-owning households are in each TV station’s catchment area. This allows me to construct a measure of potential audience size. The median TV station could reach approximately 75,000 households. Even the smallest TV audiences substantially exceeded the show audiences of local venues. Finally, the data is complemented with station specific price data from “rate cards,” which are used to compute hourly revenue figures for each market.

The same variation is used to estimate how the availability of TV signal affects demand for local live entertainment. Places that are exposed to TV signal are not necessarily the same as labor markets that produce TV shows. Figure 4 shows the variation in TV signal in 1950 and illustrates areas that narrowly miss out on TV signal due to the freeze in licensing.

**Theatre Data** Records on the seating capacity of live performance venues allows me to measure audience size in the pre-TV period. This information comes from a historic companion book for the entertainment profession, the 1921 “Julius Cahn-Gus Hill theatrical guide.” The book aims to provide “complete coverage of performance venues in US cities, towns and villages.” I hand collect data on seating capacity and ticket prices of over 3,000

---

27 This uses an irregular terrain model (ITM) to calculate signal propagation. In this model, signal reach depends on the technical properties of an antenna (channel, frequency, height, etc.) and on terrain that blocks airwave travel (e.g. mountains). The ITM has also been used in a number of other studies (Olken (2009); Enikolopov et al. (2011); Durante et al. (2015))

28 For the US, alternative signal data has been collected in Gentzkow (2006). He uses modern media markets as proxy for historic TV signal. Modern media markets are linked to the first TV station that falls within that radius and the launch date of that station is used as the date where TV became available in the area of todays media market.

29 According to the author “Information has been sought from every source obtainable - even from the Mayors of each of the cities.” Undoubtedly the coverage will be better for larger venues and small or pop-up venues will be missed. Since we focus on star venues this omission may be a lesser concern.
performance venues that, taken together, cover more than 80\% of US local labor markets. Spot checking confirms that the data accurately cover physical performance establishments.\(^{30}\) On average, a performance venue has 872 seats, but capacity varies between a few hundred seats to several thousand. The most iconic performance venue at the time was the New York Hippodrome, which was hailed as the “world’s largest theater.” Its seating capacity of over 5,000 seats was substantially bigger than most similar venues. The largest venue in a labor market had, on average, 1,165 seats. Combining audience data from live venues and for local TV shows, I build a novel measure for the audience reach of local star entertainers. Moreover, market reach is computed in dollar terms using the newly collected information on ticket prices and advertisement rates. Where price data is unavailable, prices are imputed based on audience size. These measures captures the potential audience and revenue of a single performance show and will allow me to quantify the shock in market reach from the launch of television.

**County Fairs** To measure spending and attendance at local entertainment outlets I collect information on the ticket sales and revenues for over 4,000 fairs spanning 11 years (1946-1957) and the majority of US labor markets. The data come from the “Cavalcade of Fairs,” which contains detailed records on county fairs and is published annually as a supplement to Billboard magazine. Fairs provide a range of amusement activities, usually including a carnival with rides, food stalls, activities and a grandstand show with performances by local dance squads and music groups, sport competitions and similar highlights. The records split spending into three categories: fair ticket receipts, show entrance receipts (e.g. grandstand) and carnival receipts\(^{31}\) (e.g. fair rides, merchandise and food). I aggregate the spending categories at two levels: the county-year level and at the more aggregated local labor market-year level, which allows me to analyze demand across these regions separately for leisure activities that are differentially close substitutes for television.

### 3.3 Labor Market Data

**US Census** Data on the local labor markets for entertainers come from the US Census. The US Census collects micro-data on the full US population once every decade, for my sample I use the data from 1930-1970.\(^{32}\) The sample period covers the TV roll-out, as well as pre- and post-rollout periods. The full population of US residents is covered in 1930 and 1940 and a representative sample in later years.\(^{33}\) Data on wages, occupation and employment are available consistently for individuals over the age of 15. I therefore restrict my sample to that age group.

The core of the analysis focuses on occupations that appear on television. Three-digit occupation information identifies five relevant entertainment occupations: actors, athletes, dancers, musicians and entertainers not elsewhere classified. The last group is relevant because it includes most circus and vaudeville actors, one of the most important forms of entertainment.

\(^{30}\)The data does not cover mobile performance venues, such as circus tents. Since the analysis is concerned with the largest local performance venues, this omission is likely not a major problem.

\(^{31}\)Carnival receipts are unavailable in 1953 and 1955.

\(^{32}\)Data is provided by the Integrated Public Use Microdata Files (IPUMS, Ruggles et al. (2017)).

\(^{33}\)Extending the time period in either direction is precluded by changes in variable definitions. Prior to 1930, the Census used a significantly different definition of employed workers than in my period of interest, and from 1980 onwards, the Census uses different occupation groups.
entertainment at the time.\footnote{The original string occupation title is available in the 1940 Census and confirms that the category includes acrobats, clowns, animal trainers, etc. Unfortunately, the original job title is not available for all other years and thus prevents me from further dividing entertainers not elsewhere classified.}

In many settings, the reclassification of occupations over time poses a problem. Entertainment occupations, however, are well established and there is little change to their definitions throughout the sample period.\footnote{The 1950 occupation classification system was used in the 1940-1970 censi. For the years 1920 and 1930, IPUMS has mapped the original answers into the 1950 categories.} There are a few exceptions; most relevant for the above groups is that the athlete category is discontinued in 1970. To account for such time shifts in the occupation definition, the regressions will control for occupation-specific year effects.

In defining labor markets, I follow Autor and Dorn (2013) and define local labor markets as urban centers together with their respective commuter belts, so called “commuting zones.”\footnote{These areas are defined using strong within-labor-market commuting and weak across-labor-market commuting.} I extend the data produced by Autor and Dorn (2013) backwards and produce consistent labor markets for the Census data going back to 1930.\footnote{Autor and Dorn (2013) provides crosswalks for 1950 and 1970. I build additional crosswalks for the remaining years. For each Census I use historic maps for the smallest available location breakdown. I map the publicly available Census location identifiers into a commuting zone.} The final data covers 722 consistent local labor markets spanning the mainland US over time. On average, a commuting zone has about 400,000 inhabitants and approximately 500 workers in entertainment occupations.

Wage data are first collected in the 1940 Census and in all years refer to the previous year. From 1939 onwards, the data are available consistently throughout the period. In 1940 the full distribution of wages is reported, but from 1950 onwards top coding applies. Fortunately, the top code bites above the 99th percentile of the wage distribution and up to that threshold, detailed analysis of top incomes is possible.

The analysis studies labor demand shocks at the local labor market level and I compute outcomes at this level for each occupation and year. A first set of outcome variables follows Chetty et al. (2014) and measures wage inequality by ranking entertainers relative to a benchmark group. I compute entertainers’ position in the US wage distribution. One key advantage of the rank position is that it is scale independent and thus makes it easier to compare changes in pay inequality over time. A further benefit is that the measure avoids problems from top coding. Even with the top code in place, I observe the share of workers above a threshold, say the 99th percentile.\footnote{The rank approach works, as long as the final threshold is below the top code. This is the case for the 99th percentile of the US distribution.}

This outcome measure is computed as follows: For each local labor market, I divide the number of entertainers in a given income range by the total number of entertainers in the market. To prevent that changes in the denominator bias the results, I fix the denominator above the treatment level and for all labor market divide by the same number. The denominator therefore amounts to a normalization.\footnote{To simplify interpretation of the treatment effects, I normalize by the average number of entertainers in treated labor markets (instead of averages across all markets). The regression coefficient therefore has a natural interpretation as a percentage point change in the treated market.}

$$e_{m,t}^{1\%} = \frac{\sum_i E_{i,m,t} \cdot D_{US1\%}}{E_t}$$
The denominator is the number of top-earning entertainers in market \( m \) at time \( t \). Results without the normalization, as presented in the Appendix D2, are in line with the baseline effects. I compute additional outcome measures: top-paid entertainers as a share of the local population and top income shares of entertainers.\(^{40}\)

Finally, I identify later TV stars in the 1940 Census. This is possible because the 1940 Census includes names, date of birth and parental information. The “Radio Annual, Television Yearbook” publishes the “Who is Who of TV,” which allows me to identify TV stars in 1950. Manually searching vitas generates information on place of birth, birth date and parents. Combined with the information on names and places of residence, I identify entertainers’ 1940 Census records. For 60 out of 89 cases a unique Census record can be identified; clearly, this is only a subset of all entertainers. The advantage of these linked records is that we obtain information about entertainers’ pre-TV careers.

4 Empirical Tests

TV brought monumental change to the entertainment sector. Figure 5 shows that wages became substantially more polarized between 1940 and 1970. Before TV, most entertainers earned close to average pay but dispersion grew substantially over the following decades. By 1970, after the introduction of TV, top wages had grown disproportionally, mid-income jobs had disappeared and a larger low-paid sector had emerged. At the same time, employment in performance entertainment flat lined, while it grew quickly in other leisure activities (Figure 6). Such concentration of demand on a few stars and the pattern of rising dispersion in log pay is at odds with standard skill demand models but is consistent with superstar effects.

To test the role of superstar effects, one would ideally randomize market reach of workers across labor markets and test if the income distributions change in line with the superstar model. The staggered introduction of television gets close to this ideal by exogenously varying entertainer market reach across local labor markets. I test the predictions of the superstar model in a difference in differences regression, which compares local labor markets \( m = 1, \ldots, M \) over time \( t \). The difference in differences regression is:

\[
Y_{mot} = \alpha_m + \delta_{ot} + \gamma X_{mt} + \beta TV_{mt} + \epsilon_{mot}
\]

where \( \alpha_m \) and \( \delta_{ot} \) are labor market and occupation-year fixed effects; \( X_{mt} \) a vector of time varying labor market characteristics and \( Y_{mot} \) one of the outcomes predicted to respond in the superstar model. The treatment variable \( TV_{mt} \) counts the number of TV stations producing local shows. The treatment variable varies over \( m \) and \( t \) and identifies treatment effects from differential changes across local labor markets over time. I run the regression at the more disaggregated occupation-labor market-year level to control for potential time fluctuations in the occupation definition with occupation-year fixed effects. The standard errors \( \epsilon_{m,o,t} \) are clustered at the local labor market level, so that running the analysis at the occupation-labor market level will not artificially lower standard errors.

We expect to see a local effect of a TV station because early TV stations mainly filmed locally and hired local entertainers. Non-local shows were a poor substitute for local

\(^{40}\)If the top tail of the distribution is not observed, I use Pareto interpolation to estimate top income shares. The procedure follows a large literature that uses Pareto interpolation to estimate top income shares (Kuznets and Jenks (1953); Feenberg and Poterba (1993); Piketty and Saez (2003); Blanchet et al. (2017)) and is described in Appendix D.2.
productions for two reasons. First, the infrastructure to air shows simultaneously across stations was lacking. Sterne (1999) gives a detailed account of pioneering efforts to build a national TV network and the major technical obstacles. While in principle storing and transporting shows was feasible, the technology was costly and turned out to be unpopular because it led to poor image quality.\(^{41}\) Second, regulation specified that “the main studio be located in the principal community served” (FCC annual report 195). The coefficient \(\beta\) will capture the local effect of TV stations.

The treatment variable \(TV_{mt}\) varies over time from the launch, but also from the subsequent obsolescence of local TV filming. The first commercial television stations were launched on July 1st, 1941, but many regions did not get stations until years later.\(^{42}\) The roll-out was based on an FCC license process that specified where TV was to be launched and issued licenses accordingly. Television thus remained a niche technology during World War II, as production restrictions stopped the manufacturing of TV-related equipment. From 1945 onward, television spread rapidly and by 1949, 124 stations were active. Figure 3 illustrates which local labor markets had been treated. Launch dates across areas differed by as many as 15 years because multiple delays interrupted the roll-out.

The removal of local production gives rise to a second quasi-experiment impelled by another technological innovation: videotape. The invention of the Ampex videotape made it possible to store and transport TV productions cheaply, which transformed the TV production industry.\(^{43}\) Shows from outside the local labor market became a close substitute for local live shows. This led to the demise of local TV production and the concentration of TV production in two hubs, Los Angeles and New York.\(^{44}\) The videotape proved an instant hit with TV stations. For example, when the product was presented at the National Convention of Broadcasters in 1956, over 70 videotape recorders were ordered immediately by TV stations across the country. The same year, CBS started to use the technology, and the other networks followed suit the next year. Local TV stations’ effect on entertainers would subsequently fade, while production hubs, by contrast, started to serve vastly bigger audiences. I interact a dummy for the time period of national production with a time-invariant proxy for local production cost to capture the centralization of production.\(^{45}\) This variable will pick up incentives to move filming to a given place, while the measure avoids the potential endogenous control problem.

The remainder of this section tests the four propositions derived above: disproportionate growth of top pay; falling pay for mediocre workers; employment loss and income dispersion among top earners.

\(^{41}\) Non-local content had to be put on film and shipped to other stations, where a mini film screening was broadcast live. This costly technology, known as “kinescope”, resulted in poor image quality and was therefore unpopular. There are notable exceptions, however. A handful of stations, mainly along the East Coast, experimented with various forms of interconnection (e.g. microwave relays, stratospheric broadcasting and coaxial cable connections).

\(^{42}\) Experimental broadcasting existed since the 1920s and had familiarized the population with the new technology. Prior to the launch of commercial television the private ownership of TV sets was however minimal. In four cities experimental broadcasters where later turned into commercial television channels.


\(^{44}\) This trend was also helped by the contemporaneous roll-out of coaxial cables that allowed to transmit live shows from station to station.

\(^{45}\) See Section 3 for the construction of the proxy.
4.1 Effect on Top Earners

The core prediction of the superstar effect is the sharp wage growth at the top of the distribution (see Proposition 2.1). As a first test, I analyze how local entertainer pay at the 99th percentile responds to TV. For this quantile regression, I use the difference in differences estimator developed in Chetverikov et al. (2016). I find the launch of a TV channel has a large, highly significant positive effect on top pay. Wages at the top rise by 17 log points, or approximately 19% (see panel A in Table 1).

The magnitude of these effects is easier to interpret if compared to wage changes in the aggregate US distribution. The rest of the paper therefore focuses on the position of entertainers in the US wage distribution (see Section 3 for the variable definitions). A first test looks at \( p^{99}(\xi_t) \), the share of entertainers among the top 1% of US wage earners. In 1939 about 4% of entertainers in labor markets that subsequently received TV stations were paid at this top level. TV nearly doubles the share of local top earning entertainers (see panel B in Table 1) which implies that the share of top earners increases by roughly 4 percentage points. A related approach normalizes the local top earning entertainers by the local population. Effects on top earning entertainer per capita are comparable to the previous results (panel C of Table 1).

A related question is whether TV broadened the availability of high quality entertainment. What constitutes good entertainment is however subjective. Prices act as a reasonable proxy for the value consumers attach to different entertainers. I therefore look at willingness to pay for different entertainers and build a short panel for a subset of local TV entertainers that allows me to test whether the most valued entertainers benefit from TV.\(^{46}\) The panel reveals that TV did not lead to substantial leapfrogging in the wage distribution, the vast majority of TV stars were in the top tail of the wage distribution even before TV (Figure 7). Television therefore predominately promoted the market reach of highly talented entertainers.\(^{47}\)

4.1.1 Probing the Identification Assumption – TV filming

The identification requires that TV launches are unrelated to local demand shocks. With ordinary technology adaption this is rarely the case. In this setting however, government rules prevent the free spread of TV. The roll-out of TV is constrained by a licensing system that determined where TV would launch next. While places were not chosen at random, multiple features of the assignment rules make it likely that TV launches were unrelated to local demand shocks. A salient point for the analysis is that priority for locations was based on fixed local characteristics. The 1952 “Final Allocation Report” for instance ranks suitable locations by their local population in 1950. Once we condition on local fixed effects, differences in treatment arise quasi-randomly. In practice there are two reasons why this approach may not work perfectly. A first challenge is that the implementation of TV launches may differ from the rules. In practice, rules were strictly enforced: The TV license specified a deadline for a stations’ start date and failure to comply could result in license withdrawal. This left little room to deviate from the government-dictated roll-out schedule. A second challenge is that regulator decision rules were not published for all years. Decisions in unobserved years potentially responded to local demand shocks. A battery of robustness checks will investigate this possibility.

\(^{46}\)See Section 3 for the details on the data

\(^{47}\)An alternative interpretation is that TV promotes the same kind of talent that was required to be successful in the pre-TV era.
An initial check is to control directly for time-varying changes in local labor markets in the regression. I run two specifications, one controlling for time varying local characteristics and one that allows for local specific trends (column 2 and 3 in Table 1). If spurious trends where driving the findings, the treatment effect should disappear. Both specifications, however, find effects very similar to the baseline. The second approach, is a very demanding specification as it adds more than 700 additional location specific trends and standard errors increase accordingly.

**Freeze Stations** A halt in licensing can be used to test whether areas that are about to receive TV are affected by spurious local demand shocks. The regulator shut-down stopped the planned roll-out in 1948 and introduced quasi-random variation in TV launch dates. During the shut down, all ongoing license procedures were put on hold and many locations narrowly missed out on receiving TV. The time pattern of approvals is shown in Figure 8 and shows the sharp drop in approvals.

The reason for the sudden shut-down was an error in the assignment model. The FCC used a signal propagation model to calculate which broadcast frequencies and catchment areas were safe to use, avoiding interference between stations. An error in the model resulted in interference occurring among licensed stations. To avoid such problems from becoming worse, the FCC ordered a review of the model and put all ongoing license procedures on hold.

I digitize data that allow me to observe where licenses were imminent but did not proceed as planned because of the FCC review. Figure 3 shows the affected local labor markets. During the freeze period, the FCC undertook extensive field studies and expert hearings to improve the scientific standard of their signal model. As a result, licensing did not resume until 1952 and the onset of TV was delayed by nearly four years in many markets.48

We can use such “stations that did not happen” to test whether the introduction of television coincided with spurious location specific shocks to top pay. To test this prediction, I implement a dynamic version of the previous difference in differences regression:

$$Y_{mot} = \alpha_m + \delta_{ot} + \gamma X_{mt} + \sum_t \beta_t TV_{mt} + \epsilon_{mot}$$  \hspace{1cm} (13)$$

where now the treatment variable $TV_{mt}$ are stations that were blocked by the freeze. We can therefore use another natural experiment to test the identifying assumption of the difference in differences setting.

Local labor markets that narrowly missed out on a TV station experienced no top income growth for entertainers. The point estimate is a precise zero. Figure 9a illustrates the time path of the coefficients on “stations that did not happen”. Areas affected by the freeze and untreated labor markets follow the same time path. This test is arguably more convincing than a pre-trend test or a test based on placebo occupations, as we can test for local shocks in the same occupations and year. For peace of mind, such related checks of pre-trends and placebo occupations are reported in Appendices D.1.3 and D.1.2.

---

48Initially the freeze was only expected to last about a year. However, additional technical developments prolonged the freeze period. Beside reconsidering the assignment of existing frequencies, the FCC started to experiment with making additional frequency bands available to television. Moreover, the FCC wanted to ensure that the new system was compatible with the arising transmission of colored images. It thus bundled the testing and processing of these issues.
Common Trend Test  Common trend tests add further credibility to the identification assumption. In my setting I observe treatment and control groups in an untreated state both before and after the removal of local TV show filming. I use both pre- and post-treatment periods to test the common trends assumption. If common-trends hold, the treatment effect arises when local TV productions are introduced and disappears when they are removed.\footnote{This requires that the temporary increase in available market size has no lasting effect.} This test uses the dynamic difference in differences specification in equation 13, with local TV stations as regressor. The time-path of $\beta_t$ reveals how differences between treatment and control group change over time and are plotted in Figure 9b. A difference in treatment and control areas appears during local TV filming and disappears after the end of local TV. By 1969 the differences between treatment and control group returned to the pre-treatment level, which suggests that the common trend assumption holds.

4.2 Concentration of Consumer Demand

Next, I directly test if changes in labor demand align with the prediction of the superstar model. A unique feature of this setting is that we observe demand for local entertainers, which has the advantage that shifts in labor demand can be measured even if wages respond sluggishly. Recall that the superstar model predicts growing demand for star entertainers and falling demand for mediocre workers.

To the best of my knowledge, no existing dataset with sufficiently disaggregated data to examine demand within the entertainment category at the local level is available. I therefore hand-collect novel data to document local spending on entertainment. The records cover over a decade of spending records at thousands of country fairs. Details on the data are documented in Section 3. Of course, county fairs represent only a fraction of overall entertainment spending, yet these data afford a look at demand for entertainment at the local labor market level, annually from 1946 to 1957. During this period there is substantial regional variation in exposure to TV. Moreover, I collect data on show audiences to explore demand concentration at the top end of the market. The market size variable measures the potential audience of a single performance in a given labor market.

First, I estimate how TV affected market reach of star entertainers. I run the baseline specification, regressing log of audience size of the biggest local shows on the introduction of local television production. The results show that the most widely watched entertainers experienced a dramatic growth in audience size (see panel A. of Table 2). The launch of a television station increased the audience of the largest shows by about 150 log points. Converting the log points to a growth rate shows a growth rate of over 300%, which translates to a fourfold increase in market size. Next, I estimate the change in market reach in dollar terms, this quantifies the effect on marginal revenue product that went hand in hand with the growth in audience size. In dollar terms market reach of stars roughly tripled (see panel B. of Table 2). The launch of a TV station thus dramatically increased the market value of top talent.\footnote{Note that audience grows faster than revenues, reflecting a drop in the unit price for entertainment. Star entertainment became more widely available and simultaneously more affordable.}

Second, I estimate how TV signal affected demand for local live entertainment. Differently from the previous regressions, this analysis uses TV signal rather than TV filming as regressor. The two differ because signal often travels beyond the local labor market where filming takes place. I therefore compare demand for local entertainment while signal becomes available
locally. The introduction of TV signal reduces county fair ticket revenues by about 5% (column 1 of Table 3). The decline in revenue reflects a similar decline in fair attendance (column 2). People staying away from county fairs can therefore account for fairs’ declining ticket revenues. These results are, however, noisy as they hide substantial heterogeneity.

Demand effects differ markedly for different types of entertainment. More specifically, demand for entertainment that is similar to TV falls significantly, while demand for entertainment that is very different from TV holds up. I collect data on receipts for two extremes of substitutability: grandstand shows and traditional carnivals. Grandstand shows were quite similar to many TV shows at the time and included vaudeville acts, thrill shows, dance groups and beauty pageants. Traditional carnival activity, including candy sales and amusement rides, engage more than the visual senses and are inherently less substitutable by television. As columns 3 and 4 of Table 3 report, grandstand show receipts see a large decline, while carnival receipts are unaffected.

Next, I analyze the same data at a finer local level, which arguable allows me to measure TV exposure with greater accuracy. This is feasible because the precise location of fairs is available in my data. Running the regressions at the county level increases the power of the set-up. However, this faces the drawback that we must assume that fairgoers come only from the county where the fair is held. Results at the county level are in line with the previous findings (panel B of Table 3). TV again decreases demand for fair entertainment. The point estimates are smaller than before, which plausibly reflects the fact that we are mismeasuring TV signal exposure for a sizable fraction of fairgoers who live outside the county where the fair is located. Overall, these result confirm that TV signal reduces demand for local live entertainment and increased demand for star entertainers. The marginal revenue productivity of entertainers shifts substantially in favor of stars.\footnote{\textsuperscript{51}}

### 4.3 Effect on Non-Stars

#### 4.3.1 Cannibalization of Demand for Non-Stars

**Hollowing out** A second prediction of the superstar model is that technology causes demand to shift from a profession’s mediocre workers towards its stars. As a result, we should observe declining pay for non-stars (see Proposition 2.2). The reduced consumption of county fair entertainment already points in this direction and shows that the revenue product of non-star entertainers declines because of TV. Next, I will study how this shift affects wages of non-stars. Consider, entertainers who are below the top 90th percentile of the US wage distribution but still in the upper quartile ($e_{75,90}$). These are entertainers who receive above-average pay but are far from the top of the entertainer distribution. The launch of TV production has a significantly negative effect on this group, as the number of jobs that pay in this range declines by around 50%. The results look similar between the median and the 75th percentile (results are reported in Figure 10). Television therefore leads to a substantial decline in well-paid jobs below the star level. This suggests that it is substantially worse to be a mediocre entertainer during the TV era.

The corollary to disappearing mid-paid jobs is the growing low-pay sector. Analyzing the share of entertainers paid below the median, we observe a modest rise in the share of entertainers with wages at the very bottom of the distribution, with little change in the

\textsuperscript{51}Notice that the productivity results rule out that the observed wage effects are purely driven by a shift in bargaining power.
second quartile. Television thus reduces the payoff of non-star talent and creates a growing low-pay sector.

**Employment** A third prediction of the superstar model is that falling demand for mediocre entertainers leads to a decline in employment. Proposition 2.3 shows that decreased employment sets the superstar model apart from standard models, where demand for skill shifts outward.

To test for employment effects among entertainers, I compare local labor markets with differential access to TV signal. The Census records information on employment information for longer than wages, I therefore expand the sample period backward by a decade to 1930.

The introduction of a TV channel leads to sizable employment losses of around 13% (Table 4 column 1, panel A for the extended sample and panel B for the baseline sample). This result is in line with a falling price for talent and thus reduced revenues at local entertainment outlets. Lower returns push entertainers out of the market and entertainment moves closer to a winner-takes-all market. The result is sharply at odds with models where technical change causes a positive demand shock. Such a positive demand shock, in contrast, would raise employment.

### 4.3.2 Probing the Identification Assumption – TV signal

Since the previous results use different identifying variation, it is salient to probe again whether the identifying assumption holds. As before, I first control for proxies of local demand shifts with time varying local characteristics and local trends. Columns 2 and 3 of Table 4, respectively, show that the coefficient of interest remains unchanged, implying that we are identifying fairly sharp changes around the time of treatment. With the expanded sample a conventional common pre-trend test is feasible. I implement this test by introducing a lead to the treatment variable in the regression. This captures differential changes in treatment and control areas, right before the onset of the actual treatment. The coefficient on the lead is small and insignificant and indicates that the treatment and control groups look similar in the run up to the treatment (column 4).

The TV roll-out freeze can again be used to test the identifying assumption. Panel C in Table 4 exploits this variation and reports the employment effect of TV signals that did not happen because of the freeze. Such signal has no effect, confirming that there are no spurious trends in areas next in line for licensing.

### 4.3.3 Fractal Inequality

A final implication of a superstar model is that technical change widens the wage difference between stars and their slightly less talented peers, while skill-biased demand shocks move wages proportionally (see Proposition 2.1). A non-parametric test of this prediction repeats the baseline difference in differences regression, focusing on lower percentiles. Take, for example, entertainers who are below the top 1% but still among the top 5% of the US wage distribution \( e_{95.99} \). I find that television leads to a 12% increase in \( e_{95.99} \). This effect is only one tenth the size of the effect at the very top. Television therefore disproportionally benefits the superstars. To confirm this pattern we can look at the next lower wage bin. The

\[ \text{For employment, we cannot rely on pre- and post-periods to identify counterfactual trends as TV signal, unlike local filming, is not removed.} \]
effect of television on entertainers in the top 10% but below the top 95th percentile \( (e_{e,90,95}) \) is insignificant with a negative point estimate, again confirming that television's effect fades quickly as we move away from the top stars in the market.

Combined with the above findings on stars and mediocre entertainers, these results show that the impact of TV across the wage distribution is U-shaped (Figure 10). TV has a large positive impact at the very top and the effect shrinks as we move down the distribution, turning negative below the 90th percentile. The plot also shows growth in very low-paid entertainment jobs.

Next I focus on pay dispersion within entertainment. This is closely related to the previous results but focuses on measures widely used in the literature on inequality. Proposition 2.4 suggests two measures: i) percentile ratios and ii) top income shares.

The percentile ratio picks up the wage difference at two points of the distribution. I construct this ratio for each of the local entertainer labor markets and for all permutations of the 99th, 90th, 75th and 50th percentiles. All of these ratios show wages fanning out in response to TV filming. The hypothesis that these ratios are unaffected is rejected at the 95% confidence level for any ratio (see Table 5). The second test for growing pay dispersion in the top tail is the ratio of top income shares (as used by Gabaix et al. (2015)). Prior to TV, the fraction of income going to the 1% highest earners in a local labor market was, on average, 3.8%. TV filming increased the top income share by 3.7 percentage points, which means the top 1% income share almost doubled. Most of the gains accrued to the very highest earners in the top 1%. The top 0.1% of entertainers saw their income share rise by 2.4%. This group is only one tenth of the top 1% but accounts for over half of the rise for the top 1% income share. The share of income going to the top 1% grows by 100% and the equivalent share for the top 0.1% grows by 300%, but the top 10% share grows by only 30%. A formal test of equal growth rates of either ratio is strongly rejected. This is at odds with models of skill-biased demand where wages grow proportionally across all percentiles but is in line with the superstar model where wage growth is skewed towards the top of the distribution.

In the appendix, I confirm these results with a set of quantile regressions (see Appendix D.1.4). These estimates look at changes in the wage quantiles within entertainment and show the same pattern. The effect of technology declines remarkably quickly in the top tail. TV appearances help a small group of top stars, has moderate effects on backup stars and has no discernible benefit for other top earners.

4.4 Links Between Markets

So far, I treated local labor markets as independent markets. In practice this may not hold as both output and workers may move. Links between labor markets will change the interpretation of the results, it is thus important to understand the magnitude of such links.

A first spillover comes from trade in output. If trade in output equalizes prices across local labor markets, we would not pick up any local effects. A convenient feature of live entertainment is that shows are consumed where they are produced. Live shows are therefore a non-tradable service and we do not need to worry about trade in output.\(^{54}\)

\(^{53}\) The equivalent number for the US economy as a whole is about 10%. It is however unsurprising that within a given region and industry income is less dispersed.

\(^{54}\) TV shows, by contrast, can be watched beyond the local labor market. This occurs if TV signal travels beyond the local labor market. I have data on signal reach and can thus account for links between markets.
A second spillover comes from mobility of workers. First, note that the biggest labor markets receive TV first and, as a result, place ranking by audience reach remains unchanged and generates little new incentive to relocate. To back this claim up empirically, I can use data on mobility reported in the US Census.\textsuperscript{55} I find TV has very few effects on mobility. The regression coefficient of television’s effect on the probability of moving is a precise zero. In fact, the point estimates are even negative. We can rule out that mobility increased by more than 1% (columns 1 - 3 of Table 7). Even at the upper end of the confidence interval, migration cannot explain a substantial share of the baseline findings.\textsuperscript{56}

An alternative approach is to focus on labor markets that are further apart. Moving is arguably easiest between neighboring markets, so we should see most of the migration take place across that margin. By excluding control areas that neighbor treated labor markets, we will clarify the importance of such spill-over effects.\textsuperscript{57} Results that exclude neighboring areas are close to the baseline (see panel B of Table 7), suggesting again that relocation between neighboring markets does not explain the findings.

5 Magnitude of Superstar Effects

The test for superstar effects confirms that labor market responses align with the model’s predictions, but this does not reveal whether superstar effects are large or small, nor how much of top income growth can be explained by superstar effects. Expressing the results in terms of elasticities is the most effective way to determine the magnitude. This elasticity also captures the superstar model’s key structural parameters. The magnitude of the elasticity depends on the relative scarcity of talent and market size, as well as on the complementarity of these factors (see equation 7). The regression analogue to equation 7 is:

\[
\ln(w_{m,t}^{99}) = \alpha_0 + \alpha_1 \ln(s_{m,t}^{99}) + \epsilon_{m,t}^{99}
\]

where $w_{m,t}^{99}$ is the 99th percentile of the entertainer wage distribution in market $m$ and year $t$ and $s_{m,t}^{99}$ the size of the market that such entertainers can reach. In a first step, I estimate this relation using a naive OLS estimator on a single cross-section. This uses the variation in the size of the biggest local theatre size in 1939 as regressor. A large literature has estimated similar regressions by correlating firm size and pay. In line with those results my cross-sectional OLS estimate of $\alpha_1$ is highly significant with a point estimate of 0.23 (see panel A. of Table 8) Moving from a local labor market with a small theatre to a market twice the size, increases pay for a top earner by 23%. The effect may of course reflect differences in local labor markets, rather than the effect of market reach. Indeed, after controlling for

\textsuperscript{55}The measure of mobility is noisy for two reasons. First, the migration question does not distinguish between moves within and across labor markets, about 50% of moves are within the same county, such moves do not affect the analyses. Moreover, the Census question changes over time. It asks whether a person has moved in the last X years, but X differs between Census years. Noise in the outcome variable will inflate standard errors but not necessarily bias the estimates. The usefulness of the estimation results therefore depend on their precision.

\textsuperscript{56}Take the extreme case where all additional inflow comprises entertainers in the top 1%. This would imply that the share of top-earning entertainers increase by about 1%, less than a quarter of the observed increase. Migration’s role is sufficiently small that the inequality effects are mainly driven by changes in returns to skill.

\textsuperscript{57}This excludes 2,990 labor markets, or 1/4 of the sample.
local characteristics, the effect disappears almost entirely (column 2 of the same Table).

To estimate the causal effect of market reach, I turn to the exogenous shock in market reach from TV. I use an IV approach that instruments market size with the roll-out of TV. The first stage regression, the effect of TV on audience size, is large and highly significant (recall Table 2) and the associated first stage F-statistic is well above conventional cutoffs with a value of 20. The IV estimator of the elasticity $\alpha_1$ is also highly significant with a point estimate of 0.17. This implies that wages at the 99th percentile grow 17% when market size doubles. While this wage effect is sizable, the effect is 30% lower than the cross-sectional OLS estimate above. This suggests that the causal effect of market reach is smaller than the correlation of market size and top pay suggests.

The literature on product market concentration measures changes in market size in terms of revenues, rather than measuring concentration of customers. To link my results to this literature, I compute show revenues. The dollar value of an entertainment show is calculated as the product of the audience measure and ticket prices for theaters or advertisement prices for TV shows. Notice that aggregate changes in prices will not affect the regressor, since the log specification absorbs such price changes in the time fixed effects. Results are affected by changing price dispersion between shows. Using revenue as regressor makes it easier to compare results across settings but has the drawback that it uses an outcome, prices for talent, as part of the regressor of interest. Non-withstanding this limitation, I estimate the elasticity of top pay to market value and find again a highly significant effect with a point estimate of 0.22. One dollar greater concentration in the product market therefore leads to 22 cents higher pay for star workers.

6 Market Structure, Policies and Superstar Effects

How do the above findings translate to other settings? The roll-out of TV provides rich heterogeneity for testing superstar effects in different institutional settings. With this variation superstar effects in labor markets with imperfect competition, differentially skilled workforce and different tax rates can be analyzed.

6.1 Monopsonistic Employers

Consider the case of imperfect competition. The licensing process limits employer entry into labor markets. This variation can be used to test the interaction of superstar effects with imperfect competition. The benchmark superstar model is perfectly competitive and market concentration arises without rents but predictions change sharply with imperfect competition. Employers with market power will not pass-on the surplus from greater scalability of production. To test this prediction I allow for differential effects of markets

---

\footnote{Estimate the OLS with panel data would compare wages across local labor market wages over time, as market reach changes. However, in my data variation in market reach within a local labor market over time comes exclusively from the launch of TV. My data on theatre capacity does not vary over time and the panel OLS is therefore mechanically close to the IV estimate (results are available upon request).}

\footnote{Price data is only available for a subset of observations. I infer prices based on a data from TV station ad-pricing in 1956 and theater ticket prices in 1919. I estimate the size-price gradient separately for TV and theaters and assume that this relation is time-invariant.}

\footnote{Note that this estimate is bigger than the elasticity with respect to audience size. A fact that arises because the launch of TV reduced the per-head cost of top entertainment. The reduced form of both elasticity IV estimators is the same, a smaller first-stage therefore increases the IV estimate.}
with a single TV station and markets with multiple TV stations. Since this dummifies
the previously continuous treatment variable, I first report the baseline regression with a
dummy treatment variable (column 1 of Table 9). I then introduce the additional dummy
for labor markets with multiple TV stations (column 2). The effect on labor markets with
multiple TV stations is therefore the sum of the two dummies. The differences between
monopsonistic and competitive labor are striking. Markets with a monopsony employer see
almost no top income growth, while gains are large when there is more than one employer.
Using information on the freeze, I identify labor markets that would have had competition
from a second station if the freeze had not blocked the competitor. The results confirm the
baseline findings; locations that would have experienced employer competition but miss out
see next to no top-wage growth (column 3). Only when employers face competition, does
greater market scale translate into rising wages.

Note that this result suggests that employers did not substantially share rents with their
workers. The entry of an additional TV erodes the incumbents monopoly power, yet this
does not harm pay of top workers. To the contrary, the entry of a competitor substantially
raises wages. Any loss of rent-sharing rewards among workers is therefore off-set by growing
wages from competition.

We can also study the reverse, the possibility that a TV station’s exogenous entry breaks
up previously non-competitive structures. To investigate this possibility, I allow the effect
of TV to differ across labor markets with different numbers of pre-TV employers. The
result, as reported in column 4 of Table 9, is a fairly precise zero. There is no differential
effect across this dimension. This suggests that the pre-TV labor market of entertainers was
reasonably competitive or that the differences that arose from imperfect competition are
negligible relative to the effect of greater scalability. Another possibility is that the number
of employers in the pre-period is a poor measure for competition. To address this, I use an
alternative proxy of labor market competitiveness, population density. Here again, I find
no effect (column 5), which suggests that pre-TV labor market competitiveness does not
greatly influence superstar effects. Given the magnitude of the top income growth, this is
unsurprising.

6.2 Policies

A leading policy to battle inequality is investment in education. Arguably, modern production
technologies require greater skill and are therefore driving up demand for skilled workers. In
line with this argument, the wage premium for skilled workers has been rising (Acemoglu
and Autor (2011); Katz and Murphy (1992)). Investment in education would increasing
the relative supply of skilled labor and thereby reduce inequality. In superstar models by
contrast, the level of education does not affect inequality. In such models, the rank position in
the ability distribution determines pay differences. Changes in the skill level of the workforce
have no material effect on inequality in this model. I can test this prediction empirically by
interacting the treatment with the local high school graduation share, which is admittedly a
rough proxy for education levels but has been widely used in the literature on inequality.
The interaction is insignificant, suggesting that the superstar effect is independent of the skill
level in the local labor market (column 6 of Table 9). Since the standard errors are large,
these results can however, not be interpreted as conclusive evidence against skill investments.

Taxes are another popular tool to reduce top income inequality. If part of the superstar
effect is a result of increasing work effort by star workers, higher tax rates may reduce wage
inequality by reducing stars’ incentives to increase their effort. The empirical literature on taxes and superstars has mainly focused on migration. Mobility of taxable income of stars responds significantly to differential tax incentives across states or countries (Kleven et al. (2014); Moretti and Wilson (2017); Kleven et al. (2013)). Mobility may, however, only be a small part of superstars’ behavioral response. In all of these studies, the share of movers is small and the associated distortion from migration might be dwarfed by labor supply changes by stayers. Piketty et al. (2014) suggest that markets where a lot is at stake encourage rent extraction, which would imply large income elasticities in superstar markets. Similarly, Scheuer and Werning (2017) also argue that tax rates lead to large elasticities in superstar markets. In contrast to the rent story, they argue, elasticities are high because taxation could distort the assignment of workers to markets which would generate additional distortions.

I test whether superstar effects differ under different tax regimes. This test exploits variation in top income tax rates across US states. Data on states’ historical tax rates are not centrally collected. I compiled such data from the study of historical state taxation in Penniman and Heller (1959), who collect detailed information on income tax legislation across US states during the sample period. Using this information I construct a dummy variable that is equal to one for high-tax states, aka states with tax rates above the median.

I test how higher tax rates affect the rise in top incomes in a superstar setting. This estimate combines the effect of out-migration and reduced labor supply by stayers. Column 7 of Table 9 shows there is no significant difference between high- and low-tax states. While the standard errors are large, the point estimate on the interaction term is quantitatively close to zero. There is thus no evidence that high taxes lead to substantial distortions in superstar markets, nor that taxes are able to substantially slow the rise of superstar earning.

7 Conclusion

Little is known about the causes of the vast changes in top incomes observed in recent decades. Superstar effects link these changes to technical innovation, particularly in communication technologies, that make it easier to operate over distances. This paper provides causal evidence on the effect of growing production scalability on wages and provides an empirical test of the superstar model.

To test the superstar model, I exploit quasi-experimental variation in market reach in the entertainment industry and show that the staggered introduction of TV substantially changed audience sizes for entertainment shows. Star entertainers increased their audiences fourfold through TV. I show that this expansion generated superstar effects. TV doubled the number of top earners in entertainment. The effect was concentrated on the stars of the profession, with much smaller effects on slightly less talented workers. In line with this result, wage dispersion among top earners rose. I also find evidence for negative effects on mediocre workers. Stars’ ability to reach large audiences reduced the return for mediocre workers and pushed some out of the market. I find evidence for both of these effects. Competition from TV reduced the number of mid-paid entertainer jobs and reduced employment by about 13%.

61 I use a binary variable because marginal tax rates are difficult to interpret in this context. Deductibility rules generate a wedge between MTR and headline rates. This is less of a problem for comparing high- and low-tax states to the extent that deductibility rules don’t change whether a state is a low- or high-tax state.
The increase in production scalability has profound effects on inequality at both the top and bottom of the distribution. These results are at odds with conventional models of technical change but align with the predictions of a superstar model. To assess the magnitude of superstar effects, this paper provides top income elasticities with respect to market size. The estimates imply that one extra dollar in product market concentration leads to 22 cent higher pay at the 99th percentile. Similarly, the share of income going to the top 1% increases by 30% if market size doubles.

A key implication of the evidence for superstar effects is that rising market concentration could be a sign of technical progress. Conclusions that market concentration necessarily indicates malfunctioning markets might therefore be premature. Market power is better measured by the ability to set prices than by the ability to command a large market share. To evaluate inefficiencies associated with top income concentration, it will be important to distinguish cases where superstar effects bring better quality to a greater share of consumers from cases where market concentration results from competition break-down.
References


[Notes] The figure shows model predictions from the superstar model. Panel A shows equation 8 and illustrates the change in the wage distribution when production becomes scalable between $t$ and $t+1$: $\delta' = s \cdot \delta$ and $\gamma' = s \cdot \gamma$ with $s > 1$. Panel B illustrates the same change in terms of employment growth across wage bins. The figure shows equation 10 for parameterization $g_w = 0.2x^{(1.3)} - 1$. 

$$\ln(Pr[x > \text{wage}])$$

$$\ln(\text{wage})$$

(b) Employment Growth at Different Wage Levels

Wage Bins

Employment Growth
Figure 2: Effect of Technical Change on Wage Distribution – Skill Biased Demand Model

\[ \ln(Pr[x > wage]) \]

[Note] The figure shows an increase in the skill premium for a group of workers with heterogeneous but perfectly substitutable talent.
Figure 3: Intensity of TV Exposure in 1949

[Notes] The size of a symbol indicates the number of stations per local labor market. Active stations are blue circles, frozen stations red triangles. Source: FCC reports as reported in Television Factbooks and TV Digest. Location is derived from geocoding addresses.
Figure 4: TV Signal of Licensed and Frozen Stations

[Note] Signal coverage is calculated using an Irregular Terrain Model (ITM). Technical station data from FCC files, as reported in TV Digest and Television yearbooks, are fed into the model. Signal is defined by a signal threshold of -50 of coverage at 90% of the time at 90% of receivers at the county centroid.
Figure 5: Entertainer Wage Distribution 1940 and 1970

[Note] Log real wage distribution for performance entertainers from the lower 48 states of the US from the 1940 and 1970 Censuses. Real wages in 1950 USD using Census sample weights. Density is estimated using the Epanechnikov smoothing kernel with a bandwidth of 0.4. Common top code applied at $85,000.
[Note] The mean for performance entertainers is 49 and for other leisure occupations 468. Data are from the US Census and cover the mainland US. Employment is measured per 100,000 inhabitants. For consistency with early Census vintages, the employment measure includes the unemployed when they report an occupation. Performance entertainment is defined in the main text. Other Entertainment includes “drink and dine” occupations and “interactive leisure” professionals.
Figure 7: Position of Future TV Stars in 1939 US Wage Distribution

[Note] TV stars are defined in the “Who is Who of TV” in “Radio Annual, Television Yearbook 1950.” These individuals are linked to their 1939 Census records. 1939 wages are corrected for age, education and gender. The position in the distribution is calculated based on the residual of a regression of log wages on a cubic in age, 12 education dummies and a gender indicator.
Figure 8: Number of TV Licenses Granted

[Note] Data from TV Digest report on FCC licensing activity. Missing construction permit dates are inferred from start of operation dates.
Figure 9: Dynamic Treatment Effect

(a) Blocked Stations

(b) TV station

[Note] Figure plots coefficients from two dynamic difference in differences regressions. Panel a) shows the coefficient on $F_{rozenTV_{m,t}}$ (comparison groups are untreated areas) and panel b) shows the coefficient on $TV_{m,t}$. Top-paid entertainers are in the top 1% of the US income distribution. Vertical lines mark the beginning of local TV ("TV") and the end of local TV ("Videotape"). The area shaded in light blue marks the 95% confidence interval. Standard errors are clustered at the local labor market level.
[Note] Each dot is the point estimate of a separate regression and shows a TV station’s effect on the change in entertainer employment in a given wage range. Percentile bins are defined in the overall US wage distribution. Standard errors are clustered at the local labor market level with 95% confidence intervals.
## B Tables

Table 1: Effect of TV on Top Earners

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: ln(Entertainer Wage at 99\textsuperscript{th} Percentile)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local TV station</td>
<td>0.171</td>
<td>0.158</td>
<td>0.154</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.023)</td>
<td>(0.030)</td>
</tr>
<tr>
<td>Effect size</td>
<td>18.6%</td>
<td>17.1%</td>
<td>16.6%</td>
</tr>
<tr>
<td><strong>Panel B: Share Entertainer in US Top 1%</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local TV station</td>
<td>4.14</td>
<td>4.31</td>
<td>5.93</td>
</tr>
<tr>
<td></td>
<td>(1.26)</td>
<td>(1.27)</td>
<td>(2.21)</td>
</tr>
<tr>
<td>Effect size</td>
<td>92%</td>
<td>96%</td>
<td>132%</td>
</tr>
<tr>
<td><strong>Panel C: Entertainer in US Top 1% Per Capita</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local TV station</td>
<td>0.40</td>
<td>0.40</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.10)</td>
<td>(0.10)</td>
</tr>
<tr>
<td>Effect size</td>
<td>133%</td>
<td>133%</td>
<td>103%</td>
</tr>
</tbody>
</table>

[Note] The outcome variable in Panel A is the entertainer wage at the 99th percentile. In Panel B the outcome variable is the share of top-paid entertainers, holding the denominator fixed, as described in the text. Panel C analyses the number of top-paid entertainer divided by the local population in 10,000s. Each cell reports the regression coefficient of a separate regression. Panel A uses a quantile regression for within group treatment (Chetverikov et al. (2016)). For this procedure, data are aggregated at the treatment level and use 2,264 local labor market-year observations. Observations are weighted by cell size; cells where the 99th percentile cannot be computed are dropped. Panels B and C use a difference in differences regression and are based on 13,718 occupation-local labor market-year level observations. The treatment is the number of TV stations in the local area. Reported baseline outcomes are the average of the dependent variable in treated areas in years without treatment. All regressions control for local labor market fixed effects, time fixed effects and local production cost of filming in years after the invention of the videotape. Panel B and C also control for year-occupation fixed effects. The sample period spans 1940-1970. Demographics are median age, % female, % black, population density and trends for urban areas. Entertainment are Actors, Athletes, Dancers, Entertainers Not Elsewhere Classified, Musicians. Observations are weighted by local labor market population. Standard errors, reported in brackets, are clustered at the local labor market level. Panel A has 707 clusters, B and C 722 clusters. Data sources are the US Census 1940-1970 and TV factbooks for location of TV stations.
Table 2: Effect of TV on Market Access of Local Stars

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: ln(audience size)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local TV station</td>
<td>1.499</td>
<td>1.526</td>
<td>1.146</td>
</tr>
<tr>
<td></td>
<td>(0.240)</td>
<td>(0.223)</td>
<td>(0.220)</td>
</tr>
<tr>
<td>Effect size</td>
<td>348%</td>
<td>360%</td>
<td>215%</td>
</tr>
<tr>
<td><strong>Panel B: ln(revenue)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local TV station</td>
<td>1.095</td>
<td>1.116</td>
<td>1.146</td>
</tr>
<tr>
<td></td>
<td>(0.207)</td>
<td>(0.168)</td>
<td>(0.220)</td>
</tr>
<tr>
<td>Effect size</td>
<td>199%</td>
<td>205%</td>
<td>215%</td>
</tr>
<tr>
<td>Clusters</td>
<td>722</td>
<td>722</td>
<td>722</td>
</tr>
<tr>
<td>Demographics</td>
<td>–</td>
<td>Yes</td>
<td>–</td>
</tr>
<tr>
<td>Local labor market trends</td>
<td>–</td>
<td>–</td>
<td>Yes</td>
</tr>
</tbody>
</table>

[Note] Dependent variable is: panel A. the potential audience of the largest show in the commuting zone, either physical or TV households in the service area of the local station. Panel B. revenue of show. The number of observations are 2656. All regressions control for commuting zone fixed effects, year fixed effects and film amenities after 1956. Standard errors are clustered at the commuting zone level. Demographics are median age, income, share female, share minority, population density and different trends for urban areas. Sources: signal data from Fenton and Koenig (2018) and TV ownership from the US Census, see text for details.
Table 3: Effect of TV on Spending at Local County Fairs

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fair visits</td>
<td>Fair ticket</td>
<td>Show receipts</td>
<td>Carnival receipts</td>
</tr>
<tr>
<td>TV signal</td>
<td>-0.051</td>
<td>-0.047</td>
<td>-0.059</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.024)</td>
<td>(0.022)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>Clusters</td>
<td>722</td>
<td>722</td>
<td>722</td>
<td>722</td>
</tr>
<tr>
<td>Time &amp; Labor Market FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Panel A: Local Labor Market Level**

**Panel B: County Level**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TV signal</td>
<td>-0.013</td>
<td>-0.014</td>
<td>-0.018</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Clusters</td>
<td>3,111</td>
<td>3,111</td>
<td>3,111</td>
<td>3,111</td>
</tr>
<tr>
<td>Time &amp; County FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

[Note] Outcome variables are total revenues at county fairs in location $m$ in year $t$ at annual frequency from 1946 to 1957; location refers to labor markets in Panel A and counties in Panel B. All variables use the the inverse hyperbolic sine transformation and monetary variables are in 1945 US Dollars. Treatment is the number of TV stations that can be watched in the commuting zone. Data on carnival receipts (col 4) are unavailable for 1953 and 1955. Panel A is based on 8,664 observations and, respectively, 7,220 in column 4, while Panel B uses 37,332 and 31,110 observations. Standard errors are clustered at the local labor market level in Panel A and at the county level in Panel B. Source: Billboard Cavalcade of Fairs 1946-1957 and irregular terrain model for TV signal as reported in Fenton and Koenig (2018).
Table 4: Effect of TV on Entertainer Employment

<table>
<thead>
<tr>
<th></th>
<th>Ln(Employment in Entertainment)</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel A: TV Signal 1930-1970</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TV signal&lt;sub&gt;t+1&lt;/sub&gt;</td>
<td></td>
<td>0.039</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.033)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TV signal&lt;sub&gt;t&lt;/sub&gt;</td>
<td>-0.133</td>
<td>-0.127</td>
<td>-0.125</td>
<td>-0.123</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td>(0.059)</td>
<td>(0.061)</td>
<td>(0.060)</td>
<td></td>
</tr>
<tr>
<td><strong>Panel B: TV Signal 1940-1970</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TV signal&lt;sub&gt;t&lt;/sub&gt;</td>
<td>-0.128</td>
<td>-0.114</td>
<td>-0.134</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td>(0.061)</td>
<td>(0.063)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel C: Placebo TV Signal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo TV signal&lt;sub&gt;t&lt;/sub&gt;</td>
<td>0.053</td>
<td>0.044</td>
<td>0.053</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.083)</td>
<td>(0.083)</td>
<td>(0.084)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clusters</td>
<td>722</td>
<td>722</td>
<td>722</td>
<td>722</td>
<td></td>
</tr>
<tr>
<td>Labor Market FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Time-Occupation FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Demographics</td>
<td>-</td>
<td>Yes</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Local Labor Market Trends</td>
<td>-</td>
<td>-</td>
<td>Yes</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

[Note] The outcome variable “ln(Employment in Entertainment)” is the inverse hyperbolic sine of employment in entertainment. Each cell reports the coefficient from a difference in difference regression. TV signal is a dummy that takes value 1 if signal is available in a commuting zone. Subscript “t+1” refers to the lead of the treatment. All specifications control for local labor market fixed effects and time-occupation fixed effects. Panel A covers 1930-1970, panels B and C 1940-1970. Standard errors, reported in brackets, are clustered at the local labor market level. Source: TV signal from Fenton and Koenig (2018).
Table 5: Effect of TV on Wage Percentile Ratios in Entertainment

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentile Ratio ($\frac{P_x}{P_y}$)</td>
<td>$P_{99}/P_{95}$</td>
<td>$P_{99}/P_{75}$</td>
<td>$P_{99}/P_{50}$</td>
<td>$P_{95}/P_{50}$</td>
<td>$P_{75}/P_{50}$</td>
</tr>
<tr>
<td>Local TV station</td>
<td>0.04</td>
<td>0.131</td>
<td>0.206</td>
<td>0.168</td>
<td>0.0745</td>
</tr>
<tr>
<td></td>
<td>(0.0161)</td>
<td>(0.0377)</td>
<td>(0.0491)</td>
<td>(0.0542)</td>
<td>(0.0154)</td>
</tr>
<tr>
<td>Time &amp; Labor Market FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

[Note] Outcome $\frac{P_x}{P_y}$ is the percentile ratio of percentiles $x$ and $y$ of the local entertainer wage distribution. Percentiles are calculated using the provided sample weights. Regressions control for year and labor market fixed effects and local production cost of filming in the years after 1956. Observations are weighted by cell size. Standard errors, reported in brackets, are clustered at the local labor market level.

Table 6: Effect of TV on Top Income Shares in Entertainment

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Share of Income</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top 0.1%</td>
<td>2.37</td>
<td>3.71</td>
<td>6.08</td>
</tr>
<tr>
<td></td>
<td>(1.27)</td>
<td>(1.69)</td>
<td>(2.12)</td>
</tr>
<tr>
<td>Local TV station</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time &amp; Labor Market FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>% growth</td>
<td>239%</td>
<td>96%</td>
<td>33%</td>
</tr>
<tr>
<td>% growth = top 1% growth (p-value)</td>
<td>0.0043</td>
<td>0.0000</td>
<td></td>
</tr>
</tbody>
</table>

[Note] Top p% is the share of income going to the top p percent of entertainers in a given local labor market-year, in percentage points. The shares are based on Pareto interpolation where necessary. A test of whether the growth rate of the top p% income share is the same as the growth rate of the top 1% income share is reported in the final row. The test is implemented in a regression with the ratio of top income shares as dependent variable. All regressions control for local labor market and year fixed effects. Standard errors are clustered at the local labor market level.
Table 7: Effect of TV on Mobility Between Labor Markets

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Share Entertainers who Migrated</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local TV station</td>
<td>-0.003</td>
<td>-0.003</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(0.0040)</td>
<td>(0.0044)</td>
<td>(0.0062)</td>
</tr>
<tr>
<td><strong>Panel B: Share Entertainer in US Top 1% (excl. neighbor)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local TV station</td>
<td>4.30</td>
<td>4.46</td>
<td>6.16</td>
</tr>
<tr>
<td></td>
<td>(1.31)</td>
<td>(1.30)</td>
<td>(2.27)</td>
</tr>
<tr>
<td>Demographics</td>
<td>–</td>
<td>Yes</td>
<td>–</td>
</tr>
<tr>
<td>Local labor market trends</td>
<td>–</td>
<td>–</td>
<td>Yes</td>
</tr>
</tbody>
</table>

[Note] In Panel A the outcome variable is the fraction of entertainers who moved. Panel B estimates the effect of TV stations on the share of Entertainers in the US top 1% of the wage distribution, excluding labor markets that neighbor treated labor markets. Specification details are as in Table 1. Panel A is based on 5,812 observations, Panel B on 10,792 observations.
Table 8: Elasticity of Entertainer Top Pay to Market Size

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\ln(\text{wage}^{99\text{th}}))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\ln(\text{Audience size}))</td>
<td>0.234</td>
<td>0.023</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.036)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\ln(\text{Value of market ($)})</td>
<td>0.220</td>
<td>0.194</td>
<td>0.198</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.022)</td>
<td>(0.036)</td>
</tr>
<tr>
<td>First-stage F-statistic</td>
<td>33.3</td>
<td>25.7</td>
<td>20.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel A: OLS - Cross-section 1939</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel B: IV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\ln(\text{Audience size}))</td>
<td>0.166</td>
<td>0.150</td>
<td>0.149</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.019)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>First-stage F-statistic</td>
<td>33.3</td>
<td>25.7</td>
<td>20.0</td>
</tr>
<tr>
<td>Panel C: IV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\ln(\text{Value of market ($)})</td>
<td>0.220</td>
<td>0.194</td>
<td>0.198</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.022)</td>
<td>(0.036)</td>
</tr>
<tr>
<td>First-stage F-statistic</td>
<td>57.10</td>
<td>38.1</td>
<td>28.7</td>
</tr>
<tr>
<td>Time &amp; Labor Market FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Demographics</td>
<td>–</td>
<td>Yes</td>
<td>–</td>
</tr>
<tr>
<td>Local labor market trends</td>
<td>–</td>
<td>–</td>
<td>Yes</td>
</tr>
</tbody>
</table>

[Note] The outcome variable is the logarithm of entertainer wage at the 99th percentile in a local labor market. Panel A uses 573 observations, while all other panels use 2,148 observations, missing demographics reduce the sample to 2,017 observations in column 2. The instrument is the number of local TV stations during local TV filming. Audience size is the capacity of the largest local theatre, or the TV audience of a local station and after 1956 the US TV audience. All regressions control for local production costs interacted with a dummy for the period after 1956. The first-stage F-statistic is the Kleibergen-Paap F-statistic that allows for non-iid standard errors. Standard errors are clustered at the local labor market level.
Table 9: Effect Heterogeneity by Market Structure

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Share Entertainer in US Top 1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local TV station (dummy)</td>
<td>5.90</td>
<td>0.753</td>
<td>-0.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.06)</td>
<td>(1.91)</td>
<td>(0.36)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple local TV station (dummy)</td>
<td>9.07</td>
<td>10.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.99)</td>
<td>(4.70)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frozen competitor</td>
<td>1.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local TV station</td>
<td>4.25</td>
<td>4.25</td>
<td>4.83</td>
<td>4.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.25)</td>
<td>(2.25)</td>
<td>(4.56)</td>
<td>(1.77)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local TV station × theatre count</td>
<td>-0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local TV station × population density</td>
<td>-0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local TV station × % with high-school degree</td>
<td>-1.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(12.23)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local TV station × high tax state</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.65)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Note] Sources and specification as in baseline. High-tax states are defined as states where the marginal tax rates of the top income bracket exceed the median; data availability restricts observations to 12,977 in this column. Theatre count are the number of employers listed in the Cahn-Gus Hills theatrical guide.
C APPENDIX: Extensions

C.1 A General Superstar Model

The aim of this section is to derive a unifying superstar model (SM) that allows to nest the different existing versions of superstar models. It will allow me to characterize the common assumptions and illustrates how recent advances in the SM literature link together. The backbone of the superstar model (SM) is an assignment model where workers are assigned to tasks (or markets). The model features heterogeneity on both sides of the market, workers have different levels of ability \((a)\) and tasks \((t)\) vary in complexity.

Labor Supply

Labor supply is different for each level of ability. For simplicity assume labor supply is inelastic. In that case the labor supply of each ability is the number of people with this level of ability:

**Assumption L1:** Worker ability is distributed continuously on a unit interval with PDF \(a \sim f_a(a)\)

Talent has no natural unit, assuming that \(a\) is distributed on a unit interval is therefore not a very restrictive assumption. If the value of \(a\) has no cardinal meaning, we could re-scale any measure of \(a\) to fit the unit interval.

Production

Worker and tasks differ in their productivity. The time of performing task \(t\) for worker \(a\) is given by \(C(a, t)\) and the corresponding hourly productivity is therefore \(1/C(a, t)\). Productivity only depends on the own ability and is independent of the supply of other worker types. In other words workers are perfect substitutes in the production of tasks. Assume that productivity of a worker has the following properties:

**Assumption P1:** absolute advantage \(C_a < 0\)

**Assumption P2:** comparative advantage \(C_t a < 0\)

Alternative models require a different version of comparative advantage, log-submodularity. This assumption does not imply nor is it implied by the previous comparative advantage assumption:

**Assumption P2.a:** comparative advantage II: \(\partial \ln C_a(a, t) / \partial t < \partial \ln C(a, t) / \partial t\)

Labor Demand

Demand for the different types of labor is determined by the distribution of tasks. To derive demand for tasks, we need to model the use of tasks. Assume tasks contribute to an aggregate production function \(Y(f_t(i), f_t(j), ...)^{62}\). Moreover we assume that there is a continuum of tasks \(i \in [0, 1]\) and \(Y\) has well defined derivatives for all \(i\).

**Assumption D1:** firms maximize profits by choosing tasks:

\[
\max_{f_t(i) \forall i} Y(f_t(i), f_t(j), ...) - \int_0^1 P(t)f_t(t)dt
\]

and the market for output clears.

---

62 An equivalent approach thinks of tasks as contributing to a utility function.
\[
\int_0^1 P(t)f_t(t)dt = K
\]

the first order conditions of profit maximization pin down demand for each task \((f_t)\). Instead of using such endogenous task demand, many studies take a short-cut and treat the task demand as exogenously fixed. This alternative assumption on task demand is:

Assumption D1.a: Tasks are distributed continuously on a unit interval with PDF \(t \sim f_t(t)\)

In studies of cross-sectional inequality the two assumptions lead to very similar results. The distinction becomes relevant when introducing technical change. With endogenous task demand we can study how technical change shifts the demand for tasks, while treating task demand as exogenous means that we have to assume how technical change affects task demand. This is effectively a reduced form approach to skill biased labor demand shocks.

**Equilibrium in the Labor Market**

The equilibrium in the labor market assigns workers to tasks. Given assumption P1, P2.a, L1 and D1 and competitive labor markets we can derive the equilibrium.

**Proposition 1:** The SM equilibrium is characterized by two conditions:

1. the assignment meets positive assortative matching (PAM):
   \[t'(a) > 0\] (15)

2. wages guarantee IC by:
   \[\frac{\partial \ln(w(a,t))}{\partial a} = -\frac{\partial \ln(C(a,t))}{\partial a}\] (16)

The first equilibrium result is that more able workers perform harder tasks.\(^{63}\) For the proof of PAM: see Sattinger notes (uses L1, P2). The intuition is that because markets are perfectly competitive and output of workers is perfectly substitutable within each task, the allocation of talent is completely determined by comparative advantage. Since better workers have a growing edge the harder a task gets, equilibrium assignment perfectly sorts worker ability and task difficulty. The second equilibrium condition ensures that the assignment is incentive compatible. The equality states that wages grow in line with productivity, which is a classic result of assignment models. Workers are paid their marginal product because there is perfect competition in the production of tasks.

**Sattinger (1979)**

D1: assume task demand is exogenous (D1.a) and follows a Pareto distribution with CDF: \(t = Bp^{-\beta}\)

L1: similarly talent follows a Pareto distribution: \(a = Ap^{-\alpha}\)

P1: production of workers is assumed to follow: \(\frac{1}{c(a,t)} = at^\gamma\)

\(^{63}\)Costinot & Vogel (2010) derive further structure of the assignment function from the market clearing conditions. They require 1-to-1 matching, see proofs.
**Sattenger (1975)**

D1: assume task demand is exogenous (D1.a) and is distributed on the unit interval following a continuous distribution with CDF: $t \sim F(t)$

**Terviö (2008)**

Terviö’s model is very similar to Sattenger (1975), a distinction is that Terviö allows for discontinuous CDFs for both workers and tasks. This implies, that his model allows for the case where either firms or workers have market power.

D1: assume task demand is exogenous (D1.a) and follows some distribution with CDF and is distributed on the unit interval: $t \sim F(t)$

A further difference to Sattenger is that the paper derives all results in terms of percentiles in the distribution, instead of talent units. This change is mainly expositional and does not material affect the conclusions.

In the empirical application Terviö adds an assumption on worker productivity:

P1: the production function takes the form: $\frac{1}{C(a,t)} = at$

**Gabaix & Landier (2008)**

D1: they assume task demand is exogenous (D1.a) and follows a Pareto distribution with CDF: $t \sim Bp^{-\beta}$

L1: similarly they impose a functional form for talent. They assume that it follows the general class of “ordinary functions,” which in the tail meets: $a'(p) \sim Ap^{-\alpha-1}$

P1: they make a functional form for the production function: $\frac{1}{C(a,t)} = at^\gamma$

They impose 1-to-1 matching. Moreover, just as Terviö, they use the expositional change in variables and solve the model in terms of percentiles $p$.

**Teulings (1995)**

D1: Aggregate production is CES: $Y = \left[ \int_0^1 f_t(t)^{-(\eta-1)} dt \right]^{\frac{1}{\eta-1}}$ with elasticity of substitution $\frac{1}{\eta}$ and $\eta > 0$. This is a single industry economy and spending thus equals industry output $K = Y$. The demand for each task $t$ therefore becomes:

$$P(t) = Y^\eta f_t(t)^{-\eta}$$

(17)

Given this assumption we can derive a third equilibrium condition. The task market clears if the difficulty of tasks increases proportionally with the cost of performing tasks.

$$\frac{\partial \ln(C(a,t))}{\partial t} = -\eta \frac{\partial \ln(f_t(t))}{\partial t}$$

**Costinot & Vogel (2010)**

D1: Aggregate production is similar to Teulings and follows a CES: $Y = \left[ \int_0^1 f_t(t)^{-(\eta-1)} dt \right]^{\frac{1}{\eta-1}}$ and $K = Y$. Different from Teulings, they additionally allow that tasks are excluded from production. They prevent that the marginal product of a task goes to infinity if demand goes to zero by introducing a participation threshold $\gamma$. Only tasks above $\gamma$ are used.
Additionally they impose 1-to-1 matching. This allows them to arrive at a closed form solution for the slope of the matching function (see proof below)

\[ t'(a) = \frac{f(a)}{C(a,t)} \frac{P(t)^{\frac{1}{2}}}{Y} > 0 \]

**Rosen (1982)**

The Rosen model changes the terminology, but mostly this re-labels the same mathematical concepts. So far, high ability workers produced more output per hour. In Rosens model high ability workers create higher quality. This simply relabels \( \frac{1}{C(a,t)} = q(t) \), instead of denoting quantity, \( q(t) \) now denotes the quality of the service worker \( a \) produces in task \( t \). The price for a unit quality of task \( t \) is \( p(t) \).

D1: Rosen assumes all tasks as perfect substitute in aggregate production. Hence, each quality unite \( q(t) \) produces the same amount of output, independent of the task. Tasks only differ in how often they are used in production. As before demand for a task is denoted \( f_t(t) \). Total output of quality unity is therefore \( Y = \int q(t)f_t(t)dt \). The aggregate producer problem becomes:

\[ \max_{m1} \int [q(t) - p(t)] f_t(t)dt \]

\[ \int f_t(t)p(t)dt = K \]

We also re-label tasks, recall that the role of tasks is to change the difficulty of producing output. In the Rosen model producing \( q(t) \) becomes more costly with the size of the market. Hence, we can think of tasks in terms of market size. This is the innovation of this paper, it allows workers to serve markets of varying size, while still maintaining one-to-one matching.

Equilibrium

Given the perfect substitutability of \( q(t) \) in aggregate production, tasks are only in demand if \( p(t) = q(t) \). At this price, aggregate producers are indifferent about the mix of tasks used in aggregate production. While in the above versions of the model aggregate production led to a task demand function, here the task demand will come entirely from comparative advantage.

The optimal assignment of worker over tasks must meet IC. Workers maximize their income \( (w(a,t) = p(t)f_d(t) = q(t)f_d(t) = f_d(t)/C(a,t)) \) by choosing the optimal task \( t \) given their ability level \( a \). The FOC, which guarantees IC, is therefore:

\[ w_t(a,t) = \frac{1}{C(a,t)f_d'(t)} - \frac{C_t(a,t)}{C(a,t)^2} f_d(t) = 0 \]

this implied that \( f' > 0 \). Since the cardinal value of tasks has no meaning, we can use any monotone transformation of the task label. Since tasks here denote market size it is

---

64 Notice that we do not impose \( K = \int f_d(t)w(t) \), spending in the sector does not necessarily equal income. As a consequence productivity growth means that fewer workers are needed to produce the demand K. The intuition for this assumption is that there is an outside good and spending on can therefore differ from income in the sector. Formally, K then depends on relative prices of the outside good and the elasticity of substitution between the superstar good and the outside good. For simplicity I suppress that complication.
which uses the envelope theorem. To assess how the income distribution compares to the
talent and
\[ \frac{\partial f_d(t)}{\partial a} = t'(a) = -\frac{\partial^2 \ln C(a,t)}{\partial t \partial a} > 0 \]  
This result pins down the matching of workers to tasks. From comparative advantage
wages are more dispersed than talent and
\[ w_{a}(a,t) = -\frac{\partial \ln C(a,t(a))}{\partial a} t(a) > 0 \]  
Which uses the envelope theorem. To assess how the income distribution compares to the
ability distribution, take the derivative again: . After rewriting the second derivative is:
\[ w_{aa}(a,t) = -\frac{\partial^2 \ln C(a,t(a))}{\partial a \partial a} t(a) - t'(a) \left[ \frac{\partial \ln C(a,t(a))}{\partial a} + \frac{\partial^2 \ln C(a,t(a))}{\partial a \partial t} t(a) \right] \]
\[ = -t'(a)(\hat{C}_a + t(a) \frac{d}{dt} \hat{C}_a) \]
with \( \hat{x}_y = \frac{\partial \ln(x)}{\partial y} \). The second derivative is positive as long as \(-\hat{C}_a > t(a) \frac{d}{dt} \hat{C}_a \). This implies that the marginal product of a worker exceeds the marginal output gain from increasing market size \( t \) for that worker. Another way of thinking about the same restriction is that as long as diseconomies of scale are not too extreme, wages are more dispersed than talent. This result is the insight of Sattinger (1975) that wages are more dispersed than talent and skewed to the right. In the Rosen model this result is extended to a model with endogenous task demand.

Rosen illustrates two interesting comparative statics of this model. First, consider the
case where producing at a large scale becomes easier, falling diseconomies of scale imply
\[ \frac{\partial^2 \ln C(a,t)}{\partial a \partial a} \downarrow \]. This change makes wages more convex in talent \( (w_{aa}(a,t) \uparrow) \), top income inequality therefore increases. The intuition for this result is that workers who operate in the largest market benefit most when the cost of large markets is relaxed. Since we have PAM, it is the high ability workers who benefit most from falling diseconomies of scale.

A second comparative static is the case of growing comparative advantage \( (\frac{\partial^2 \ln C(a,t(a))}{\partial a \partial t} \uparrow) \). From the PAM condition 18 it follows that \( t'(a) \uparrow \) for all levels of \( a \). Hence all workers work in bigger markets, moreover markets for the best worker grow the most. The expansion of market size for all types of workers implies wages for some workers will fall. Call the new market size distribution \( \hat{t}(a) > t(a) \), the budget constraint becomes \( \int \hat{t}(a)p(t)dt = K \). Since \( \hat{t}(a) > t(a) \) for all \( a \), the budget constraint requires that \( p(t) \) declines. Recall that wages are \( w(a,t) = p(t)t \), a drop in \( p(t) \) may therefore offset the rise in \( t \) and thus lead to falling

\[ 65 \] To arrive at the final equality use \( \frac{d}{dt} \hat{C}_a = \frac{\partial^2 \ln C(a,t)}{\partial a \partial a} \frac{1}{t(a)} + \frac{\partial^2 \ln C(a,t)}{\partial a \partial t} \), which assumes \( t(a) \) is invertible, aka one-to-one matching.
wages, even though all workers became more productive. The model can deliver falling wages without technical regress.

C.1.1 Contradiction with Canonical Skill Biased Technical Change Models

The canonical model of technical change in the labor market is a CES model of heterogeneous workers. A key implication of a CES function is that aggregate output exhibits constant returns to scale (CRS). This is incompatible with Rosen model, which features both comparative advantage and perfect substitutability in aggregate production. Proof:

Assume both CRS and comparative advantage holds. Recall that CRS implies that if we increase all inputs by $\kappa$, total output grows by $\kappa/C$:

$$1/C(\kappa a, \kappa t) = \kappa/C(a, t)$$

Denote productivity $\pi$ by $\pi(a, t) = 1/C(a, t)$. With comparative advantage two workers $a > a'$ and tasks $t > t'$, productivity meets:

$$\pi(t, a) \pi(t', a') > \pi(t, a') \pi(t', a)$$

Let $a = \kappa a'$ and $t = \kappa t'$, using CRS the inequality becomes $\kappa \pi(t', a') > \pi(t, a') \pi(t', a)$, since $\pi_t < 0$ and $\pi_a > 0$ the inequality also holds if we lower a single input in $\pi$ on the LHS. Hence

$$\kappa \pi(t', a') > \pi(t, a') \pi(t', a) > \pi(t', a') \pi(t', a)$$

Now let $\kappa = \pi(t', a')$ and we found a contradiction. The core CRS assumptions of the canonical skill biased technical change model is therefore incompatible with Rosen’s superstar model.

C.2 Equilibrium of the Superstar Model

Worker talent and firm follow a continuous and differentiable distribution. This will guarantee competitive like behavior of the market, despite the fact that each worker and firm type is unique. Denote talent at percentile $p$ by $T_p = W^{-1}(p)$ and firm size at percentile $i$ by $S_i = F^{-1}(i)$. For simplicity I assume that only one worker can appear on each stage, each firm therefore hires at most one worker.\(^{66}\) This assumption implies one of the three key assumptions of superstar effects, the imperfect substitutability of workers with team of workers. The production function is given by $\tilde{Y}(S, T)$. We can use the CDF above to equivalently express output in terms of stage size $(S)$ and the talent rank of the worker $(p)$. $Y(S, p)$ has the following properties: $Y_S > 0$, $Y_p > 0$, $Y_{pp} \leq 0$ and $Y_{Sp} > 0$ where subscripts denote partial derivatives. The final assumption guarantees that talented workers have a comparative advantage in bigger markets. This assumption is another key assumption of the model, also known as single crossing assumption. It will generate positive assortative matching in equilibrium.

Each stage manager maximizes profits by hiring a worker with talent $T_p$, taking its own firm characteristic as given. It will be convenient to express the hiring decision as choosing a percentile $p$ from the talent distribution. The firm problem is therefore given by:

\(^{66}\)The model extends to a setting were many workers are matched to a given market (see Sattinger (1993); Eeckhout and Kircher (2018))
\[
\max_p Y_i(p) - w(p)
\]

where \(w(p)\) is the wage for a worker at percentile \(p\) of the talent distribution.

The equilibrium is characterized by a matching function that assigns workers to firms, a participation threshold that satisfies the participation constraint (PC) and a wage profile that guarantees the assignment is incentive compatible (IC). The equilibrium of a superstar model is given by

i) assortative matching: \(i = p\)

ii) a wage schedule: \(w'(p) = Y_p(S_i, p)\)

iii) a participation threshold: \(\bar{p}\)

Condition i) is a consequence of the single crossing condition \(Y_{Sp} > 0\). For a proof see for example Sattinger (1975). To derive condition ii) I start from the fact that the equilibrium is incentive compatible. Incentive compatibility guarantees that for each firm \(i\) the optimal worker \(p\) meets:

\[
Y_i(p) - w'(p) \geq Y_i(p') - w'(p') \quad \forall \ p' \epsilon [0, 1]
\]

The second set of constraints are participation constraints (PC). They guarantee that both firms and workers are staying in the industry. Denote the reservation wage of workers \(w^{res}\) and the reservation profits \(\pi^{res}\). We assume they are the same for all workers and firms. The PC condition is thus given by:

\[
Y_i(p) - w(p) \geq \pi^{res} \quad \forall \ p \epsilon [0, 1]
\]

\[
w(p) \geq w^{res} \quad \forall \ p \epsilon [\bar{p}, 1]
\]

The participation constraint binds with equality for the lowest talented market participant. Let’s define the lowest percentile of the talent distribution that participates in the market as \(\bar{p}: w(\bar{p}) = w^{res}\). Individuals with lower levels of skill will work in an outside market where pay is independent of talent and given by \(w^{res}\).

The number of IC constraints can be reduced substantially for these kind of incentive compatibility problems. If the IC holds for the adjacent \(p'\) all the other ICs will hold as well. We can therefore focus on the percentiles just above and below \(p\). The IC for the adjacent \(p' = p + \epsilon\) can be further simplified if \(Y\) is differentiable in \(p\). Divide equation 19 by \(\epsilon\) and let \(\epsilon \rightarrow 0\).

\[
\frac{w(p) - w(p + \epsilon)}{\epsilon} \leq \frac{Y(S_i, p) - Y(S_i, p + \epsilon)}{\epsilon}
\]

\[
w'(p) = Y_p(S_i, p)
\]

The IC condition can thus be written as a condition on the slope of the wage schedule.

Condition ii) pins down the wage distribution up to a constant. Wages are increasing for more talented workers and under mild assumptions returns are convex, that is small differences in talent translate into growing wage differences at the top.\(^{67}\) Consider worker at percentile \(p\), a firm will pay this worker \(w'(p)\) more than a slightly less talented worker. The wage distribution is convex:

\[
w''(p) = \frac{Y_{pp}}{Y(S_p)} + Y_{pp} > 0 \quad \text{as long as the return to talent is not diminishing sharply and talent at any percentile is sufficiently scarce} - Y_{pp} < \frac{Y^{res}}{Y(S_p)}
\]

\(^{67}\)
wage increase is exactly equal to the additional output that the worker produces over the next best worker. Note that a firm pays what the worker is worth to it, while the same worker would be worth less to a smaller firm (since $Y_p^S > 0$). Despite the worse outside option of the worker, the firm passes all the output gains to the worker. This may seem surprising but holds because we assume firm types are distributed continuously. The outside option of the worker is thus only infinitesimally worse and the worker receives the full productivity gains of moving to a bigger firm. Take the alternative case where the distribution of firm types has jumps, some theatre venues are thus discretely bigger than their competition. Here the firm would keep all of the productivity gains. The lack of competition among employers therefore dampens wages. Condition iii) pins down employment levels in the market. All workers with talent above $p$ are active. In the extreme case of a winner takes all market $\bar{p} \to 1$ and only the most talented worker prevails.

To study how wages respond to technical change, assume that output is given by $Y(S_i, p) = \pi \cdot S_i^\gamma \cdot p$, with $\pi$ the price for a unit of talent. From condition ii) the slope of the wage schedule becomes:

$$w'(p) = \pi \cdot T'(p) \cdot S_p^\gamma$$

(23)

Wages at percentile $p$ are found by integrating equation 23. Wages are pinned down up to a constant ($b$) that represents the outside option. For simplicity I will normalize $b = 0$. The equilibrium superstar wage schedule is thus given by:

$$w(p) = \int_{\bar{p}}^{p} \pi \cdot T'(j) \cdot S_j^\gamma dj$$

(24)

In equilibrium the supply and demand for talent $D(\pi)$ clear at the market clearing price for talent ($\pi$):\[68]

$$D(\pi) = \int_{\bar{p}}^{1} T(j) \cdot S_j^\gamma dj$$

(25)

The wage of a worker at percentile $p$ depends on all worker-stage pairs below her, but is not directly affected by anything that happens at higher percentiles. Market size changes at the bottom end of the distribution will thus affect everyone. The logic for this is that bigger venues will pay a higher price per talent unit to attract the best talent. Each venue cares only about distinguishing itself from the next worse employer and thus pays a mark-up on their wage. In this sense all employers look “downward” in the distribution. An increase in wages at the bottom thus has a domino effect and will push up all wages. An increase in the market size at the top however doesn’t directly affect percentiles below. There is an indirect impact, as greater abundance of talent will put pressure on the price of skill $\pi$ and thus push wages downward.

To analyze wage changes we need more information on the functional form of the firm size and talent distribution. Assuming a Pareto distribution allows for a tractable solution to the model. This is a strong assumption that is not required but helps the exposition of the model. The conclusions hold more broadly, Appendix C.4 discusses the assumption in more detail and derives a more general solution. The CDF of talent and stage size ($W(p)$)

---

\[68\] A downward sloping demand guarantees that an equilibrium exists. The RHS is increasing in $\pi$, since more workers enter the market when the returns are high. $\bar{p}$ therefore increases when $\pi$ falls. The assumption is that the supply response happens along the participation margin. The model has however been extended to include an hours response Scheuer and Werning (2017).
and $F(i)$ is given by $T_p = (1 - p)^{-\beta}$ and $S_i = (1 - i)^{-\alpha}$, where $\beta$ and $\alpha$ are the shape parameters of the Pareto distributions. A higher value of these parameters means there is greater dispersion. Substituting the distributional equations into equation 24, wages become:

$$w_p = \int_{\bar{p}}^{p} \pi \cdot \beta \cdot (1 - j)^{-\beta - 1} \cdot (1 - j)^{-\alpha \gamma} dj = \left[ \frac{\pi \beta}{\alpha \gamma + \beta} \cdot (1 - j)^{-\beta - \alpha \gamma} \right]_{\bar{p}}^{p}$$  (26)

For top incomes we can assume that the effect of $\bar{p}$ is small and at the top income is thus proportional to:

$$w_p = N \cdot (1 - p)^{-\alpha \gamma}$$  (27)

with $N = \frac{\pi \beta}{\alpha \gamma + \beta}$. Wages are Pareto distributed and more dispersed than talent by factor $\alpha \gamma$.

### C.3 Growing Dispersion in Size

Assume the dispersion in the size distribution takes the form $\alpha' = s \cdot \alpha$ with $s > 1$. By substituting $\alpha'$ into equation 27 we can see that the wage schedule becomes:

$$w_{p}^{t+1} = N^{t+1}(1 - p)^{-(s \cdot \alpha \gamma + \beta)} = w_{p}^{t} \psi(1 - p)^{(s - 1)\alpha \gamma}$$  (28)

with $\psi = \frac{\pi^{t+1}}{\pi^{t}} \frac{(\alpha \gamma + \beta)}{(s \cdot \alpha \gamma + \beta)}$, superscripts denote the time period. As benchmark consider the case were labor supply is perfectly inelastic, hence $\bar{p}$ is fixed. There is therefore no entry and exit and percentiles do not carry time superscripts. Allowing for entry has negligible effects on top wages and the results thus would carry through approximately. At the bottom of the distribution entry would matter. 69

Dividing both sides by $w_{p}^{t}$ we get wage growth ($g_{p}^{w}$) from a superstar effect:

$$g_{p}^{w} = \psi(1 - p)^{(s - 1)\alpha \gamma}$$  (29)

Notice that the growth rate depends on $p$. Wages grow more at the top of the distribution. The effect on the wage schedule is illustrated in figure D1. Two features stand out. For one, superstar effects pivot the wage schedule. Second the impact across the wage distribution is U-shaped. Gains for the superstars come at the expense of less talented workers, while wages at the bottom remain pinned down by the outside option. Two factors drive these effects. For one, top entertainers can use their talent more intensely through the new technology and more consumers get to see the star entertainer. This effect is captured by last term in equation 29. The biggest wage gains occur at the top of the distribution and the returns to being a superstar have risen.

There is a second effect that comes from the greater availability of stars. The wider availability of stars has reduced the price for a unit of talent ($\pi^{t} > \pi^{t+1}$). 70 If the audience of an entertainer is unchanged marginal revenue product therefore decreases. The wage at

---

69 Allowing for entry and exit would reduce wages by a constant for all participating workers. The drop in the talent price $\pi$ induces exit. Workers that exit earn the outside option (here 0). The first participating worker will be just indifferent and also earn zero. As above we assume that the changes at the bottom are immaterial for income at the top.

70 The intuition for this result is that the average market size increases with $\alpha$. Formally, we can look at the RHS of equation 25. It is now given by $\int_{p^{res}}^{1} \lambda p^{\beta - \alpha + 1} dp = \lambda [1 - (p^{res})^{\beta - \alpha + 1}] / (\beta - \alpha + 1)$ which is increasing in $\alpha$ (and $\lambda$ is a constant). The skill price $\pi$ has to fall to bring the market into equilibrium.
any given stage declines. This effect is captured by $\psi < 1$ and is a level shift downward in earnings for all entertainers. Taking these effects together the wage schedule has shifted downwards and pivoted upward.

C.4 Distributional Assumptions

The core result, that a unit of talent becomes more valuable as $S$ increases, holds independent of the distributional assumptions. As it becomes feasible to serve bigger markets, the wage-talent profile pivots and becomes steeper. For the general case we can show this by evaluating condition $ii)$ of proposition 2.1 at the equilibrium values and differentiate with respect to $S$:

$$w_{pS}(p^*) = Y_{pS}(p^*) + Y_{pp}(p^*) \frac{\partial p^*}{\partial S} = \frac{w''(p^*)}{\theta'(p^*)} > 0$$

The second equality uses positive assortative matching to invert the assignment function $p = \theta^{-1}(S)$ and differentiates to yield $\frac{\partial p^*}{\partial S} = \frac{1}{\theta'(p^*)}$. The effect of market size on the wage slope is positive. This follows from the convex wage schedule discussed above and the positive assortative matching of talent and market size. We don’t need to appeal to the envelope theorem here. The envelope theorem doesn’t apply in an assignment model. An employer who increases the market size is able to poach a better worker from a competitor and thus has first order effects on other market participants. Even without appealing to the envelope theorem we can sign the equation as long as the assignment function is invertible.

The closed form solution derived in the text also holds for a broader set of assumptions. The assumption that talent is Pareto distributed can be relaxed. Extreme value theory has shown that many functions look similar in the upper tail. Gabaix and Landier (2008) show that we can apply this to the assignment problem and solve for top incomes, given mild assumptions on the talent distribution. They noted that for “regular” continuous distributions the tail of the distribution can be approximated by: $T'(p) = -Bp^{\beta-1}$. This holds exactly, as for Pareto distributions, or up to a slowly varying function.

The assumption that firm size is Pareto has empirical foundations. A large literature has found that the Pareto distribution fits the data well. The remarkable fit of the distribution has become known as "Zipf’s law". For various settings this has been documented in Axtell (2001) and Fujiwara et al. (2004). An alternative approach to assuming a size distribution, is to treat size as an endogenous variable. This is the path pursued by “differential rent models.” Geerolf takes this approach and shows that under plausible assumptions the resulting market size follows a Pareto distribution (see Geerolf (2016)). By imposing the Pareto distribution this paper can be thought of as a reduced form version of such a model.

C.5 Skill Biased Technical Change and Pay Dispersion

The skill biased technical change model features two groups of workers, high (H) and low (L) skilled workers. To give the model the best possible shot at fitting the data assume that workers can have different amounts of H and L, call the quantity of skill $t$. Assume that $t$ is distributed with an invertible CDF $G_H(t)$ and $G_L(t)$ respectively. Within a skill group workers are perfect substitutes and the firm therefore cares only about the total units of H and L employed. Production is given by a CES function with $A_i$ the productivity of skill group $i$: 
\[ Y(H, L) = \left[ A_H \left( \sum t^H \right)^\theta + A_L \left( \sum t^L \right)^\theta \right]^{1/\theta} \]

Because workers are perfect substitutes the law of one price applies. There is a single market clearing price for a unit of low and high talent, call them \( w_H \) and \( w_L \). The price of high talent is given by:

\[ w_H = A_H \left[ \frac{\sum t^H Y}{Y} \right]^{\theta - 1} \]

And the wage of a high skilled individual with quantity of skill \( t^H \) is given by:

\[ w_{t^H} = w_H \cdot t^H \]

We can now show that this model can generate a convex wage distribution. Call the distribution of wages \( F^w(w) \). In the top tail this is given by:

\[ F^w(W) = Pr(w_{t^H} > W) = Pr(t^H > \frac{W}{w_H}) = G^H \left( \frac{W}{w_H} \right) \]

The top tail of the wage distribution follows the same distribution as \( t^H \). The wage at percentile \( p(W_p) \) is than given by:

\[ W_p = w_H G^{-1}_H(p) \]

With an appropriate assumption on \( G^H \) we can therefore match any wage distribution, including one that is convex. The result that a superstar model leads to a convex wage distributions is therefore not unique to superstar models.

We can however use the effect of technical change to distinguish the two models. Consider the a skill biased technical change. The standard assumption in this model is skill biased technical progress makes high skilled workers more productive (\( \tilde{A}_H > A_H \)). The wage per talent unit therefore becomes:

\[ \tilde{w}_H = \tilde{A}_H \left[ \frac{\sum \tilde{t}^H \tilde{Y}}{Y} \right]^{\theta - 1} > w_H \]

Next consider wages. The baseline case assumes that labor supply is inelastic, hence the talent distribution \( G^H(t) \) is unchanged. Allowing for a labor supply response complicates notation and generates little additional insight. The wages at \( p \) are given by:

\[ \tilde{W}_p = \tilde{w}_H G^{-1}_H(p) \]

We now can show that technical change leads to very limited change in the distribution of wages. The growth of wages is given by:

---

\(^{71}\)Here we assume that low skill workers do not features in the top tail of the wage distribution.

\(^{72}\)The higher wage induces entry of workers where \( \tilde{w}_t \) growths above the outside option b. These are workers with low levels of \( t \) and as a result the distribution of talent changes at the bottom end. For ordinary talent distributions this has little effect on the top tail of \( G^{-1}_H(p) \). The result that follow therefore carry through approximately at the the top of the distribution.
Wage growth is the same across all percentiles in the top tail. Technical change leads to a level shift in the wage schedule.

### C.6 Technical Change and Pay Dispersion in the Income Tail

This section derives proposition 2.4. To derive part i) of the proposition, take the wage at two percentiles, say the 90th and 99th. The ratio is given by:

$$ R_{99,90} = 0.01^{-(\alpha \gamma + \beta)} $$

Changes in the price for talent $\pi$ will not affect this ratio, while growth in $\alpha$ does. Taking derivatives we get $\frac{\partial R_{99,90}}{\partial \alpha} > 0$. The 99-90 percentile ratio thus grows with superstar effects.

Next consider the effect on top income shares. This metric has been used extensively to document growing inequality at the very top (Piketty and Saez (2003); Piketty (2014)). The top income share is defined as the sum of incomes of individuals above percentile $p$ divided by total income ($G$):

$$ s_p = \int_p^\infty \int_i^{\bar{i}_p} w_j djdi/G $$

With a skill biased demand shock the growth in the top income share is given by:

$$ g^{s_p} = \frac{s_p^{t+1}}{s_p^t} = \frac{G^{t+1}}{G^t + 1} \pi^t \frac{\int_p^\infty \int_i^{\bar{i}_p} T'(j) \cdot S_j^t djdi}{\pi^t \int_p^\infty \int_i^{\bar{i}_p} T'(j) \cdot S_j^t djdi} = \frac{g^\pi}{g^G} (1 + g^b) $$

The second step uses the definition of top income shares and equation 24. The final step collects terms and cancels. Top income shares grow as long as the price for talent grows faster than GDP. There is an additional effect from entry ($g^b$) which is likely negligible, it affects top income shares positively because the wage setting process is downward looking. Strikingly, the growth rate of the top income share at $p$ is independent of $p$. All top income shares are growing at the same rate. The top 1% is therefore contributing a constant fraction to the income share of the top 10%. The ratio of the income share that goes to the top 1% and 10% is therefore unaffected by SBD shocks.

This contrasts with the impact of superstar effects. Note that the top income share above percentile $p$ for a Pareto distributed variable is given by $s_p = (1 - p)^{1-\lambda}$, with $\lambda^{-1}$ the shape parameter of the distribution. Here wages follow a Pareto distribution with shape parameter $\lambda = \alpha \gamma + \beta$. The growth in the top income share from superstar effects is therefore given by:

$$ g^{s_p} = \frac{s_p^{t+1}}{s_p^t} = \frac{(1 - p)^{1-s\alpha \gamma - \beta}}{(1 - p)^{1-\alpha \gamma - \beta}} = (1 - p)^{-(s-1)\alpha \gamma} $$

---

73 This result has been used extensively to calculate top income and wealth shares. Even for variables that do not follow a Pareto distribution, there is still a lambda now varying with $p$. Many income variables are approximately Pareto and lambda is only slowly varying and the result holds approximately.
The second equality uses the property of a Pareto variable, while the final equality cancels terms. Top incomes shares are again growing. But the pattern is different from SBD shocks. Here the growth rate is increasing in \( p \). This implies that the income share of the top 0.1% grows faster than the share that goes to the top 1%, which in turn grows faster than the share of the top 10%. A growing fraction of the top 10% income share is taken home by the top 1%.

C.7 Proof: No Cannibalisation in SBTC Models

This section proves that technical progress rules out falling wages in the SBTC model. I study a flexible SBTC model with arbitrary many skill groups \( 1 \ldots n \). The production function is given by:

\[
F(\alpha_1(\theta)L_1, \alpha_2(\theta)L_2, ..., \alpha_n(\theta)L_n)
\]

Where \( L_i \) is type of labor \( i \) and \( \alpha_i \) the associated productivity and \( \theta \) is the driver of technical change. We allow for exit and therefore impose that no worker type is indispensable in production:

\[
\frac{\partial F}{\partial L_i} < \infty \quad \forall L_i
\]

Technical change may affect different parts of the distribution differently, in particular we allow for extreme bias technical change that predominantly helps star workers. We do not ex-ante rule out that changes in technology reduces productivity for some types of workers. However, we impose that the overall effect of technology is positive, hence we assume there is no technical regress in production:

\[
\frac{\partial F}{\partial \theta} = \sum L_i \frac{\partial \alpha_i}{\partial \theta} \frac{\partial F}{\partial L_i} > 0
\]

We want to show that this implies that:

\[
\frac{\partial \alpha_i}{\partial \theta} \geq 0 \quad \forall i
\]

We proceed by contradiction and assume this was not the case, hence \( \frac{\partial \alpha_i}{\partial \theta} < 0 \) for some \( i \). To see that this violates restriction 31, assume that all \( L_j = 0 \) for all \( j \neq i \) and \( L_i > 0 \) for \( i \). This implies \( \frac{\partial F}{\partial \theta} < 0 \), violating the assumption that technical progress cannot lead to falling productivity.

C.8 Superstar Effects and the Link of Talent and Pay

An important feature of superstar models is that pay is more dispersed than talent. The difference in pay across two individuals thus exceeds the gap in the marginal product. To see this define the ratio of the marginal products as \( \omega = \frac{w'(\hat{n})}{w'(n)} \). This ratio captures the difference in the marginal value of talent at two points of the distribution, I call it the talent premium. It might look similar to the skill premium in a SBD model. Note that the talent premium has an important difference to the skill premium. The skill premium compares the wage of two workers, while the talent premium compares the derivative of wages of two workers. It turns out that analyzing the talent premium can be misleading if one is
interested in wage dispersion. As we saw in equation 24, wages are downward looking and thus depend on all percentiles below. Changes at lower percentiles will affect pay at higher levels of the distribution. As a result the talent premium doesn’t capture the full extent of wage inequality.

Consider for instance the case were all markets double in size. The talent premium would be unaffected, wage inequality would however rise. Workers higher up in the talent distribution benefit twice, once from their own rise in marginal product and once from the greater renumeration at the percentiles below. The first effect is the same for everyone, while the second accumulates as we move up the distribution. Growing market size thus generates faster income rises at the top. To the contrary, the talent premium remains unchanged as it is based on a marginal change in talent and thus cancels out the accumulated effect of the downward looking wage distribution. For a local change in market size it does however give the right result. We can see that if the top market grows, top wages would go up by $\gamma$.

In a SBTC model pay within a skill group is always proportional to the amount of skill a worker possess. Here we compare the link between reward for talent and dispersion in pay in a superstar model. Let’s define the wage premium as the ratio of wages at percentile $\tilde{n}$ and $n$. This measures the difference in pay at two percentiles of the distribution. Using 27 together with the distributional assumptions above, we get:

$$ln\left(\frac{w(\tilde{n})}{w(n)}\right) = ln\left(\frac{T'(\tilde{n})}{T'(n)}\right) + \gamma \times ln\left(\frac{S(\tilde{n})}{S(n)}\right) + ln\left(\frac{\tilde{n}}{n}\right) = ln(\omega) + ln\left(\frac{\tilde{n}}{n}\right)$$

(32)

The wage premium is thus closely linked to the talent premium. The wage premium exceeds the talent premium due to the final term. This term captures that wages are downward looking. In other words it captures that differences in wages at $\tilde{n}$ and $n$ depend on the infra-marginal wages between them. We can simplify the above result further by noting that in a Pareto distribution percentiles are proportional, with the Pareto parameter the the factor of proportionality. We thus get:

$$ln\left(\frac{\tilde{n}}{n}\right) = -\frac{1}{\epsilon} ln\left(\frac{w(\tilde{n})}{w(n)}\right)$$

This result follows from the fact that wages are Pareto distributed, as we can see in 32. The Pareto parameter is $\epsilon = \alpha\gamma - \beta$. We can substitute this result back into equation 32 and get:

$$ln\left(\frac{w(\tilde{n})}{w(n)}\right) = ln(\omega)\left(\frac{\epsilon}{\epsilon + 1}\right)$$

(33)

The wage premium thus exceeds the talent premium by a constant factor. In a superstar model wages amplify differences in returns to talent. The model thus breaks the proportionality of talent and pay. The mark up is particularly stark if $\epsilon$ is large. That is the case if wages are very dispersed. In terms of model primitives this corresponds to the case where the relative scarcity of talent and market size ($\alpha/\beta$) is large and the returns to market size don’t diminish quickly ($\gamma$).
C.9 Superstar Parameters and Employment Elasticities

Structural parameters from the superstar model can be identified off the relation of top pay to market size. Elasticities of employment at top wages to market size are therefore not immediately comparable. However, these two estimates are linked. This section shows how the employment elasticities can be used to identify structural superstar parameters. To do so I will establish the link of employment elasticities to wage elasticities. The link is very simple if the wage distribution is Pareto and the superstar effect is order preserving. This is a useful benchmark and we can relax those assumption somewhat below.

First note that a top earner is defined as an earner above a threshold:

$$TE^0 = (1 - F(\bar{w})) = G(\bar{w})$$

Where we define G(x) as the share of individuals above x. Consider a small increase in the number of top earners. If the order of individuals in the distribution has remained the same we can re-write the expression.\(^{74}\) The top earners in period 1 are the top earner of period 0 plus individuals that were previously just below the top earner threshold. Let’s denote the lowest period 0 wage of a period 1 top earner by \(\tilde{w}\). The number of new top earners thus becomes:

$$TE' \approx TE^0 + g(\bar{w})(\tilde{w} - \bar{w})$$

It follows:

$$\Delta TE \approx f(\bar{w})(\bar{w} - \tilde{w})$$

where the last equality holds for small changes in w.

If the shape of the CDF is known this equation allows to translate a change in employment to an associated shift in wages. Assuming a Pareto distribution will again proof useful. A convenient property of the Pareto distribution is that the tail of the distribution has a well defined shape with \(f(x) = \alpha/x\). Using this fact we can re-write the above equation in terms of elasticities \(\varepsilon_{i,j}\) with \(\alpha\) the Pareto coefficient:

$$\frac{\Delta TE}{TE} \approx \frac{f(\bar{w})}{G(\bar{w})} \Delta w$$

$$\varepsilon_{TE,m} = \alpha \varepsilon_{w,m}$$

This gives us a simple expression for the link between the elasticity of number of top earners and the elasticity of pay. We can apply this expression to the results in this study and get an alternative estimate for the elasticity of top pay to market size. I estimate the \(\alpha\) parameter on the pre-TV wage distribution using the full count, non top-coded Census of 1940. I experiment with a number of estimation strategies with similar results.\(^{75}\) Independent of the approach the estimated Pareto coefficient is close to but bigger than 3. To err on the conservative side, I will use a value of 3 for the analysis.

---

\(^{74}\)Rosens’ model of superstar effect is an example where this assumption holds. This is however a strong assumption that can be relaxed with additional data on the position of TV-stars in the pre-TV wage distribution. This is to follow.

\(^{75}\)The baseline results use Kuznets’ approach to estimate the Pareto parameter. This approach uses the fact that average income above a threshold is proportional to the threshold. With \(\alpha\) the coefficient of proportionality. I also run the Atkinson & Piketty approach and use different threshold values. All estimates are above 3, with most between 3.02 and 3.16.
Using the results above the estimates to compute the elasticity of top pay employment to market size we find $\varepsilon_{TE,m} = 0.45$. Using the relation derived here we can translate this into an elasticity of income. The implied elasticity of top wages to market size is $\varepsilon_{w,m} \approx 0.15$. A doubling in market size will thus raise top wages around 15%. This is remarkably close to the wage elasticity that we estimated directly from the data (recall it was 0.12).

The wage elasticity would be bigger if we relaxed the assumption that the effects are order preserving. Without such homogeneous treatment effects individuals from further down in the wage distribution could become earning superstars. This would require a large wage rise for these people and thus potentially increase the estimated elasticity. To assess how much this matters in practice, we need to know where in the income distribution local TV stars came from. Figure 7 plots this. The figure matches local TV stars to their pre-TV earnings in the 1939 Census. It plots the position of these stars in the pre-TV income distribution after correcting the wage for age, gender and education effects. The figure makes clear that most of the star entertainers were earning high incomes in the pre TV period. The order preserving assumption thus looks like a reasonable approximation. Allowing for heterogeneous treatment is unlikely to change the conclusions substantially.

D APPENDIX: Empirics

D.1 Robustness checks

D.1.1 Top Income Metrics

The baseline outcome variable normalizes the number of top earners by aggregate employment in entertainment. This has the convenient effect that the result is a percentage change. The numerator doesn’t vary at the local labor market level, changes in this variable should therefore be captured by the year fixed effect. We may however worry that since the variable enters multiplicatively, the additive year fixed effect doesn’t completely control for changes in the denominator. In column 2 Table D3 I therefore re-run the baseline regression using the count of top earners as outcome. In an average labor market 18 individuals are in the top percentile. TV more than doubles the number of top earners. Column 1 repeats the baseline regression. The normalization changes the units of the results, but the basic conclusion remains unchanged. This confirms that the normalization has no substantive effect on the result.

Figure D6 illustrated the evolution of various alternative top income measures. The figure shows the the 99th percentile of the Census wage distribution over time. This is the threshold that defines top earners in the baseline estimates. The figure contrasts this threshold with alternative top income thresholds. These include the thresholds calculated by (Piketty and Saez, 2003) and the 95th percentile of the wage distribution and the 95th percentile of the entertainer wage distribution. All of these are below the wage top-code applied in the data. The series move similarly. In practice it will therefore matter little how a top earner is defined. Table D3 confirms this formally. It repeats the previous analysis using other top income measures. Column 1 repeats the baseline estimate. Column 3 uses

---

76 This is based on panel B of table 1 and table 2.

77 The information on TV stars comes from the 1949 “TV and Radio Annual.” The magazine publishes a who is who of the industry. Biographical information is used to link those individuals to their 1940 Census records. The information is thus based on the subset of TV stars who are listed in the who is who and can be found in the 1940 Census.
the top income percentile as defined by (Piketty and Saez, 2003). With this definition of top earners slightly more entertainers are top earners. The effect of TV remains however unchanged. The the number of people in the top percentile about doubles.

Column 4 and 5 look at the wage distribution among entertainers. By definition 1% of entertainers will earn wages above the 99th percentile of the entertainer wage distribution. Mechanically the share of top earners thus can’t change. Instead the analyses looks at where these individuals live. If TV had a positive effect on top incomes, the number of top earning entertainers increases in areas where TV productions are filmed and declines elsewhere. With the Census data it is not possible to analyze the 99th percentile of the entertainer wage distribution. This value is above the top code in some years. While we saw that the 99th percentile of the overall wage distribution stays below the top code, the same doesn’t hold true in entertainment wage distribution because entertainer wages are more skewed than overall wages. The analysis therefore looks at entertainers above the 95th percentile of the entertainer wage distribution. Analyzing within entertainers wage dispersion has the appealing advantage that it is a measure of inequality in the affected sector. This measure is however problematic if TV induces substantial exit in the entertainment sector. Exits would shift the 95th percentile even in the absence of any effect of television on top earners. If television results in an exit of the bottom 10% of entertainers, the 95th wage percentile would rise. If there was no further effect on top earners, we would find that fewer entertainers are top earners after the introduction of television. Hence, this measure will lead to a downward biased in the estimate of TV. Indeed in column 3 the number of top earners increases by less. The increase here is 20% over the baseline. To address the endogeneity issue column 4 keeps the 95th percentile fixed at the 1940 level. This measure is thus unaffected by exit of entertainers. This estimate is indeed substantially bigger than column 3. These results confirm that television led to a substantial increase in top earnings in entertainment.

D.1.2 Pre-Trend

A challenge for estimating pre-trends with this sample is that wage data in the Census is first collected in 1939. Since the Census is decennial this only allows for a single pre-treatment period. To estimate pre-trends I therefore combine the Census data with data from Internal Revenue Services (IRS) tax return data. In 1916 the IRS published aggregate information on top earners by occupation-state bins. Data for actors and athletes are reported. I link the Census data with the tax data and run the regressions at the state level. Table D5 reports the results. Column 1 repeats the baseline estimate with data aggregated at the state level. Despite the aggregation at the state level the effect remains highly significant. Column 2 adds the additional 1916 data from the IRS. The results stay unchanged. Column 3 shows the differences in top earners in treatment and control group for the various years. It shows a marked jump up in top earners in the treated group in the year of local TV production. The coefficient on the pre-trend is not significant because the standard errors are large. If anything the pre-period saw to a decrease relative decrease in top earners in the treatment areas. Even if taken at face value the pre-trends thus can’t explain the identified positive effect of TV.
D.1.3 Placebo Occupations

Television only changed the production function of a handful of occupations, we can therefore use alternative occupations as placebo group. The ideal placebo group will pick up changes in top income in the local economy. The main high pay occupations are therefore used as placebo group, these professions are medics, engineers, managers and service professionals. If TV assignment is indeed orthogonal to local labor market conditions, we would expect that such placebo occupations are unaffected. Results for the placebo group are reported in A. TV does not show up in top pay of the placebo occupations. The only occupation group with a significant positive effect are performance entertainers. Column 1 shows that the placebo group doesn’t experience any growth in top incomes. Moreover, the estimated effect on performance entertainers remains similar to the baseline in Table 1. Column 2 allows for separate impact of television across the different placebo occupations. Only performance entertainers experience the significant and large top earner rise.

With the inclusion of the control groups, I can run a full triple difference regression. In this specification there are treated and untreated workers within each labor market. We already controlled for location specific trends before, this specification will go further and allow for a non-parametric location specific time fixed effect. An example where this might be necessary is if improved local credit conditions result in greater demand for premium entertainment and simultaneously lead to the launch of a new TV channel. This may lead to an upward bias in the estimates. My treatment now varies at the time, labor market and occupation level. This allows me to control for pairwise interactions of time, market and occupation fixed effects. These will address the outlined credit access problem as the fixed effects will now absorb location specific time effects.

Column 3 shows the results. The effect on performance entertainers remains close to the baseline estimate. The additional location specific time and occupation fixed effects therefore don’t seem to change the findings. This rules out a large number of potential confounder. The introduction of a "superstar technology" thus has a large causal effect on top incomes and this effect is unique to the treated group.

D.1.4 Quantile Regressions

A further method of testing the effect of TV across the distribution is through quantile regressions. A number of recent papers have extended the use of conditional quantile regressions to panel settings. In the linear regression framework additive fixed effects lead to a "within" transformation of the data. In the non-linear quantile framework additive linear fixed effects will not result in the standard "within" interpretation of the estimates. Adding fixed effects may therefore not be sufficient for identification. Chetverikov et al. (2016) develop an quantile estimator that handles group level unobserved effects if treatment varies at the group level. Similarly, Powell (2016) develops a panel quantile estimator that mimics the "within" transformation of fixed effects for the quantile regression.

A shortcoming of the quantile regression is that the estimates are sensitive to entry and exit. The magnitude of the quantile effect is therefore hard to interpret. However, the relative magnitude across percentiles is still informative and the test relies exclusively on such relative patterns. Recall that SBD predicts a homogeneous growth rate, while the superstar model predicts larger wage growth rates at the top. To test whether either model matches the data, I run quantile regressions at various percentiles. I restrict myself to quantiles for the median and above since the results were derived by using an approximation for the top
of the distribution. I follow the procedure in Chetverikov et al. (2016) to implement the difference in difference for quantile regressions. The estimated coefficients are plotted in figure D8, alongside the prediction of the SBD model. The effect is biggest at the top of the distribution and effects are notably smaller at the lower percentiles. This result is in line with the superstar model but contradicts a model of SBD. Table D7 reports the panel quantile estimates using the Powell (2016) approach.

D.2 Data construction

D.2.1 Local labor markets

- The analysis defines local labor markets as commuting zones (CZ). A labor market is an urban center and the surrounding commuters belt. The CZs fully cover the mainland US. The regions are delineated by minimizing flows across boundaries and maximizing flows within labor markets.

- David Dorn provides crosswalks of Census geographic identifiers to commuting zones (Autor and Dorn, 2013). I use these crosswalks for the 1950 and 1970 data.

- No crosswalk is available for the 1960 geographic Census identifier in the 5% sample and the 1940 Census data. Recent data restoration allows for more detailed location identification than was previously possible (mini-PUMAs).

- To crosswalk the 1940 data, I use maps that define boundaries of the identified areas. In GIS software I compute the overlap of 1940 counties and 1990 CZ. In most cases counties fall into a single CZ. A handful of counties are split between CZ. For cases where more than 3 percent of the area falls into another CZ, I construct a weight that assigns an observation to both commuting zones. The two observations are given weights so that they together count as a single observation. The weight is the share of the county’s area falling into the CZ. The same procedure is followed for 1960 mini PUMAs.

- Carson city county (ICSPR 650510) poses a problem. This county only emerges as a merger of Ormsby and Carson City in 1969, but observations in IPUMS are already assigned to this county in 1940. I assign them to Ormsby county (650250).

- CZ 28602 has no employed individual in the complete count data in 1940.

D.2.2 Worker data

The data uses the public use micro data of the US decennial census from 1940-1970 (excluding Hawaii and Alaska). Most variables remain unchanged throughout the sample period. IPUMS has taken great care to provide consistent measures of variables that did change.

- There are 722 commuting zones (CZ) covering the mainland USA. These regions are consistently defined over time.

- There are 28 relevant occupations. 1950 occupation codes are
  - Treatment group: 1, 5, 31, 51, 57
– Placebo group: 0, 32, 41, 42, 43, 44, 45, 46, 47, 48, 49, 55, 73, 75, 82, 200, 201, 204, 205, 230, 280, 290, 480

- controls are population aggregates in the area: share high skilled (high school and above for people over 25), share non white, median age, sample size per CZ, median wage and age
- Aggregates are calculated using the provided sample weights
- variables used incwage, occ1950 (in combination with empstat), wkswork2, hrswork2

D.2.3 Employment

- Occupation based on the 1950 classification of IPUMS (Occ1950). This data is available for years 1940-1970. For previous years the data is constructed using IPUMS methodology from the original occupation classification.

- Occupational definitions change over time. IPUMS provides a detailed methodology to achieve close matches across various vintages of the US census. Luckily the occupations used in this analysis are little affected by changes over time. More details on the changes and how they have been dealt with are: The pre 1950 samples use an occupation system that IPUMS judges to be almost equivalent. For those samples IPUMS states: "the 1940 was very similar to 1950, incorporating these two years into OCC1950 required very little judgment on our part. With the exception of a small number of cases in the 1910 data, the pre1940 samples already contained OCC1950, as described above." For the majority of years no adjustment all is therefore necessary. Changes for the 1950-1960 period - Actors (1950 employment count in terms of 1950 code: 14,921 and in terms of 1960 code: 14,721), other entertainment professions are unaffected. Changes from 1960-1970: Pre 1970 teachers in music and dancing were paired with musicians and dancers. In 1970 teachers become a separate category. My analysis excludes teachers and thus is unaffected by this change. Athletes disappear in 1970 coding. The analysis therefore only uses the athlete occupation until 1960. The only change that has a major effect on worker counts is for "Entertainers nec". In 1970 ca. 9,000 workers that were previously categorized as "professional technical and kindred workers" are added and a few workers from other categories. The added workers account for ca. 40 percent of the new occupation group. The occupation specific year effect ought to absorb this change. I have also performed the analysis excluding 1970 and find similar results. Moreover I find the TV effects for each occupation individually. The classification changes therefore seem to have little effect on the results.

- The industry classification also changes over time. I use the industry variable to eliminate teachers from the occupations "Musicians and music teacher" and "Dancers and dance teachers." The census documentation does not note any change to the definition of education services over the sample period, however the scope of the variable fluctuates substantially over time. From 1930 to 1940 the employment falls from around 70,000 to 20,000, from 1950 to 1960 it increases to around 200,000 and falls back to around 90,000 from 1960 to 1970.

- The definition of employment changes after the 1930 Census. Before the change, the data doesn’t distinguish between employment and unemployment. In the baseline
analysis I therefore focus on the period from 1940 onwards. For this period the change
doesn’t pose a problem. An alternative approach is to build a harmonized variable for
a longer period, this includes the unemployed in the employment count for all years.
I build this alternative variable and perform robustness checks with it. The results
remain similar. For two reasons the impact of this change on the results is smaller
than one might first think. First, most unemployed don’t report an occupation and
don’t fall into the sample of interest. Second, the rate of unemployed is modest
compared to employment and thus including them doesn’t dramatically change the
numbers.

- The control group are workers in top earning professions outside entertainment (lawyer,
  medics, engineers, managers, financial service). The relevant occupations are available
  across most years. Exceptions are 1940 where a few occupations in engineering,
  medicine and interactive leisure are grouped together and in 1970 where the floor men
  category is discontinued. I control for those changes with year-occupation fixed effects
  in the regressions. The effects occur within occupations rather than between them,
  results for all occupations separately are available upon request.

- Number of workers are based on labforce and empstat. Both variables are consistently
  available for 16+ year olds. Hence the sample is restricted to that age group.

- occupation is recorded for age>14. I use this information for all employed. This is
  available consistently with the exception of institutional inmates who are excluded
  until 1960. The magnitude of this change is small and the time fixed effect will absorb
  the effect on the overall level of employment.

D.2.4 Wage data

- Census data on wages refer to the previous calendar year

- In 1940 and 1960+ every individual replies to this question - in 1950 only sample line
  individuals do (sub-sample)

- Labor earnings are used to be consistent with the model (wages, salaries, commissions,
  cash bonuses, tips, and other money income received from an employer). This differs
  from Piketty et al who use earnings data of tax units. As described above, I use wage
  data and focus on individual data rather than earnings of a tax unit. This choice makes
  economically sense for this setting. The superstar theory is concerned with individual
  labor earnings and abstracts from household composition and capital income.

- Wage data is in real 1950 terms

- The 1940 100% sample is not top coded, other years are. The 99th percentile threshold
  is always below the top code, hence the top code doesn’t pose a problem here.

- Top earners are individuals above the 99th percentile of the US wages distribution
  who report positive earnings. See the text for details on the variable construction.

---

78 There are a number of cases were the unemployed report an occupation. This occurs if they have
previously worked. I construct an employment series that includes such workers for the entire sample period.
This measure is a noisy version of employment as some job losers continue to count as employed. Since the
share of these workers is small, the correction has only small effects on the results.
• As a robustness check I use earners above the 99th percentile within their occupation.

• I calculate measures for top income dispersion in entertainment for each market by year. Measures of income dispersion are not additive across occupations and I therefore calculate a single dispersion coefficient per year-labor market observation. This pools the data for the five occupations affected by TV.

D.2.5 Pareto Interpolation

• Top income shares can be computed straight from the data if the full population is covered. Without information on the full population the standard approach in the literature is to use Pareto approximations (e.g. Kuznets and Jenks (1953); Atkinson et al. (2011); Atkinson and Piketty (2010); Blanchet et al. (2017); Piketty and Saez (2003); Feenberg and Poterba (1993)). This assumes that the income distribution is locally Pareto and interpolates incomes between two observed individuals, moreover it allows to extrapolate the top tail of the distribution. In a Pareto distribution two parameters, pin down the wage distribution. In practice there are a number of challenges. Key to the dispersion is the “Pareto coefficient.” There are at least four challenges in estimating the parameter. The first is misspecification, we do not belief that wages exactly follow a Pareto distribution. Second, outcomes are an order statistic which violates the iid assumption. Third measurement error in wages affects the regressor. Fourth in samples the population rank of an observation is not observed. I address these issues by analyzing the performance of popular methods in years where the full population data allows for validation.

• The beauty of the Pareto distribution is that a it is a straight line in the log space. This holds because the CDF of a Pareto distribution is linear in logs: \( 1 - F(w) = (w/\omega)^{-1/\alpha} \). Once we know two points on the line we can reconstruct the slope and intercept of the line and have fully characterized the distribution. The slope captures the “Pareto coefficient”. The slope is given by: \( \alpha_{i,j} = [\ln(income_i) - \ln(income_j)] / [\ln(rank_i) - \ln(rank_j)] \). Since we usually observe many points we could calculate many Pareto coefficients and combine them in an optimal way. Fortunately economist have thought about the best way of fitting a line through a cloud of points. We can fit a line to estimate the Pareto coefficient by running a regression of the form\(^{79}\):

\[
\ln(income_i) = \beta - \alpha \cdot \ln(rank_i) + \epsilon_i
\]

• It turns out that OLS is a poor approach here. The Gauß Markov assumptions are violated making OLS inefficient and bias. The outcome variables are order statistics, resulting in heteroskedasticity and correlation of errors across observations. Moreover, the log transformation implies that \( E(\epsilon_i) = E(\log \epsilon_i) \neq 0 \), making OLS biased. The latter problem can be addressed by replacing the regressor with the Harmonic index (Blanchet (2016)). And efficiency can be achieved with MLE.\(^{80}\) Polivka (2000) and Armour et al. (2015) give an overview how MLE can be applied to this problem. A further challenge is misspecification. The Pareto distribution is used as an approximation and may not fit the data perfectly. In particular the distribution may fit better at the top.

\(^{79}\)Here \( \beta = \ln(income) - \ln(rank) \) where lower bars represent the lower bound of the interval considered.

\(^{80}\)Since the covariance structure of order statistics is known, GLS yields the same result.
than the bottom of the distribution. Even at the top of the distribution changing Pareto coefficients may be required to fit the data (Blanchet et al. (2017)). Misspecification is particularly problematic for the more efficient estimators (Finkelstein et al. (2006)). I will test the performance of three estimators using real-world data by drawing samples from the full-count Census. This allows us to assess how estimators cope in data with i) small samples, ii) top coding and iii) bunching at tax thresholds and round numbers. I test the following estimators:

- Estimator with \( n \) total observations, \( T \) top coded observations, \( \text{rank}_j \) the rank in the wage distribution (1 being the top), \( w_j \) wage at rank \( j \) and \( \omega \) the smallest wage in the sample:
  
  \[ \hat{\beta}^{MLE} = \frac{1}{n} \sum_{j=1}^{n} \log\left(\frac{w_j}{\omega}\right) \]

- MLE (top code adjusted): \( \hat{\beta}^{MLETC} = \frac{T}{n} \sum_{j=T}^{n} \log\left(\frac{w_j}{\omega}\right) + T \log\left(\frac{w_{TC}}{\omega}\right) \)

- OLS: \( \log(w_j) = \delta - \beta^{OLS} \ast \ln\left(\frac{\text{rank}_j}{n+1}\right) + \epsilon_j \)

- Close to cut-off: \( \hat{\beta}^A = \left(\sum_{j=1}^{3} \frac{\ln(w_j/w_{j-1})}{\ln(\text{rank}_j/\text{rank}_{j-1})}\right)^{-1} \)

- Extrapolation: The standard method of calculating top income shares fits a Pareto curve through the observed data and computes income shares as area under the curve. For the Pareto distribution the fraction that falls in the tail is captured by a single Parameter. We can thus compute any top income share once we know the tail index of the Pareto distribution. For other distributions the tail index varies for different percentiles, in that case we have one shape parameter that allows to compute the top 1% income share and a different one to compute the top 0.1% share. A well known feature of extreme value theory is that in the tail many regular distribution only differ by a slow moving function from the Pareto. Using the Pareto parameter estimate just below the cut-off may thus yield a reasonable approximation even if the data generating process is not Pareto.

- Table D8 shows the results. They suggest that OLS and MLE perform relatively poorly in small samples of the data of interest. I find that the best performing estimator is the average of the alpha values just below the top code. The difference to OLS and MLE estimates is the weight attached to values far from the top-code. OLS and MLE give a non zero weight to observations further away from the top-code. This approach will yield greater bias if the Pareto distribution is not a perfect fit and observations far from the top-code are poor proxies for the distribution beyond the top-code. Consistent with this, I find that the OLS and MLE perform worse in smaller samples. For the application here I therefore focus on Pareto interpolation based on observations closest to the top-code. It should be stressed that this result is specific to the data in this context. More general results for Pareto inference with real-world data should be conducted to establish the wider relevance.

- For each local labor market and year I derive the Pareto coefficient. At the bottom of the income distribution the Pareto distribution has been found be a poor fit, I therefore discard Pareto parameters based on observations at the bottom quarter of
the distribution. The results are however robust to including those observations. Next, I use the local labor market-year specific Pareto coefficient to estimate top income shares. Here I make use of the fact that for a Pareto distribution top income shares are given by: $S_{p\%} = (1 - p)^{\frac{\alpha - 1}{\alpha}}$.

D.2.6 Controls

- Control variables are: share blacks, male, high skilled and median age and income. Most variables are available consistently throughout the sample period. Income and education are only available from 1940 onwards. The race variable has changing categories and varying treatment of mixed race individuals. I use the IPUMS harmonized race variable that corrects for those fluctuations were possible.

D.2.7 IRS Taxable Income Tables

Data from the Internal Revenue Service (IRS) allows me to extend income data backward beyond what is feasible with the Census.\textsuperscript{81} To obtain records for entertainers, I digitize a set of taxable income tables that lists income brackets by state and occupation. The breakdown of the data by occupation and state is only available for the year 1916.

D.2.8 Marginal Tax Rates

I compile data on top income tax rates at the state level from “State Income Tax Administration” (Peniman & Hellar 1959). The study describes the history of state income taxation and collects data on the top income tax rates by state in 1957, as well as information on changes in the tax code since World War II. As far as possible, I use information on tax rates in 1945. This predates most of the TV roll-out and avoids potential endogeneity concerns. Most of the data are collected in 1957 but tax reforms are noted. If no reform is reported I use the 1957 tax rate. I exclude Delaware, where substantial reforms took place between 1945 and 1957. The state tax is levied on top of federal taxes and the top bracket varies from 0 to 11.5 percentage points. This rate however does not reflect the effective marginal tax rate faced by an individual. Allowances and deductions, including for taxes paid to the federal government, lower the effective marginal tax rate in most states. The exact level of the headline tax rate is likely misleading. There are however clear differences in how states use the ability to tax incomes. Many states charge little or no additional income taxes, while others charge significant amounts. I make use of this visible distinction of low/no tax states vs high tax states and classify states as high tax if they charge taxes above the median tax rate. Deductions are unlikely to turn a high tax state into a near-zero tax state. The distinction of high vs low tax state thus captures a meaningful difference in the marginal tax rate faced across the country.

\textsuperscript{81}Such tax tables have been used by Kuznets and Piketty to construct time series of top income shares for the US population.
D.3 APPENDIX: TABLES & FIGURES

Figure D1: Superstar Wage Distribution

Note: Wages based on a superstar model ($w_p = \pi \cdot \kappa \cdot (1 - p)^{-(\alpha \gamma - \beta)}$). $\alpha$ is the shape parameter of the market size distribution ($\alpha' > \alpha$). The percentiles shown are the upper tail of the wage distribution. With exit they correspond to the percentiles in the pre-distribution.
Figure D2: Effect of Technical Change on Wage Distribution - Skill Biased Demand Model

[Note] The figure shows the wage distribution above the 70th percentile. The talent distribution has been chosen to match the 1940 wage distribution. The change in the skill premium matches the growth in the share of top earners.
Figure D3: Superstar Effect on Top Earner

[Note] Details as in figure D1. $w^{US1\%}$ is a wage threshold that defines a top earner, e.g. the national top percentile. $E_{1\%}$ and $E'_{1\%}$ are the share of entertainers above the threshold. $\Delta E_{1\%}$ is the change in top earners when market size becomes more dispersed (move from $\alpha$ to $\alpha'$).
Figure D4: Theatre Seating Capacity

[Note] Performance venues are the venues listed in Julius Cahn-Gus Hill’s 1921 theatrical guide. Size refers to the average seating capacity of the largest two venues in the commuting zone.
Figure D5: P95-P50 Gap

[Note] Figure reports the ratio of wages at the 95th and median. Percentiles are from the wage distribution reported in the US decennial Census for the lower 48 states.

Figure D6: Top Income Percentile Values
Figure D7: Dynamic Treatment Effect of TV station - Placebo Occupations

[Graph showing the dynamic treatment effect of TV station and videotape on share top paid professionals (in ptp) over years 1939 to 1969.]
Figure D8: Quantile Effects of Television

[Note] Each dot is based on separate quantile regression. The quantile regressions control for local labor market and year fixed effect. I use the technique developed in Chetverikov et al. (2016) to do so. This amounts to calculating percentiles for each year-labor market observation and regressing those percentiles on the treatment. The first step uses the provided sample weights, while the second weights by cell size. If the top code bites for the analyzed percentiles, the cell is discarded. The dashed line represents the benchmark prediction of a skill biased demand model.
Table D1: Effect of TV on Top Earner - Placebo Occupations

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local TV station</td>
<td>0.023</td>
<td>0.019</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Outcome mean</td>
<td>9.08</td>
<td>9.08</td>
<td>9.08</td>
</tr>
<tr>
<td>Effect size</td>
<td>2.3%</td>
<td>1.9%</td>
<td>1.6%</td>
</tr>
</tbody>
</table>

**Panel A: ln(Entertainer Wage at 99th Percentile)**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local TV station</td>
<td>0.21</td>
<td>0.66</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>(0.52)</td>
<td>(0.89)</td>
<td>(0.52)</td>
</tr>
<tr>
<td>Outcome mean</td>
<td>5.55</td>
<td>5.55</td>
<td>5.55</td>
</tr>
<tr>
<td>Effect size</td>
<td>4%</td>
<td>12%</td>
<td>20%</td>
</tr>
</tbody>
</table>

**Panel B: Share Entertainer in US Top 1% (ptp)**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local TV station</td>
<td>0.438</td>
<td>0.524</td>
<td>0.865</td>
</tr>
<tr>
<td></td>
<td>(0.221)</td>
<td>(0.234)</td>
<td>(0.319)</td>
</tr>
<tr>
<td>Outcome mean</td>
<td>10.86</td>
<td>10.86</td>
<td>10.86</td>
</tr>
<tr>
<td>Effect size</td>
<td>4%</td>
<td>5%</td>
<td>8%</td>
</tr>
<tr>
<td>Cluster</td>
<td>722</td>
<td>722</td>
<td>722</td>
</tr>
<tr>
<td>Demographics</td>
<td>–</td>
<td>Yes</td>
<td>–</td>
</tr>
<tr>
<td>Local labor</td>
<td>–</td>
<td>–</td>
<td>Yes</td>
</tr>
</tbody>
</table>

[Note] Each cell is the regression coefficient of a separate regression. Panel A uses a quantile regression for within group treatment Chetverikov et al. (2016). For this procedure data is aggregated at the treatment level and uses 2,887 local labor market - year observations. Observations are weighted by cell-size, cells where 99th percentile cannot be computed are dropped. Panel B and C use a difference in difference regression and are based on respectively 62,042 and 62,746 observations at the treatment level. The treatment is the number of TV stations in the local area. Reported baseline outcomes are the average of the dependent variable in treated areas in years without treatment. All regressions control for local labor market fixed effects, time fixed effects, local production cost of filming in years after 1956, in Panel B and C additionally for year - occupation fixed effects. The sample period spans 1940-1970. Demographics are median age, % female, % black, population density and trends for urban areas. The outcome variable in Panel B is the share of top paid entertainers calculated as described in the text, Panel C is the number of top paid entertainer divided by the population in a local labor market. Entertainer are Actors, Athletes, Dancers, Entertainers Not Elsewhere Classified, Musicians. Observations are weighted by local labor market population. Standard errors are reported in brackets, they are clustered at the local labor market level.
Table D2: Effect of TV on Top Earner - Alternative Top Income Measures

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Count Entertainer in US top 1%</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local TV station</td>
<td>30.91</td>
<td>32.09</td>
<td>19.31</td>
</tr>
<tr>
<td></td>
<td>(8.92)</td>
<td>(9.92)</td>
<td>(8.31)</td>
</tr>
<tr>
<td>Outcome mean</td>
<td>15.53</td>
<td>15.53</td>
<td>15.53</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel B: Share Entertainer in US top 1% (denominator fixed)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local TV station</td>
<td>6.51</td>
<td>6.73</td>
<td>9.21</td>
</tr>
<tr>
<td></td>
<td>(1.90)</td>
<td>(1.89)</td>
<td>(3.44)</td>
</tr>
<tr>
<td>Outcome mean</td>
<td>6.39</td>
<td>6.39</td>
<td>6.39</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel C: Share Entertainer in US top 1%</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local TV station</td>
<td>0.178</td>
<td>0.193</td>
<td>0.194</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.038)</td>
<td>(0.063)</td>
</tr>
<tr>
<td>Outcome mean</td>
<td>0.28</td>
<td>0.28</td>
<td>0.28</td>
</tr>
<tr>
<td>Cluster</td>
<td>722</td>
<td>722</td>
<td>722</td>
</tr>
<tr>
<td>Demographics</td>
<td>–</td>
<td>Yes</td>
<td>–</td>
</tr>
<tr>
<td>Local labor</td>
<td>–</td>
<td>–</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>market trends</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note* See table 1. Panel B denominator is the average number of entertainers per labor market in occupation o at time t. Denominator in Panel C is the total number of entertainers in local labor market c at time t.
### Table D3: Alternative Top Income Measures

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Share in US top 1%</td>
<td>Count top 1%</td>
<td>Share in top 5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local TV station</td>
<td>90.19</td>
<td>132.5</td>
<td>30.91</td>
<td>31.64</td>
<td>120.0</td>
</tr>
<tr>
<td></td>
<td>(26.25)</td>
<td>(35.92)</td>
<td>(8.92)</td>
<td>(16.36)</td>
<td>(47.85)</td>
</tr>
<tr>
<td>threshold</td>
<td>Census</td>
<td>Piketty &amp; Saez</td>
<td>Census</td>
<td>Entertainer</td>
<td>Entertainer</td>
</tr>
<tr>
<td></td>
<td>mean outcome</td>
<td>94.27</td>
<td>109.09</td>
<td>18.39</td>
<td>150.02</td>
</tr>
<tr>
<td></td>
<td>% growth</td>
<td>96%</td>
<td>121%</td>
<td>168%</td>
<td>21%</td>
</tr>
</tbody>
</table>

[Notes] Different thresholds for top earners: column (1) top 1% in overall distribution based on Census wage, (2) top 1% in overall distribution based on Piketty and Saez (2003) (3) count of entertainer in top percentile, (4) 95th percentile of entertainer wage distribution, (5) 95th percentile of entertainer in 1940. Source: Data US Census and Piketty & Saez. Specification and sample same as baseline.

### Table D4: Effect of TV on Top Earner - Micro Data

<table>
<thead>
<tr>
<th>Probability in Top 1%</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TV × Performance Entertainer</td>
<td>0.74</td>
<td>0.76</td>
<td>0.79</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>(0.23)</td>
<td>(0.26)</td>
<td>(0.36)</td>
<td>(0.22)</td>
</tr>
<tr>
<td>TV × Interactive Leisure</td>
<td>-0.49</td>
<td>(0.34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TV × Drink &amp; Dine</td>
<td>-0.65</td>
<td>(0.48)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TV × Professional Services</td>
<td>0.32</td>
<td>(0.21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TV × Medics</td>
<td>-1.54</td>
<td>(0.60)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TV × Engineer</td>
<td>-0.09</td>
<td>(0.26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TV × Manager</td>
<td>0.43</td>
<td>(0.28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Location &amp; Occupation-Year FE</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Demographics</strong></td>
<td>–</td>
<td>Yes</td>
<td>–</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Local labor market trends</strong></td>
<td>–</td>
<td>–</td>
<td>Yes</td>
<td>–</td>
</tr>
</tbody>
</table>

[Notes] The outcome is a dummy that takes the value 100 if an individual is in the top 1% in the US distribution. Columns 1-3 are based on 83,748 individuals and column 4 on 3,438,042 individuals. Placebo occupations are non affected free time professions: drink & dining and active leisure and typical high pay professions: management, medicine, engineering, professional services (finance, accounting, law). The number of observations are 100308. Regressions use provided Census weights and cluster by local labor market.
Table D5: Effect of TV on Top Earner - State Level

<table>
<thead>
<tr>
<th></th>
<th>Share in Top 1%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Local TV station (1940)</td>
<td>-9.62</td>
</tr>
<tr>
<td>Local TV station (1950)</td>
<td>20.94</td>
</tr>
<tr>
<td></td>
<td>(8.09)</td>
</tr>
<tr>
<td>Local TV station (1960)</td>
<td>-9.95</td>
</tr>
<tr>
<td>Local TV station (1970)</td>
<td>-13.33</td>
</tr>
<tr>
<td>Observations</td>
<td>912</td>
</tr>
</tbody>
</table>

[Notes] Data US Census (1940-1970) and IRS in 1916. The regressor is the number of TV stations in 1950 in the state, allowing for time varying effects. In column 3 the omitted year is 1916. Standard errors are clustered at the state level.

Table D6: Earning Effect - triple diff

<table>
<thead>
<tr>
<th></th>
<th>Share in Top 1%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>TV × Placebo Occupation</td>
<td>-0.41</td>
</tr>
<tr>
<td>TV × Performance Entertainer</td>
<td>4.87</td>
</tr>
<tr>
<td>TV × Interactive Leisure</td>
<td>-3.40</td>
</tr>
<tr>
<td>TV × Drink &amp; Dine</td>
<td>-3.80</td>
</tr>
<tr>
<td>TV × Professional Services</td>
<td>5.23</td>
</tr>
<tr>
<td>TV × Medics</td>
<td>-3.24</td>
</tr>
<tr>
<td>TV × Engineer</td>
<td>-1.12</td>
</tr>
<tr>
<td>TV × Manager</td>
<td>3.55</td>
</tr>
<tr>
<td>Location &amp; Occupation-Year FE</td>
<td>Yes</td>
</tr>
<tr>
<td>Pairwise Interaction: Location, Year, Occupation FE</td>
<td>–</td>
</tr>
</tbody>
</table>

[Notes] Data and specification are as in 1. Placebo occupations are non affected free time professions: drink & dining and active leisure and typical high pay professions: management, medicine, engineering, professional services (finance, accounting, law). The number of observations are 100,308.
Table D7: Quantile Effect of TV

<table>
<thead>
<tr>
<th>Wage Percentiles</th>
<th>99th</th>
<th>95th</th>
<th>75th</th>
<th>50th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local TV station</td>
<td>260.3</td>
<td>85.00</td>
<td>22.33</td>
<td>19.13</td>
</tr>
<tr>
<td></td>
<td>(92.23)</td>
<td>(3412.5)</td>
<td>(445.3)</td>
<td>(101.2)</td>
</tr>
</tbody>
</table>

[Notes] The reported coefficients are estimates using the quantile estimator for within group transformation developed in Powell (2016).

Table D8: Small Sample Performance of Pareto Shape Parameter Estimators

<table>
<thead>
<tr>
<th>Estimator</th>
<th>sample 10%</th>
<th>local 5% sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.460</td>
<td>0.460</td>
</tr>
<tr>
<td>OLS</td>
<td>0.558</td>
<td>0.715</td>
</tr>
<tr>
<td>MLE</td>
<td>0.617</td>
<td>0.629</td>
</tr>
<tr>
<td>MLE (top code)</td>
<td>0.640</td>
<td>0.618</td>
</tr>
<tr>
<td>Close to cut-off</td>
<td>0.478</td>
<td>0.480</td>
</tr>
</tbody>
</table>

The true $1/\alpha$ is the value implied by the top 5% income share. The simulation draws samples from the entertainer wage distribution in the 1940 US full count Census. The samples are top coded at the 99th percentile of the distribution. Column 1 fits estimators on 10% samples dropping observations in the bottom half of the sample. Column 2 draws a smaller sample equivalent to a 5% sample of local labor markets. Estimates that imply an infinite mean are discarded ($\alpha < 1$).