Technical Change and Superstar Effects: Evidence from the Roll-Out of TV*

Felix Koenig§

June 2019

Abstract

Top incomes have risen sharply across Western economies in recent decades and we have limited understanding why. This paper provides a quasi-experimental test for the Superstar explanation which holds that technical change is causing top income growth by enabling workers to reach bigger markets. I use the historic experience with such technical change in the entertainment for a test of the Superstar Effect. The launch of TV in the 1940s markedly increased market reach of entertainers and regulatory and technical constraints led to quasi-random variation in exposure across local labor markets. I use this variation in a differences in differences analysis and find strong support for the Superstar Effect. Top incomes in entertainment grow sharply, around 17% at the 99th percentile and the pattern of labor market changes closely aligns with Superstar Effects: the wage distribution becomes more right skewed, while mid paid jobs disappear and overall employment falls. I quantify the magnitude of Superstar Effects by estimating the model's key elasticity and find that when market size doubles, top pay grows about 16%.

Keywords: *Superstar Effect, Inequality, Top Incomes, Technical Change*
JEL classification: J31, J23, J24, M52, O33, D31

*I thank Daron Acemoglu, Josh Angrist, David Autor, 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, Alan Manning, Ben Moll, Niclas Moneke, Markus Nagler, Barbara Petrongolo, Steve Pischke, David Price, Arthur Seibold, Marco Tabellini, Lowell Taylor, Catherine Thomas, Joonas Tuhkuri, Anna Valero and John van Reenen as well as seminar participants at LSE, MIT, Goethe University, ZEW, RES Junior Symposium, IZA summer school, EMCON, Ski and Labor, EEA Annual Congress and EDP Jamboree. §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

Sharp top income growth has been a striking feature of labor market around the globe in recent decades. In the US for instance the share of income accruing to the top 1% of the distribution has risen from about 8% in the 1970 to over 20% in recent years. Why has most income growth occured at the top of the distribution? Such questions about the causes of inequality have been at the heart of economics since the beginning of the field, Tinbergen for instance concluded that "the fairly satisfactory state of affairs with respect to the statistical description of income distribution contrasts with an unsatisfactory state in the area of economic interpretation (Tinbergen, 1956)." Even today, we only have a handful of economic explanations for the rise of top incomes – prominently among them is the theory that technical change is driving inequality by creating Skill Biased Technical Change (SBTC) and Superstar Effects (Acemoglu & Autor, 2011, provide an overview). SBTC arises when technology enhances the productivity of specific "groups" of workers, in recent decades skilled workers, which leads to a growing wage gap between skilled and unskilled workers. Superstar Effects on the other hand emerge when technology facilitates production on a larger scale. Such technical change enables the most talented workers in the profession – the "superstars," to attract a greater share of customers, which leads to income concentration at the top and a moves the labor market towards a winner takes all market. Despite the prominence of the Superstar Effect in the public and academic debate, there is little direct empirical evidence for such effects.

The empirical test of the Superstar Effect is complicated by endogeneity issues. A first key challenge is to separate the effect of technology from other simultaneous changes in inequality. This issue is highlighted in the debate about SBTC effects, where early work indicated that SBTC can explain much of the rise in inequality in the US in the second half of the 20^{th} century, however a "revisionist literature" points out that it is difficult to isolate the impact of SBTC from other trends in the labor market (See for instance Katz and Murphy, 1992; Card and DiNardo, 2002; Lemieux, 2006; Autor, Katz, and Kearney, 2008). Recent studies have tackled this issue by testing the SBTC theory using natural experiments and find evidence that new technologies lead to skill biased demand shifts (Bartel, Ichniowski, & Shaw, 2007; Akerman, Gaarder, & Mogstad, 2013; Michaels & Graetz, 2018). By contrast, surveys of the literature point out that the empirical evidence for Superstar Effects is thin and predominantly focuses on cross-country or cross-industry comparisons

(Katz & Autor, 1999; Lemieux, 2008; Kaplan & Rauh, 2013). Such studies point out that global technological trends ought to generate similar top income trends globally (Piketty & Saez, 2006; Kaplan & Rauh, 2013). However, this approach assumes that that there are no other confounding forces that could explain different income trends across settings.

In this paper I use a historic natural experiment to test the Superstar Effect directly. The test leverages an iconic shift in production scalability: the launch of TV in the mid 20th century, which vastly increased the audience available to entertainers. 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.

This approach allows me to addresses two key endogeneity issues faced by the literature. A first attraction is that regions are differently exposed to technical change, allowing me to test the effect *within* a country and industry, while holding aggregate trends in regulation, labor market-institutions or pay-setting norms constant. Differential regional exposure arises because technical constraints forced early TV filming to take place near broadcast antennas. The launch of a local TV station thus led to a sharp rise in market reach for local entertainers, similar to the construction of a hypothetical giant theater that would hold an entire local population. I use the staggered local deployment of TV stations for a difference in differences analysis across local labor markets. A second appeal of this setting is that it allows me to tackle the challenge posed by endogenous technology adaption, which invalidates the identification assumption of a difference in differences analysis for most technologies.[1] In the TV case, technology is not adapted by market participants but deployed by the government. I exploit deployment rules that are unrelated to local labor market conditions for identification and verify the validity of this approach during an unexpected interruption of the roll-out where a number of locations narrowly miss out on a planned TV launch.

The data provide overwhelming support for the main prediction of the Superstar Effect, that scalable production technologies lead to top income growth. The launch of television filming leads to a highly significant increase in entertainer top incomes. Pay at the 99th percentile increases 17% and the income share of the top percentile nearly doubles. Entertainers rise visibly in the US wage distribution, which reflects the fact

[1]Evidence for endogenous technical change is presented in Blundell, Griffith, and Van Reenen (1999), the theory in Acemoglu (1998).

that the effect is specific to the entertainment sector. The share of entertainers in the top 1% of the US distribution approximately doubles with the launch of a TV station. By contrast, locations where the launch of a TV station is unexpectedly blocked see no growth in top entertainer pay.

The paper next distinguishes the Superstar Effect from alternative mechanisms, including conventional labor demand models and SBTC. To this end, I derive additional characteristic predictions of the Superstar Effect. Classic comparative static results of the superstar model rely on knowledge of the talent distribution, which makes such predictions notoriously difficult to test (e.g. Rosen, 1981). I use a tractable model of the Superstar Effect to derive alternative predictions that are independent of the talent distribution and allow me to differentiate the Superstar Effect from a wide class of alternative models. Expanding market reach moves labor market towards a winner-takes-all market, and the ensuing pattern of wage changes gives rise to additional auxiliary predictions. Specifically, Superstar Effects, lead to a decline in employment and growing income dispersion in the top tail of the distribution. The data shows striking evidence for the predicted labor market change and strongly supports the Superstar Effect.

I quantify the magnitude of Superstar Effects by estimating the key elasticity of the superstar model and find that doubling market size increases wages at the 99th percentile around 16 percent. I use these results to estimate the size of superstar effects and find that the superstar effect explains about 70% of top income variation in this setting. Previous studies have used the superstar model to explain income dynamics, focusing on rock-stars, CEOs and the US economy at large (Cook & Frank, 1995; Krueger, 2005; Terviö, 2008; Gabaix & Landier, 2008; Kaplan & Rauh, 2013). Many of these studies use anecdotes to support the superstar hypothesis, while others use the model to fit the wage data by calibrating the relevant model parameters to the correlation of market size and top pay. My setting allows me to complement such OLS estimates with an instrumental variables estimate.

The rise of superstars has also been at the centre of the public policy debate. Modern digital and online technologies are making production more scalable which has led to speculation that we could soon see a surge in Superstar Effects (e.g. OECD, 2016; Guellec & Paunov, 2017). In this study test the labor market consequences of scalable production in an industry that experienced this change decades ago. A further related debate focuses on the rise of "superstar firms." Such studies point out that a small number of firms serve a growing share of the market and apply the superstar logic to the realm of firms (Song, Price, Guvenen, & Bloom, 2019; Autor, Dorn, Katz, Patterson, & Van Reenen, 2017). Adding to this debate, this paper tests the canonical explanation for the rise of superstars and sheds light on the conditions that lead to rising market concentration. Finally, Superstars Effects have been central to the debate about tax policy, optimal tax rates depend on the economic mechanisms that generate top income growth (Kleven, Landais, Saez, & Schultz, 2013, 2014; Scheuer & Werning, 2017; Moretti & Wilson, 2017). It is therefore vital to understand whether Superstar Effects are at play in practice for the optimal design of tax systems.

The remainder of this paper derives key predictions of the superstar model in Section 2, describes the data in Section 3, reports empirical results in Section 4 and discusses the magnitude of Superstar Effects in Section 5, Section 6 concludes.

2 The Superstar Model

This section develops a tractable model of the superstar effect that illustrates the key predictions of the model and distinguish it from conventional models of labor demand. In the superstar model workers with different and unique levels of talent are matched with heterogenous tasks.[2] In the context of entertainment we can think of workers as actors and of tasks as shows. Denote the inverse CDF of talent t and show venue size s by $p^t \equiv P(t > t_p) = h(t_p)$ and $p^s \equiv P(s > s_p) = g(s_p)$ and assume that these distributions are continuous.[3] In production actors are matched with a stage and revenue of a matched pair is given by $Y(s,t)$ with $Y_s > 0, Y_t > 0, Y_{tt} < 0$ and $Y_{st} > 0$. The final assumption here is comparative advantage of more talented actors in bigger markets, which will be essential to the solution of the model.[4]

The equilibrium of this model is characterized by three conditions, which respectively reflect positive assortative matching (PAM) between actors and shows (equation 1), incentive compatibility of the assignment (Equation 2) and market clearing (derived in Appendix C.1). Together these conditions pin down the equilibrium

[2]Differences in job characteristics have also been called "market size" or "firm value". Key is that the characteristics cannot be changed at the time of hiring.

[3]This is not essential to the model but helps the presentation. With jumps in the distribution one actor would be discretely better than the next, which would generate monopoly power and lead to match specific rents.

[4]This assumption generates positive assortative matching in equilibrium. Alternatively, papers have used a log-supermodularity assumption but neither assumption is stronger in the sense that one implies the other.

values of wages $w(\hat{t})$, equilibrium assignment $\sigma(\hat{t})$, and output prices $\hat{\pi}$:

$$p^t = p^s(\sigma(\hat{t})) \tag{1}$$

$$w'(\hat{t}) = Y_t(\sigma(\hat{t}), \hat{t}) \tag{2}$$

$$\int -Y(\sigma(\hat{t}), \hat{t})h'(\hat{t})d\hat{t} = D(\hat{\pi}) \tag{3}$$

PAM in equation 1 implies that matched actors and stage partners are at the same percentiles of their respective size and talent distributions. Note that incentive compatibility in equation 2 guarantees that equilibrium wages increases in line with the marginal product of workers. The equilibrium is therefore perfectly competitive in the sense that there are no match specific rents. The final condition, equation 3, ensures that demand $D(\pi)$ equals supply at the equilibrium prices $\hat{\pi}$.[5]

The resulting superstar wage distribution has been the key focus in the literature. Distributional results have been derived for the general case but to illustrate the key mechanics of the model I will focus on a closed form solution. Assume that talent t and show size s follow Pareto distributions with shape parameters α and β respectively ($p^t = t_p^{-\frac{1}{\alpha}}$ and $p^s = s_p^{-\frac{1}{\beta}}$). A larger value of the shape parameter implies greater dispersion in the distribution. Moreover, assume that the production function is Cobb-Douglas $F(s,t) = \pi\left[s^{(1-\gamma)}t^{\gamma}\right]^{\phi}$, where $\gamma \,\epsilon\, (0,1)$ and ϕ determines the scalability of production ($\phi > 0$). These assumptions simplify the equilibrium conditions, the PAM condition 1 solves for $\sigma(\hat{t}) = \hat{t}^{\frac{\beta}{\alpha}}$. And using this result, the wage schedule in 2 becomes: $w'(\hat{t}) = \gamma\phi\hat{\pi}\hat{t}^{(\frac{1}{\xi}-1)}$, with $\xi \equiv \frac{1}{\phi}\frac{\alpha}{\gamma\alpha+(1-\gamma)\beta}$. Integrating this wage schedule and normalizing the constant of integration, or workers' outside option, to zero we arrive at the wage distribution in the superstar economy. After applying $\sigma(\hat{t})$ and the talent distribution the wage distribution can be written as:

$$p^w = \lambda w_p^{-\frac{\xi}{\alpha}} \tag{4}$$

Wages follow a Pareto distribution, with the shape parameter $\frac{\alpha}{\xi}$ and scale parameter $\lambda \equiv (\gamma\phi\pi)^{\xi/\alpha}$. Notice that the wage distribution looks similar to the distribution of talent. The shape parameter of the talent distribution is α, hence wages are more

[5]We assume that workers supply one unit of labor inelastically, an assumption that is relaxed below.

dispersed than talent if $\xi < 1$. For small values of ξ the superstar model therefore produces large wage differences, even if talent differences are small. This "talent amplifier effect" has been the focus of much early literature on superstar models (e.g. Rosen, 1981; Tinbergen, 1956; Sattinger, 1975).[6] It is however difficult to test the talent amplifier effect because such a test requires a cardinal measure of talent. The lack of an objective talent unit thus makes the model indistinguishable from an alternative model without talent amplifier effect and extremely rare talent.

2.1 The Effect of Technical Change

This issue can be addressed by looking at time series predictions of the superstar model. Consider a technical change that makes production more scalable ($\tilde{\phi} = \kappa\phi$ with $\kappa > 1$). The best known result of the Superstar Effect is sharp top income growth, yet we can go further and derive additional results that differentiate the superstar effect from a wide range of alternative mechanisms. To study the impact of Superstar Effects across the wage distribution consider the share of jobs with wage w, denoted by $f(w)$, which we derive by differentiating equation 4:

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

and denote the change in the share of jobs with wage w by $g_e(w)$ with "$\tilde{x}$" indicating new values:

$$g_e(w) = \frac{\tilde{f}(w)}{f(w)} = \frac{\tilde{\lambda}\tilde{\xi}}{\lambda\xi} w^{\frac{(\kappa-1)}{\kappa}\frac{\xi}{\alpha}} - 1 \tag{5}$$

This result shows that technical change leads to a growing fraction of top paid actors, as $g_e(w)$ is increasing in w and will be positive for large w. Such effects are becoming more pronounced at higher values of w, which shows that technical change causes disproportional gains at the top of the distribution in the superstar model. As we move away from the top of the distribution such positive effects die out. A second effect operates through the drop in λ, the greater availability of stars, reduces demand for the rest of the profession, reduces π and in the limit, a single superstar serves the entire market.[7] The impact of Superstar Effects across the wage distribution are

[6] In what follows I focuses on cases where this amplifier effect holds. This is the case as long as large show venues are scarce enough to overcome diminishing returns to scale as we move up in the distribution (aka if $\frac{\beta}{\alpha} > \frac{1-\gamma}{\gamma}$).

[7] Notice that if π is unchanged (ie if demand for entertainment is perfectly inelastic), λ would rise. I assume that demand is sufficiently elastic ($1 - \varepsilon < (\gamma\phi\pi)^{\kappa-1}$) to rule this case out, for an

summarized in Figure 1. High paid jobs emerge at the very top, while growth rates diminish as we move away from the stars of the profession. Towards the middle of the distribution mid-income jobs are disappearing as mediocrely talented workers are moved into the low paid sector, leading to growing share of workers with low pay. The impact of Superstar Effect across the wage distribution is therefore U-shaped.

The distributional consequences of Superstar Effects can be summarized by four testable propositions (derived in Appendix C.3):

Proposition 2.1. *Superstar Effects lead to*

a) Top pay growth: For two percentiles at the top of the wage distribution $p' > p$ *the growth rate* g_e *increases as we move up in the distribution:* $g_e^{p'} > g_e^{p}$

b) Fractal inequality: For top income shares (s_p) *at two percentiles p:* $\tilde{s}_{1\%}/\tilde{s}_{10\%} > s_{1\%}/s_{10\%}$

c) Mediocre worker pay: When $w \to 1$ *the growth rate turns negative* $g_e < 0$ *and the share of mid paid jobs declines*

d) Employment loss: For a participation threshold $\bar{p}$ *superstar effects imply* $\tilde{\bar{p}} > \bar{p}$

The first result in *a*) is disproportionate gains at the top of the income distribution. The second result in *b*) is "fractal inequality" and focuses on growing income dispersion *within* the top income tail, which reflects that moving up a rank in the talent distribution becomes more valuable. As a result, a growing proportion of the the income earned by the top 10% is earned by the top 1% and consequently the ratio of the two income shares ($s_{1\%}/s_{10\%} \uparrow$) increases. A third result in *c*) highlights that gains at the top come at the expense of other entertainers. Technical progress allows stars to steal business of lesser stars, leading to falling demand for mediocre entertainers. The final results in *e*) captures the winner takes all nature of superstar effects, as employment falls when the stars' growing market reach pushes other workers out of the market.

Finally, consider why and when Superstar Effects differ from canonical models of labor demand. Workers in the superstar models have a unique quality of talent, making workers imperfect substitutes, while canonical models, including SBTC, feature skill groups of perfectly substitutable workers. Among perfectly substitutable workers the law of one price limits growth in wage dispersion and makes wage growth

alternative approach to generate this result see Rosen, 1981. Even without these cannibalizing effects, the previous results on relative gains at the top hold. This is for instance the case in Gabaix and Landier, 2008.

proportional to talent, at odds with results a) and b) above.[8] In the superstar model on the other hand being slightly better has huge rewards at the top, while it barely matters for ordinary workers. The limitation of the canonical model holds even if we introduce a continuous distribution of worker talent (see Appendix C.4) and only breaks if each worker is in a unique skill group and hence when we replicate the superstar case where all workers are unique types. This extension of the canonical case is however an unappealing model as it features as many parameters as workers, making it effectively unfalsifiable. A second distinctive feature of Superstar Effects are the losses incurred by large parts of the distribution. This is again at odds with canonical labor demand models, where wages and employment increase with a positive demand shock, which contradicts results c) and d).[9] To see this take the classic case of a CES production function, here a positive demand shock for one worker group leads to wage growth across the board as skill groups as Q-complements.

In summary, we can distinguish the Superstar Effect by taking 4 predictions to the data: a) disproportional wage growth at the top, b) decreasing wages for mediocre workers, c) falling employment and d) growing dispersion of wages at the top in the data.

3 Data

The entertainment sector in the mid 20th century provides sharp variation in production scalability and a rare opportunity to simultaneously observe changes in production technology, labor market outcomes and market reach. I combine records from archival sources to measure all three changes and isolate plausibly exogenous variation in market reach.

3.1 Production Technology

The main sources of variation in production scalability arises through the launch of TV filming. At the time, filming took place locally as non-local shows were a poor substitute for local productions for several reasons. Foremost, the infrastructure to air shows simultaneously across stations was lacking (see Sterne, 1999 for a detailed

[8] Appendix C.4 shows these limitations of SBTC and illustrates the extensions that are required to make the model match superstar effects.

[9] Models where technical change improves the productivity of capital, instead of workers, (e.g. routinization) may lead to wage and employment losses if capital is a substitute for workers.

account). Moreover, while storing and transporting shows was in principle feasible, but the technology was costly and turned out to be unpopular because it led to poor image quality.[10] Finally, regulation restricted studio locations by specifying that "the main studio be located in the principal community served" (FCC annual report 195). I collect and geocode the location of TV studios using data from the "Annual Television Factbooks" to document how the the technology emerges across local labor markets. I compute the number of stations filming in the local labor market for each year.[11] Figure 2 shows the variation in the year 1949, a year with Census wage data and illustrates that there were vast differences in the exposure to this new technology.

For identification it will be essential to understand the rules that govern the local deployment of TV. I collect archival information on how locations are prioritized. Administrative records on roll-out rules were published in a few years and show that the priority ranking of the TV roll-out was based on fixed location characteristics. This lends credibility to the identifying assumption that the TV timing did not respond to local demand shocks. The identification strategy relates to studies that look at spread of TV signal across the US (e.g. Gentzkow, 2006; Gentzkow and Shapiro, 2008). Different from those, I focus on variation in TV filming and use novel administrative records and the subsequent demise of local filming to sharpen the identification strategy.

A unplanned halt in licensing leads to additional quasi-experimental variation in TV. The principal reason for the interruption was an error in the FCC's airwave propagation model. This model was used to delineate interference free signal catchment areas, but the error implied that signal interference occurred among neighboring stations. To avoid a worsening of the situation, the FCC put all licensing on hold and ordered a review of the model. I use newly collected administrative records to identify locations that were held up by the freeze and thus missed out on TV, Figure 2 shows affected labor markets.[12] Licensing only resumed in 1952 and the onset of TV was delayed by four years in many markets.[13]

[10] Non-local content had to be put on film and shipped to other stations, where a mini film screening was broadcast live, this was known as "kinescope".

[11] assume that all stations were filming locally at that time. A handful of stations are an exception and operated a local network. This 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 to avoid potential endogenous selection of filming locations within the network.

[12] Previous studies have used stations that go live after the freeze as proxy for the hold up.

[13] Initially the freeze was expected to last about a year. However, the review was delayed to ensure compatibility with arising new transmission technologies (UHF and color transmission).

Finally, I track the decline of local TV filming as a result of the invention of the videotape in 1956.[14] I use this variation to test whether local Superstar Effects disappear when the local filming is removed. The Ampex videotape made shows from outside the local labor market a close substitute for local live shows and led to the concentration of TV production in two hubs, Los Angeles and New York and the demise in other locations. The year the videotape was presented, 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 and thus led to the rapid decline of local filming. To control for places where filming centralizes and avoid an endogenous control problem, I use a pre-determined measure of production costs, which picks up location incentives that come from permanent regional characteristics. As measure I use the share of 1920 movies produced in each local labor market, based on filming locations of movies in the "Internet and Movie Database" (ImDB).

3.2 Labor Market Data

Labor market data come from the micro-data files of the US Census (1930-1970). I focus on five entertainment occupations that benefitted from the introduction of TV: actors, athletes, dancers, musicians and entertainers not elsewhere classified. For each of them I compute outcomes at the local labor market level for the 722 local labor markets that span the mainland US.[15] The Census first collected wage data in 1940 and in all years asked about the previous year. 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. A first set of variables looks at entertainers' position in the US wage distribution and follows Chetty, Hendren, Kline, and Saez, 2014 in measuring inequality by ranking entertainers' wages relative to a benchmark group. This ranking metric has the advantage that it is scale independent and thus makes it easier to compare changes in pay inequality over time and simultaneously allows to side-steps top coding issues, since the share of workers with a wage above a threshold, say the 99th percentile, can be computed. For the analysis I ensure that fluctuations in the denominator do not bias my results by fixing the denominator

[14] This trend was also helped by the contemporaneous roll-out of coaxial cables that allowed to transmit live shows from station to station.

[15] I follow Autor and Dorn, 2013 and define commuting zones as local labor markets.

above the treatment level.[16] The share of top paid entertainers is thus computed by dividing the number of top-earning entertainers in market m at time t by the number of entertainers in a standard labor market and does not vary at the treatment level.[17] Take for example the share of entertainers whose wage falls in the top 1% of the US wage distribution ($D^{US1\%} = 1$):

$$p_{m,t}^{99} = \frac{\sum_i E_{i,m,t} \cdot D^{US1\%}}{\overline{E_t}}$$

A related outcome measure is the share of income going to top percentiles in entertainment. To compute such shares I have to take a stance on the distribution beyond the top code and I follow the literature in using Pareto approximations.[18] Moreover, I compute additional outcome measures: top-paid entertainers per capita, log employment and geographic mobility of entertainers (see Appendix D.3 for details).[19]

3.3 Demand for Entertainers

The entertainment setting offers a unique opportunity to quantify market reach of workers by analyzing audience size. I digitize audience and revenue records of live and TV shows from archival sources. For live shows I use the 1921 "Julius Cahn-Gus Hill theatrical guide", which aims to provide "complete coverage of performance venues in US cities, towns and villages."[20] For TV shows I compute the number of TV households in a station's catchment area using signal data from Fenton and Koenig, 2018 and Census data on TV ownership. Moreover, I use TV stations' "rate cards" to compute the revenue of local shows. TV shows provided an enormous step-up in the revenue and audience of entertainment shows.[21] Live shows reached on average

[16] Results without the normalization, as presented in the Appendix 10, are in line with the baseline.

[17] I normalize by the average number of entertainers in treated labor markets to simplify interpretation of the treatment effects as percentage point changes,

[18] For top income shares I focus on the larger 350 markets, details are described in Appendix D.3.

[19] Note that the definition of mobility varies across Census vintages. Moreover, it does not distinguish between moves within and across labor markets. Noise in the outcome variable will inflate standard errors but not necessarily bias the estimates.

[20] The theatrical guide covers seating capacity and ticket prices of over 3,000 performance venues that cover ca. 80% of US local labor markets. 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 imperfect and small or pop-up venues will be missed. Since we focus on star venues this omission may be a lesser concern. I use the largest available audience in the labor market as proxy for stars' shows. Spot checking confirms that the data accurately cover physical performance establishments.

[21] Details on revenue data are in Appendix D.3. For TV shows, prices are imputed based on an demand elasticity estimated in a subset of 451 markets where data is available.

1,165 people before TV, while the median TV station could reach around 75,000 households.

I collect information on attendance and spending at county fairs to document demand shifts for non-star entertainment. My data spans ticket sales and revenues for over 4,000 county 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. I aggregate local spending in three categories that are differentially close substitutes for television: spending on shows (e.g. grandstand shows), fair tickets, and carnival items (e.g. candy sales and fair rides). Fair shows most closely resembled TV shows at the time, while candy sales and fair rides are by nature less substitutable with TV. County fairs started to face competition from TV entertainment when TV signal becomes available in the local area. Note that these are not the same places as have access to TV filming, as signal travels beyond the local labor market of a station. To measure TV signal I again use signal data from Fenton and Koenig, 2018.[22] Figure 3 shows TV signal in 1950 and illustrates areas that narrowly miss out on TV signal due to the freeze in licensing.

4 Empirical Tests

Before turning to identification consider the aggregate transformation of the entertainment sector throughout the roll-out of TV. Before TV, most entertainers earned close to average pay, while pay dispersion grew substantially after the introduction of TV. Figure 4a shows that by 1970 wages at the top had grown disproportionally, many 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 4b). Such concentration of demand on a few stars and the pattern of rising dispersion in log pay, precisely what the Superstar Effects predicts.

An ideal test of the Superstar Effect would randomize production technologies across labor markets. Such experiments at the labor market level are hard to implement. To get close to this ideal, I exploit the staggered introduction of television across local labor markets (m) and test the effects in a difference in differences

[22]Similar TV signal data has been widely used to study the effect of TV watching (e.g. Angelucci, Cagé, and Sinkinson, 2019; Gentzkow and Shapiro, 2008; Gentzkow, 2006).

regression:

$$Y_{mot} = \alpha_m + \delta_{ot} + \gamma X_{mt} + \beta TV_{mt} + \epsilon_{mot} \tag{6}$$

where α_m and δ_{ot} are labor market and occupation-year fixed effects; X_{mt} is a vector of time varying labor market characteristics and the treatment variable TV_{mt} is the number of TV stations producing local shows. I run the regression at the more disaggregated labor market-year-occupation 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 disaggregated level will not artificially lower standard errors.

The main source of variation in TV_{mt} comes from the staggered deployment of TV stations (See Section 3 for details). By leveraging regional differences in exposure to technical change the identification can hold aggregate effects, for example trends in regulation and norms, fixed.

The key identification assumption is that TV launch dates are unrelated to local trends. Leveraging the unexpected interruption of the roll-out I test for counterfactual trends in locations where TV launches are unexpectedly and temporarily blocked. Moreover, I exploit the rise and decline of local filming for a powerful parallel trend check. The effects of local antennas should disappear when their importance for TV filming declines after the invention of the videotape. To exploit this variation, I control for the location of filming hubs by interacting a dummy for the time period of national production with a time-invariant and pre-determined proxy for local production cost. This variable will pick up incentives to move filming to a given place, while the measure avoids the potential endogenous control problem.

4.1 Results: The Effect on Top Earners

First, I focus on the main prediction of the superstar effect, the sharp wage growth at the top of the distribution (see Proposition 2.1a). I first test the response of the 99th percentile of entertainer wages by estimating equation 6. The data confirm the headline prediction of the Superstar Effect and shows that the launch of a TV station leads to an increase of wages at the top percentile by 17 log points (see panel A in Table 1). This is a large and highly significant effect which amounts to approximately a 19% wage increase. Next, I compare the sector to the wider US economy and check if the effect is specific to entertainment. Panel B in Table 1 shows that the share of entertainers in the top percentile of the US distribution increases by 4 percentage points and relative to the pre period the share nearly doubles. A potential drawback of this approach are spurious effects from fluctuations in the denominator. However, recall that we hold the denominator fixed here to address this challenge. Alternatively, we could address this problem by looking at per capita shares, which yields similar results (panel C of Table 1).

Next I analyze who benefits from the increase in top incomes. In the superstar model most sought after performers benefit most from an expansion in market reach. To test this, I build a small panel that matches local TV stars to the willingness to pay for these entertainers before TV. I use entertainer wages in 1939 to measure willingness to pay before TV and match willingness to pay in this pre-period to successful local TV entertainers who appear in the "Who is Who of Television." Figure 5 shows that TV indeed amplifies the most successful entertainers. The vast majority of TV stars were in the top tail of the entertainer wage distribution even before TV.

Probing the Identification Assumption

The main threat to identification are spurious demand shocks that are correlated with the TV roll-out. A range of robustness checks investigate this possibility. First notice that the roll-out has rigid rules that are likely unresponsive to local demand shocks, the 1952 "Final Allocation Report" for instance prioritizes locations by their local population in 1950. Once we condition on such pre-determined local characteristics, differences in treatment should arise quasi-randomly over time. To verify this I first 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 labor market specific trends (column 2 and 3 in Table 1). The second approach adds more than 700 additional location specific trends and thus is a very demanding specification, standard errors increase accordingly. Both specifications find effects very similar to the baseline, indicating that differential local trends are not driving the findings.

Next, I use the variation from the interruption of the roll-out to test for demand shocks that coincide with the launch of television. 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. Recall that the interruption was a blanket freeze that put all license procedures on hold. For identification this in-discriminatory approach is useful as it generates variation that is independent of local economic conditions and allows to test for spurious. Figure 6 shows the resulting sharp drop in approvals. I first implement a test that leverages the fact that we observe untreated results in places that ought to be treated. I use this to test for spurious effects of blocked stations ($TV_{mt}^{blocked}$) in a dynamic difference in differences regression:

$$Y_{mot} = \alpha_m + \delta_{ot} + \gamma X_{mt} + \sum_t \beta_t TV_{mt}^{blocked} + \epsilon_{mot} \tag{7}$$

The effect of such blocked stations (β_t) is depicted in Figure 7a and shows that they have no effect. The trends in blocked and untreated areas evolve in parallel and there is no sign of spurious demand shocks around the intended time of treatment. The TV roll-out process therefore appears to unfold orthogonally to local demand shocks, shoring up confidence that the identification assumption of the diff in diff holds. Related robustness checks with placebo occupations show the same result (Appendix D.1.3).

Finally, I test the common trends assumption directly. Such a test focuses on the treated areas themselves, rather than extrapolating from places that narrowly miss out on TV and leverages the decline of local filming after the invention of the videotape. If common-trends hold, the treatment effect arises when local TV productions are introduced and disappears when they are removed. As before, the test is implemented using a dynamic difference in differences regression, analogue to 7. Differences between treatment and control areas over time are plotted in Figure 7b and reveal the expected pattern. Differences in treatment and control areas appear during local TV filming and disappears after the end of local TV. In 1969 the differences between treatment and control group reverted to the pre-treatment level, which suggests that the common trend assumption holds.[23]

Links Between Markets

So far the analysis focused on changes in inequality, for the interpretation of the results it will be useful to distinguish between effects coming from mobility and from changing returns to talent.[24] I therefore study the mobility of entertainers across local

[23] A conventional pre-trend check is reported in Appendix D.1.2.

[24] Spillovers effects from trade in output do not apply here, as live entertainment shows are consumed where they are produced.

labor markets.[25] Note that labor markets where entertainers could reach the largest audiences before TV also tend to receive TV earlier. As a result, the the ranking of places in terms of audience reach remains largely unchanged and the roll-out of TV creates limited incentives to re-locate. I test the mobility response of entertainers empirically and indeed find quantitatively small effects. The point estimates are in fact negative and confidence intervals rule out that mobility increased by more than 2% (columns 1 – 3 of Table 2). Using these results to bound the impact of migration, we can rule out that migration explains more than a quarter of the total effect, while the central estimates suggest mobility plays next to no role for the results. A related test studies mobility across neighboring labor markets where moving is arguably easiest. Results that exclude neighboring areas are close to the baseline, indicating again that migration plays a minor role for the findings (panel B of Table 2). The data thus suggests that the change in inequality is mainly driven by shifts in labor market returns.

4.2 Results: Demand for Non-Stars

A second prediction of the Superstar Effect is the decline in employment (proposition 2.1). Such effects set the Superstar Effect apart from many conventional models of the labor market. To test for employment effects, I compare entertainer employment in local labor markets with differential access to TV signal. Employment records in the Census are available for additional years, which allows me to expand the sample period backward by a decade to 1930 and results are reported alongside results for the baseline period. For this extended period data on TV signal is not available at a channel level, instead I use a dummy for access to TV signal as regressor.

With the introduction of TV employment in local entertainment contracted substantially and around 13% of jobs disappear (Table 3 column 1, panel A for the extended sample and panel B for the baseline sample). This aligns with the Superstar Effect, where labor market moves closer to a winner-takes-all market but is sharply at odds with models where technical change causes a positive demand shock, which would raise employment.

Since these specifications use variation from TV signal rather than from TV

[25]It might seem appealing to use panel data for this test. While such data is unavailable here, it would not identify the two channels. In the assignment model, migration affects the assignment of workers to jobs and an individual fixed effect regression would conflate such changes in worker-firm matches with changes in productivity.

filming, it is salient to probe the identifying assumption again. As before, results are robust to the inclusion of controls and local trends (Columns 2 and 3 of Table 3).[26] Further common trend tests suggest that the set-up is valid. A first pre-trend test introduces a lead to the treatment variable in the difference in differences regression, which captures differential changes in treatment and control areas, right before the treatment and shows no sign of such differential trends (Column 4).[27] Placebo tests with stations that did not happen due to the freeze also show no effect (Panel C of Table 3), adding confidence that there are no spurious trends during the roll-out.

The shift of demand from a profession's mediocre workers towards its stars should also be reflected in a decline in mid-paid jobs (see Proposition 2.1c). To test for such effects, consider entertainers who are below the top 90th percentile of the US wage distribution but still in the upper quartile. These are entertainers who receive above-average pay but are far from the top of the entertainer distribution. I test how TV affects this group and find significantly negative effect. 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 8). Television therefore leads to a substantial decline in well-paid jobs and makes it 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 second quartile. Television thus reduces the payoff of non-star talent and creates a growing low-pay sector.

4.3 Results: Fractal Inequality

A further implication of the Superstar Effect are widening wage difference between stars and their slightly less talented peers (see Proposition 2.1b). A non-parametric test of this prediction repeats the baseline difference in differences regression, focusing on percentiles just below the star level. Take, for example, entertainers who are below the top 1% but still among the top 5% of the US wage distribution. I find that television also benefits this group but the effect is only one tenth the size of the

[26]Median income is missing in 1930 and controls in the extended sample use the remaining variables

[27]TV signal, unlike local filming, is not removed and we thus cannot rely on pre- and post-periods to identify counterfactual trends.

effect at the very top. Television therefore disproportionally benefits the superstars and widens the pay gap in the top tail of the distribution. To confirm this pattern we can look at the next lower wage bin, between the 90th and 95th percentile. Here the effect of television is insignificant, again confirming that television's effect fades quickly as we move away from the top stars in the market. 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.

The growing fractal inequality is also reflected in growing top income dispersion within entertainment. This is closely related to the previous results but focuses on an inequality measures widely used in the literature, top income shares.[28] Prior to TV, the fraction of income going to the 1% highest earners in a local labor market was, on average, 3.8%.[29] TV filming increased the top income share by 3.7 percentage points, and thus nearly doubles the income share (Table 4). Proposition 2.1b suggests that the growth in these shares should escalate toward the top of the distribution. And indeed most of the gains in the top 1% accrued to the very highest earners. 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. While the share of income going to the top 1% doubled and the equivalent share for the top 0.1% grew 4 fold, the top 10% share grew only 30%. A test of equal growth rates in the top tail is strongly rejected, which aligns with the Superstar Effect where wage growth is strongest at the very top of the distribution.[30]

Looking at the effects across the entire distribution of entertainer pay, the data show a U-shaped impact of TV (Figure 8). At the very top of the distribution TV has a large positive impact but such positive effects decline quickly as we move away from the very top, turning negative below the 90th percentile. At the same time we see a growing low-paid sector in the industry. This characteristic pattern offers direct empirical support for the Superstar Effect.

[28] See for example Piketty and Saez, 2003; Piketty, 2014.

[29] 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.

[30] The appendix confirms the results with a set of quantile regressions (see Appendix D.1.4).

5 Magnitude of Superstar Effects

Finally, I estimate the magnitude of Superstar Effects. This moves beyond the reduced form results and quantifies the elasticity of top pay to market size. Previous work has used this elasticity to calibrate the key structural parameters of the superstar model (e.g. Gabaix and Landier, 2008; Terviö, 2008). Such studies use the correlation of top pay and market size, proxied by firm value, to calibrate the elasticity. An attractive feature of the entertainment setting is that it offers a direct measure of workers' market reach, the entertainers' audience. Moreover, audience reach varies for plausibly exogenous reasons during the TV roll-out. This allows me to supplement correlational estimates with an instrumental variables approach. Such estimates are based on the following regression equation:

$$ln(w^{99}_{m,t}) = \alpha_0 + \alpha_1 ln(s^{99}_{m,t}) + \epsilon^{99}_{m,t} \tag{8}$$

where $w^{99}_{m,t}$ is the 99th percentile of the entertainer wage distribution in market m and year t and $s^{99}_{m,t}$ the size of the market that such entertainers can reach.

First consider how TV affected the audiences of entertainers. For stars, this is equivalent to the first stage of an IV regression of equation 8, where TV is the instrument. However, the distribution of audience is also of interest in itself. According to the Superstar Effect scalable production leads to shift of audiences from more mediocre workers to the stars of the profession. This shift in audience is the mechanism that underpins the wage changes of the Superstar Effect. To study such audience shifts, I first look at log audience of shows of local stars. The launch of a television station increased the audience of the largest shows by about 150 log points, which converted to a growth rate implies a growth of over 300%, or a fourfold increase in market size (Panel A of Table 5). Additionally, I quantify the change in market reach in dollar terms and thus estimate the change in marginal revenue product that went hand in hand with the growth in audience size. In dollar terms market reach of stars roughly tripled (Panel B of Table 5). The launch of a TV station thus dramatically increased the market value of top talent.

Second, I estimate the effect of TV on ordinary local live entertainment. For this test, I study data on county fairs, which take place in all parts of the US each year. TV leads to a 5% decline in audiences and spending at local county fairs (column 1 and 2 of Table 6). These results are, noisy as they hide substantial heterogeneity across types of entertainment. Splitting the results by types of entertainment that are differently close substitutes for TV the data show very different effects. Demand for entertainment that is similar to TV, such as grandstand shows, falls significantly, while demand for entertainment that is very different from TV, e.g. candy sales and amusement rides, holds up (see columns 3 and 4 of Table 6 and panel B for regressions at the county level). This shows that even within the entertainment spending category, close substitutes to TV are most affected by the availability of TV signal. These result confirm that TV reduced demand for local live entertainment, increased demand for star entertainers and shifted marginal revenue productivity in favor of stars at the expense of ordinary entertainment.

Next, I turn to the elasticity of top pay to audience size. First, I estimate equation 8 with a naive OLS estimator on a single cross-section, using variation in the size of the biggest local theatre size in 1939 as regressor. Similar regressions have been estimated in a large literature that studies the relation of pay and firm size. In line with those results my cross-sectional OLS estimate of α_1 is highly significant with a point estimate of 0.23 (see panel A. of Table 7). 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 local characteristics, the effect disappears almost entirely (column 2 of the same Table).[31]

I compare the OLS estimate with an IV estimate that uses the roll-out of TV as instrument for audience size. The first stage of TV on audience size is highly significant, as we saw above. The associated first stage F-statistic is around 20, well above conventional cutoffs. The IV estimator of the elasticity α_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.

In most settings revenue is used to proxy market size, instead of the underlying shift in market reach. To link my results to this literature, I run the same estimation using data on revenue. A drawback of this approach is that it conflates quantity and endogenous price effects and estimates the combined reduced form effect. Such

[31] The panel OLS estimate would compare wages across local labor markets 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 and hence such a panel OLS is therefore mechanically close to the IV estimate.

estimates with log revenue as market size proxy also shows a highly significant superstar elasticity, 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.[32] This falls in the range estimated among CEOs, which varies between 0.1 and 1 (Gabaix and Landier, 2008; Frydman and Saks, 2010).

Finally, I consider how Superstar Effects vary across different settings. First, recall that two ingredients are required for Superstar Effects: scalable production and heterogeneous and unique types of talent. In this paper I study a setting that features unique talents and test how variation in scalable production affects inequality. By contrast, we would expect smaller or no superstar effects in markets where workers are closer substitutes. The results here therefore provide an upper bound for Superstar Effects in the economy at large. Moreover, differences in labor market institutions may affect the magnitude of Superstar Effects. This is particularly poignant as similar production technologies appear to create very different levels of top income inequality across Western countries (Piketty and Saez, 2006). A foremost concern among policy makers is how the Superstar Effect interacts with imperfect competition.[33] The benchmark superstar model is perfectly competitive and growing top incomes are the result of changing demand for talent. The models' predictions change sharply with imperfect labor market competition, as employers with market power will not pass-on the surplus from greater scalability of production. To test the effect of imperfectly competitive labor market I exploit entry restrictions during the TV roll-out and allow for differential effects of TV in markets with a single TV station and markets with multiple TV stations. The differences between monopsonistic and competitive labor markets are striking. Markets with a monopsony employer see almost no top income growth, while gains are large when there are competing employers. These results are confirmed when I narrow in on the variation from the roll-out interruption (Table 8). The results emphasize the importance of competition in driving top income growth. Only when employers face competition, does greater market scale translate into rising wages.

[32] 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.

[33] Appendix D.2 explores the impact of tax wedges and education levels and finds that such institutional features had limited effect on top pay growth.

6 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 Effect.

To test the Superstar Effect, 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 and the sector experiences sharp income concentration at the top. In line with the prediction of the Superstar Effect the increase in production scalability has profound effects on inequality at both the top and bottom of the distribution. The characteristic patterns of the Superstar Effect is strongly supported in the data. Income growth escalates as we move up towards the top of the wage distribution and the share of income going to the top 1% nearly doubles. Moreover, the ability to reach larger markets puts many mediocre workers out of work. The number of mid-paid entertainer jobs declined significantly and total employment fell about 13%.

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. In comparing these results with other settings, it is important that imperfect substitutability is a crucial ingredient of the Superstar Effect and may not apply across all industries. The iconic entertainment case is thus particularly suitable to provide a first identified estimate of Superstar Effects and may provide an likely provides an upper bound for Superstar Effects in the wider economy. Finally, the paper shows that a key driver of Superstar Effects is competition. Top income growth is muted substantially in settings with limited competition in the labor market.

The Superstar Effect suggests that rising market concentration is a sign of technical progress. Conclusions that market concentration indicates malfunctioning markets might therefore be premature. 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

Acemoglu, Daron (1998). "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality." In: *The Quarterly Journal of Economics* 113.4, pp. 1055–1089.

Acemoglu, Daron and David H. Autor (2011). "Skills, Tasks and Technologies: Implications for Employment and Earnings." In: *Handbook of Labor Economics* 4.PART B, pp. 1043–1171.

Akerman, Anders, Ingvil Gaarder, and Magne Mogstad (2013). "The Skill Complementarity of Broadband Internet." In: *Quarterly Journal of Economics* 7762.

Alvaredo, F. et al. (2018). *The World Inequality Database.*

Angelucci, Charles, Julia Cagé, and Michael Sinkinson (2019). "Media Competition and News Diets." In: *wip.*

Autor, David H. (2014). "Skills, education, and the rise of earnings inequality among the "other 99 percent"." In: *Science* 344.6186, pp. 843–851.

Autor, David H. and David Dorn (2013). "The Growth of Low Skill Service Jobs and the Polarization of the U.S. Labor Market." In: *American Economic Review* 103.5, pp. 1553–97.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney (2008). "Trends in U.S. Wage Inequality: Revising the Revisionists." In: *Review of Economics and Statistics* 90.May, pp. 300–323.

Autor, David et al. (2017). "The Fall of the Labor Share and the Rise of Superstar Firms." In: *NBER Working Paper.*

Bakija, Jon, Adam Cole, and Bradely Heim (2012). "Jobs and Income Growth of Top Earners and the Cause of Changing Income Inequality: Evidence from US Tax Return Data." In: *unpublished manuscript, Williams College.*

Bartel, Ann, Casey Ichniowski, and Kathryn Shaw (2007). "How Does Information Technology Affect Productivity? Plant-Level Comparisons of Product Innovation, Process Improvement and Worker Skills." In: *The Quarterly Journal of Economics* 122.November, pp. 1721–1758.

Bertrand, Marianne and Sendhil Mullainathan (2001). "Are CEOS Rewarded for Luck? The Ones without Principals Are." In: *Quarterly Journal of Economics* 116.3, pp. 901–932.

Blundell, Richard, Rachel Griffith, and John Van Reenen (1999). “Market value, and innovation in a panel of British manufacturing firm.” In: *Review of Economic Studies* 66.3, pp. 529–554.

Card, David and John DiNardo (2002). “Skill Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles.” In: *Journal of Labor Economics* 20.4.

Chetty, Raj et al. (2014). “Where is the land of opportunity? The geography of intergenerational mobility in the United States.” In: *Quarterly Journal of Economics* 129.4, pp. 1553–1623.

Chetverikov, Denis, Bradley Larsen, and Christopher Palmer (2016). “IV Quantile Regression for Group-Level Treatments, With an Application to the Distributional Effects of Trade.” In: *Econometrica* 84.2, pp. 809–833.

Cook, Philip J. and Robert Frank (1995). *The Winner-Take-All Society*, p. 288.

Edmans, Alex and Xavier Gabaix (2016). “Executive Compensation.” In: *Journal of Economic Literature* 54, p. 150.

Fenton, George and Felix Koenig (2018). “The Labor Supply Response to Innovation in Entertainment: Evidence from TV.” In: *Working paper*.

Frydman, Carola and Raven E. Saks (2010). “Executive compensation: A new view from a long-term perspective, 1936-2005.” In: *Review of Financial Studies* 23.5, pp. 2099–2138.

Gabaix, Xavier and Augustin Landier (2008). “Why Has CEO Pay Increased So Much?” In: *Quarterly Journal of Economics* 123.1, pp. 49–100.

Garicano, Luis and Thomas N. Hubbard (2009). “Specialization, firms, and markets: The division of labor within and between law firms.” In: *Journal of Law, Economics, and Organization* 25.2, pp. 339–371.

Gentzkow, Matthew (2006). “Television and voter turnout.” In: *The Quarterly Journal of Economics* 121.August, pp. 931–971.

Gentzkow, Matthew and Jesse Shapiro (2008). “Preschool Television Viewing and Adolescent Test Scores: Historical Evidence from the Coleman Study.” In: *Quarterly Journal of Economics* 123.February, pp. 279–323.

Guellec, Dominique and Caroline Paunov (2017). “Digital Innovation and the Distribution of Income.” In:

Kaplan, Steven N. and Joshua Rauh (2013). “It’s the Market: The Broad-Based Rise in the Return to Top Talent.” In: *Journal of Economic Perspectives* 27.3, pp. 35–56.

Katz, Lawrence F. and David H. Autor (1999). “Changes in the Wage Structure and Earning Inequality.” In: *Handbook of Labor Economics* 3, pp. 1463–1555.

Katz, Lawrence F. and Kevin M. Murphy (1992). “Changes in Relative Wages, 1963-1987: Supply and Demand Factors.” In: *The Quarterly Journal of Economics* 107.1, pp. 35–78.

Kleven, Henrik Jacobsen et al. (2013). “Taxation and International Migration of Superstars: Evidence from European Football Market.” In: *American Economic Review* 103.5, pp. 1892–1924.

— (2014). “Taxation and International Migration of Top Earners: Evidence from the Foreigner Tax Scheme in Denmark.” In: *The Quarterly Journal of Economics* 129.1, pp. 333–378.

Krueger, Alan B. (2005). “The Economics of Real Superstars: The Market for Rock Concerts in the Material World.” In: *Journal of Labor Economics* 23.1, pp. 1–30.

Lemieux, Thomas (2006). “Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?” In: *American Economic Review* 96.3, pp. 461–498.

— (2008). “The Changing Nature of Wage Inequality.” In: *Journal of Population Economics* 21.1, pp. 21–48.

Malmendier, Ulrike and Geoffrey Tate (2008). “Superstar CEOs.” In: *Quarterly Journal of Economics* 124.4, pp. 1593–1638.

Manson, Steven et al. (2017). *IPUMS NHGIS: Version 12.0*. Tech. rep. Minneapolis: University of Minnesota.

Michaels, Guy and Georg Graetz (2018). “Robots at Work.” In: *Review of Economics and Statistics*.

Michaels, Guy, Ashwini Natraj, and John Van Reenen (2014). “Has ICT polarized skill demand? Evidence from eleven countries over 25 years.” In: *The Review of Economics and Statistics* 96, pp. 60–77.

Moretti, Enrico and Daniel Wilson (2017). “The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists.” In: *American Economic Review* 107.7, pp. 1858–1903.

OECD (2016). *The Productivity-Inclusiveness Nexus*. Tech. rep. June.

Piketty, Thomas (2014). *Capital in the 21st Century*. Harvard University Press.

Piketty, Thomas and Emmanuel Saez (2003). “Income Inequality in the United States, 1913-1998.” In: *The Quarterly Journal of Economics* 68.February, pp. 1–39.

Piketty, Thomas and Emmanuel Saez (2006). “The Evolution of Top Incomes: A Historical and International Perspective.” In: *American Economic Review (P&P)* 96.2, pp. 200–205.

Rosen, Sherwin (1981). “The Economics of Superstars.” In: *American Economic Review* 71.5, pp. 845–858.

Sattinger, Michael (1975). “Comparative advantage and the distributions of earnings and abilities.” In: *Econometrica* 43.3, pp. 455–468.

— (1979). “Differential Rents and the Distribution of Earnings.” In: *Oxford Economic Papers* 31.1, pp. 60–71.

— (1993). “Assignment Models of the Distribution of Earnings.” In: *Journal of Economic Literature* 31.2, pp. 831–880.

Scheuer, Florian and Iván Werning (2017). “The Taxation of Superstars.” In: *Quarterly Journal of Economics* 132(1), pp. 211–270.

Song, Jae et al. (2019). “Firming Up Inequality.” In: *The Quarterly Journal of Economics* 1354.

Sterne, Jonathan (1999). “Television under construction: American television and the problem of distribution, 1926-62.” In: *Media, Culture & Society* 21.199905, pp. 503–530.

Stigler, G. J. (1941). *Production and Distribution Theories*. New York: Macmillan.

Terviö, Marko (2008). “The difference that CEOs make: An assignment model approach.” In: *American Economic Review* 98.3, pp. 642–668.

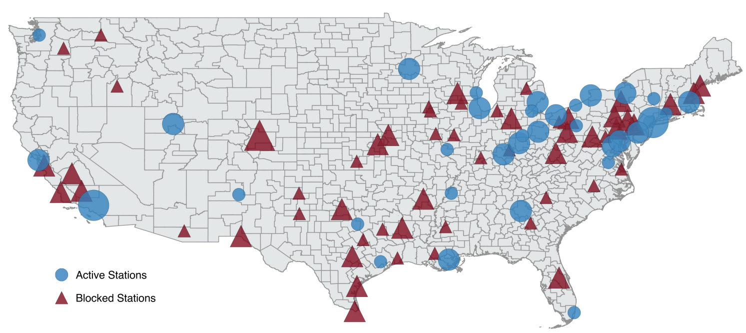
Tinbergen, Jan (1956). “On the Theory of Income Distribution.” In: *Weltwirtschaftliches Archiv* 77.2, pp. 155–175.

A Figures

Figure 1: Superstar Effect: Employment Change at Different Wage Levels

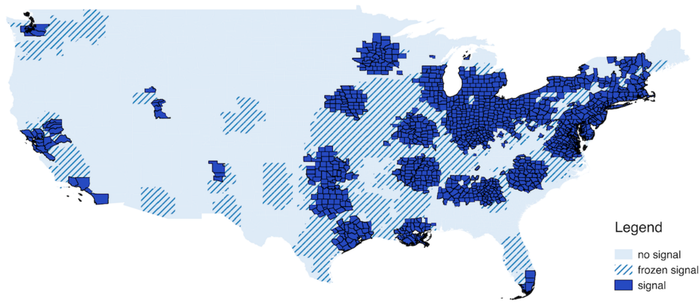
[Notes] The figure shows the Superstar Effect on employment changes across the wage distribution. The figure is based on equation 5 for parameterization $g_e = 0.2x^{(1.3)} - 1$. The figure reports growth rates across the full distribution by grouping job growth outside the range of previous support with the final bins with positive mass and thus avoids dividing by zero.

Figure 2: TV Filming of Licensed and Blocked Stations in 1949



[Notes] Symbols show the location of TV filming and the size of a symbol indicates the number of TV stations per local labor market. Active stations are blue circles, frozen stations red triangles. Source: FCC reports.

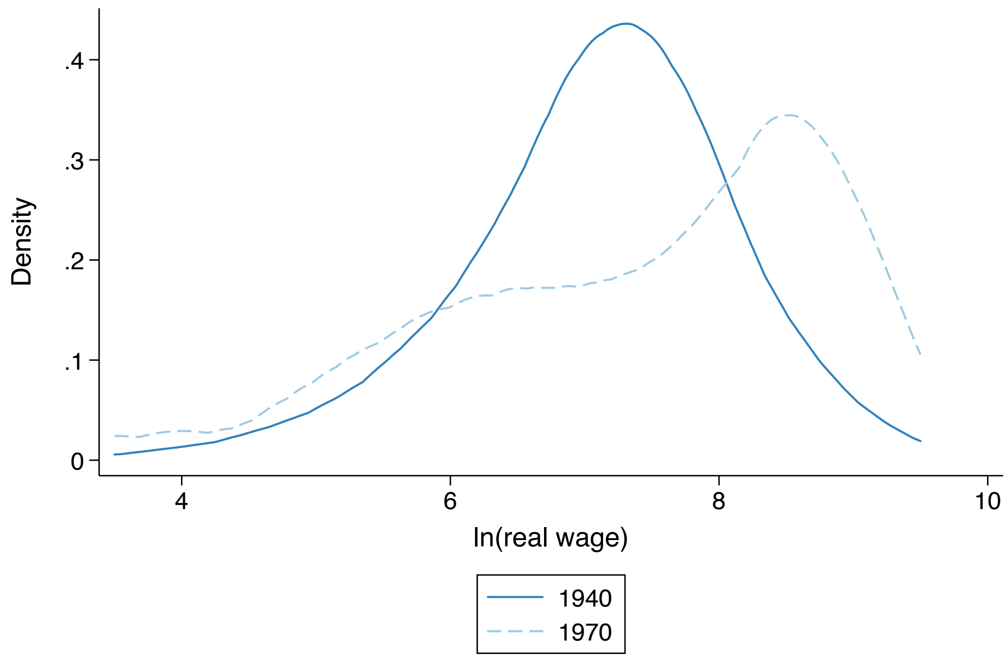
Figure 3: TV Signal of Licensed and Blocked Stations in 1949



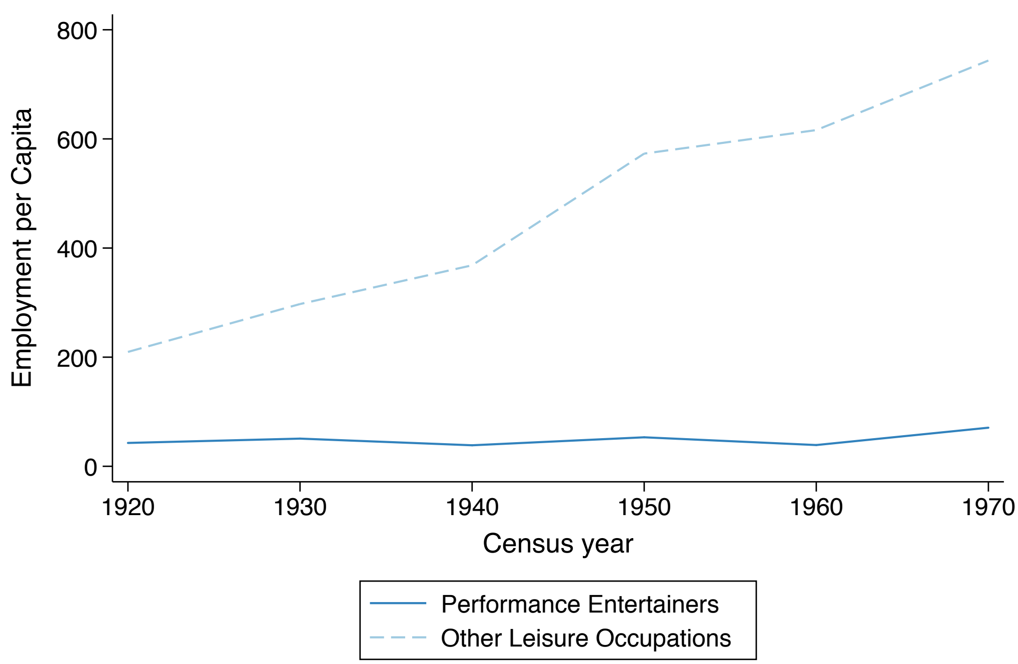
[Note] Areas in dark blue can watch TV, while shaded areas would have had TV signal from blocked TV stations. Signal coverage is calculated using an Irregular Terrain Model (ITM). Technical station data from FCC files, as reported in TV Digest and Television yearbooks, are fed into the model . Signal is defined by a signal threshold of -50 of coverage at 90% of the time at 90% of receivers at the county centroid. Source: Fenton and Koenig, 2018.

Figure 4: Change in Entertainment 1940 – 1970

(a) Entertainer Wage Distribution

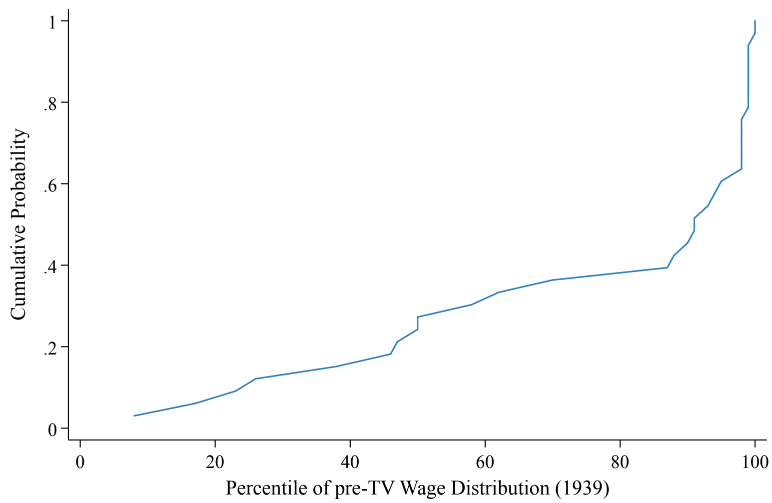


(b) Entertainer per Capita



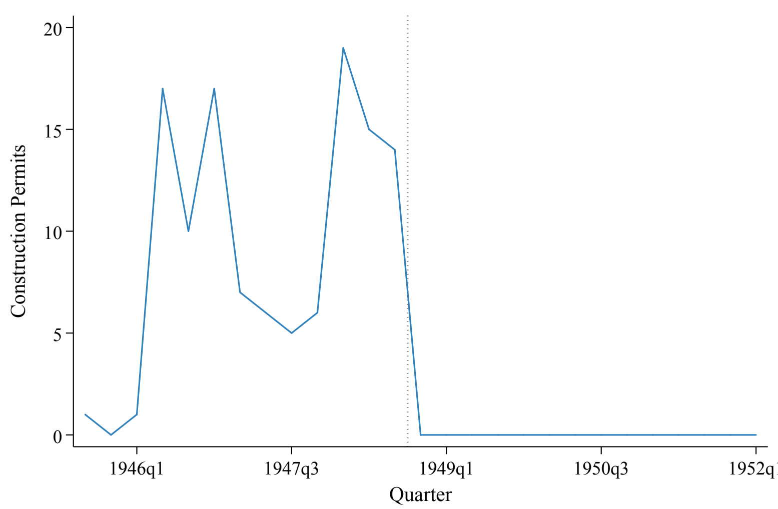
[Notes] Panel A shows the entertainemnet log real wage distribution in 1940 and 1970 from the lower 48 states. Dollar values are in 1950 USD. Density is estimated using the Epanechnikov smoothing kernel with a bandwidth of 0.4 and Census sample weights. Common top code applied at $85,000. Panel B shows employment per 100,000 inhabitants of performance entertainers (defined in text) and other leisure related occupations (drink & dine and "other entertainment occupations"). The mean for performance entertainers is 49 and for other leisure occupations 468. Sources: US Census 1940, 1970.

Figure 5: Position of Future TV Stars in the 1939 US Wage Distribution



[Note] The Figure shows the CDF of wage distribution ranks of TV stars before they became TV stars. TV stars are defined in the 1950 "Who is Who of TV". These individuals are linked to their 1939 Census wage records. 1939 wages are corrected for age, education and gender using a regression of log wages on a cubic in age, 12 education dummies and a gender indicator. Source: Radio Annual, Television Yearbook 1950.

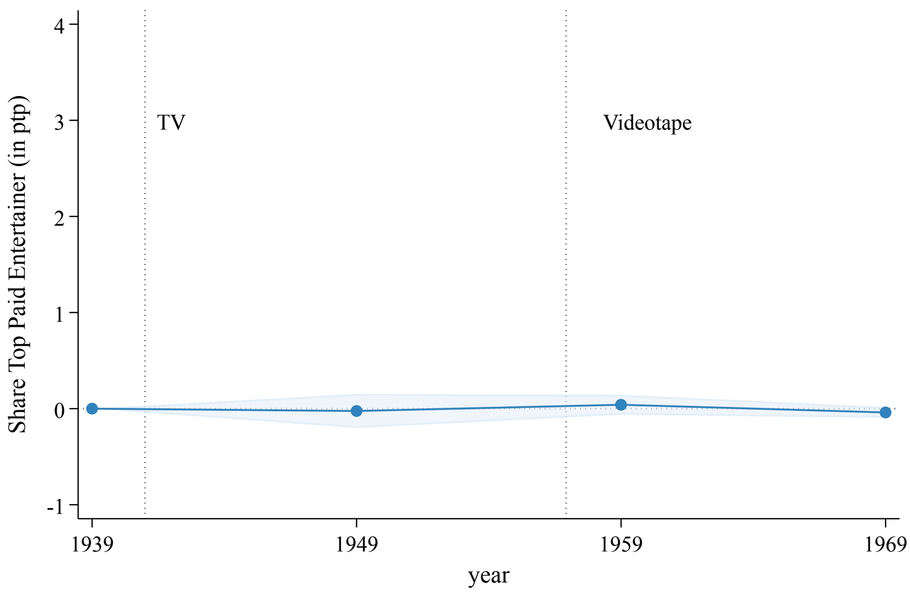
Figure 6: Number of TV Licenses Granted



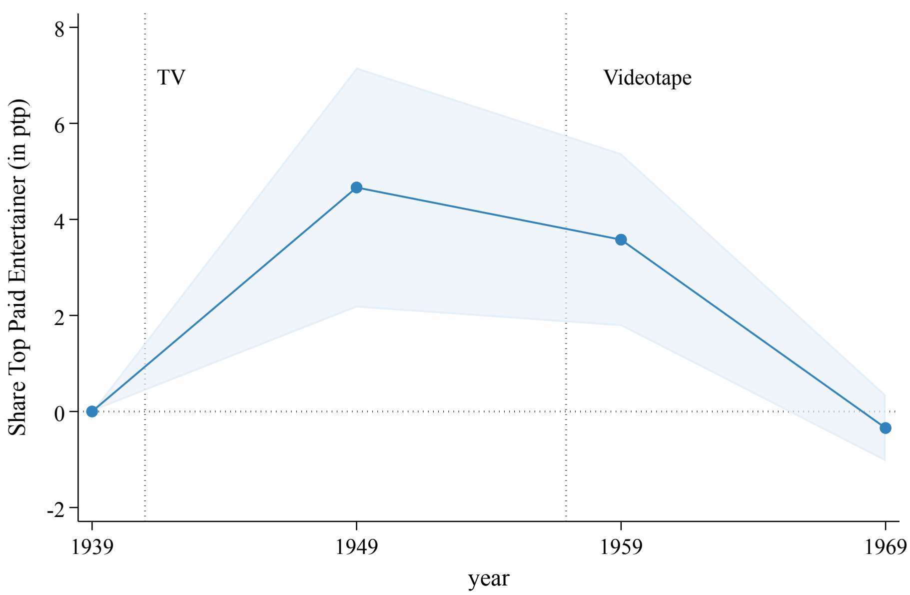
[Note] Missing issue dates of construction permits are inferred from start of operation dates. Source: TV Digest reports.

Figure 7: Dynamic Treatment Effect of TV on

(a) *Blocked TV Stations*

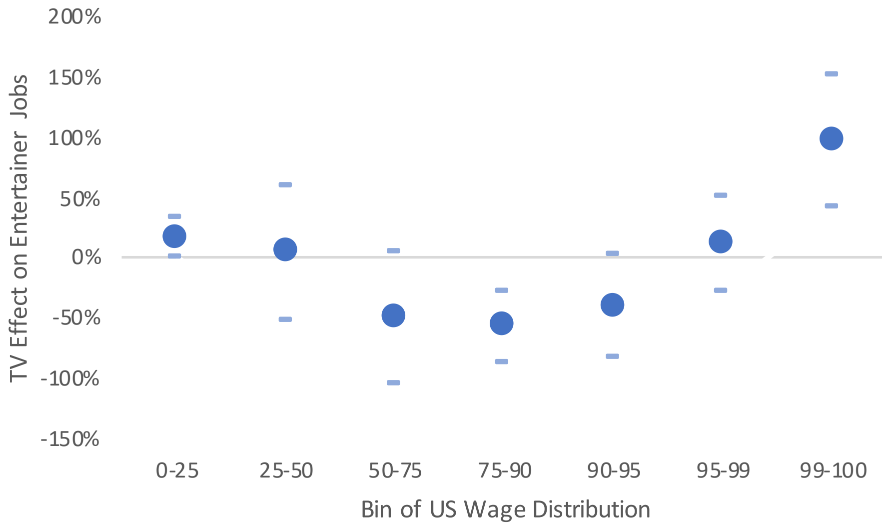


(b) *Active TV stations*



[Note] Figure plots treatment coefficients from two dynamic difference in differences regressions. Panel a) shows the coefficient on $FrozenTV_{m,t}$ (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.

Figure 8: Effect of TV on Entertainer Employment Growth at Different Wage Levels



[Note] Each dot is the treatment effect estimate of a separate DiD regression. It shows a TV station's effect on entertainer jobs at different parts of the wage distribution. Percentile bins are defined in the overall US wage distribution. Dashes indicate 95% confidence intervals. See table 1 for details on the specification. Sources: US Census: 1940-1970.

B Tables

Table 1: Effect of TV on Entertainer Top Earners

	Panel A: *Ln(99th Percentile of Entertainer Wages)*		
Local TV stations	0.171	0.157	0.149
	(0.029)	(0.066)	(0.093)
Effect/Baseline	18.6%	16.9%	16.1%
Cluster	702	702	702
	Panel B: Entertainer among Top 1% of US Earners (% of Entertainers)		
Local TV Stations	4.14	4.31	5.93
	(1.26)	(1.27)	(2.21)
Effect/Baseline	92%	96%	132%
Cluster	722	722	722
	Panel C: Entertainer among Top 1% of US Earners (Per Capita)		
Local TV Stations	0.40	0.40	0.31
	(0.10)	(0.10)	(0.10)
Effect/Baseline	133%	133%	103%
Cluster	722	722	722
Time & Labor Market FE	Yes	Yes	Yes
Demographics	–	Yes	–
Local labor market trends	–	–	Yes

[Note] Outcomes: Panel A: The entertainer wage at the 99th percentile, Panel B: share top-paid entertainers, Panel C: top-paid entertainer per capita in 10,000s. Specifications: Each cell is the result of a separate DiD regression on the number of TV stations in the local area. All regressions control for CZ & time fixed effects and local filming cost in years after the invention of the videotape . Demographics: median age & income, % female, % black, population density and trends for urban areas; local labor market trends: allow for a linear trend for each local labor market. Entertainers are actors, athletes, dancers, entertainers nec, musicians. Panel A uses the quantile DiD estimator developed by Chetverikov, Larsen, and Palmer, 2016; cells where the 99th percentile cannot be computed are dropped. The unit of analysis is the CZ – year level in A and the the more disaggregated CZ – occupation – year level in B and C to additionally control for year-occupation fixed effects. Panel A uses 2,264 observations and Panel B and C 13,718 observations, demographic data is missing for one CZ in 1940. "Effect/Baseline" reports treatment effects relative to the baseline value of the outcome variable. Observations are weighted by local labor market population. Standard errors are reported in brackets and are clustered at the local labor market level. Sources: US Census 1940-1970.

Table 2: Effect of TV on Mobility Between Labor Markets

	Panel A: Share Entertainers who Migrated		
Local TV stations	-0.014	-0.017	-0.010
	(0.015)	(0.015)	(0.020)
	Panel B: Entertainer among Top 1% of US Earners (excl. neighbor)		
Local TV stations	4.30	4.46	6.16
	(1.31)	(1.30)	(2.27)
Time-Occupation & Labor Market FE	Yes	Yes	Yes
Demographics	–	Yes	–
Local labor market trends	–	–	Yes

[Note] Dependent variables are, Panel A the fraction of entertainers who moved, Panel B share of Entertainers among the top 1% of the US wage distribution, excluding labor markets that neighbor treated labor markets. Specification details are as in Table 1, except that Panel B is run on a reduced sample of 10,792 observations. Source: US Census 1940-1970.

Table 3: Effect of TV on Entertainer Employment

	Ln(Employment in Entertainment)			
	(1)	(2)	(3)	(4)
	Panel A: TV Signal 1930-1970			
TV signal$_{t+1}$				0.039
				(0.033)
TV signal$_t$	-0.133	-0.127	-0.125	-0.123
	(0.059)	(0.059)	(0.061)	(0.060)
	Panel B: TV Signal 1940-1970			
TV signal$_t$	-0.128	-0.114	-0.134	
	(0.061)	(0.061)	(0.063)	
	Panel C: Placebo TV Signal			
Placebo TV signal$_t$	0.053	0.044	0.053	
	(0.083)	(0.083)	(0.084)	
Clusters	722	722	722	722
Time-Occupation & Labor Market FE	Yes	Yes	Yes	Yes
Demographics	-	Yes	-	-
Local Labor Market Trends	-	-	Yes	-

[Note] Dependent variable "ln(Employment in Entertainment)" is the inverse hyperbolic sine of employment in entertainment. Control variables and specifications are as described in Table 1, except that demographic controls exclude median income. TV signal is a dummy that takes value 1 if signal is available in a commuting zone. Placebo TV signal is the signal of stations that were blocked. Subscript "t+1" refers to the lead of the treatment. Standard errors, reported in brackets, are clustered at the local labor market level. Source: TV signal from Fenton and Koenig, 2018 and labor market data from US Census 1930-1970.

Table 4: Effect of TV on Top Income Shares in Entertainment

	Share of Income		
	Top 0.1%	Top 1%	Top 10%
Local TV stations	2.37	3.71	6.08
	(1.27)	(1.69)	(2.12)
Time & Labor Market FE	Yes	Yes	Yes
Effect/Baseline	239%	96%	33%
P-value: same growth as top 1% share	0.0043	—	0.0000

[Note] Dependent variable top p% is the share of income going to the top p percent of entertainers in a given local labor market-year. The shares are calculated using Pareto interpolation as described in the text. The sample includes the larger 350 labor markets and 1,069 observations. Estimates are based on a difference in difference specification. P-values from a test of equal growth rates in top income shares are also reported. This test is implemented in a regression with the ratio of top income shares as dependent variable. Standard errors are clustered at the local labor market level. Sources: US Census 1940-1970.

Table 5: Effect of TV on Market Reach of Local Stars

	(1)	(2)	(3)
		Panel A: Ln(Show Audience)	
Local TV stations	1.499	1.526	1.146
	(0.240)	(0.223)	(0.220)
Effect/Baseline	348%	360%	215%
		Panel B: Ln(Show Revenue)	
Local TV stations	1.095	1.116	1.146
	(0.207)	(0.168)	(0.220)
Effect/Baseline	199%	205%	215%
Clusters	722	722	722
Time & Labor Market FE	Yes	Yes	Yes
Demographics	–	Yes	-
Local labor market trends	–	–	Yes

[Note] Dependent variables are, Panel A: potential show audience of the largest show in the commuting zone, computed from venue seating capacity and TV households in transmission area, Panel B: potential revenue of largest show. Cells report results from separate DiD regressions across local labor markets. Control variables are as described in Table 1. The total number of CZ - year observations are 2,656. Sources: See text.

Table 6: Effect of TV on Log Spending at Local County Fairs

	(1) Ln(Fair Visits)	(2) Ln(Ticket Receipts)	(3) Ln(Show Receipts)	(4) Ln(Carnival Receipts)
	Panel A: Local Labor Market Level			
TV signal	-0.051	-0.047	-0.059	0.014
	(0.031)	(0.024)	(0.022)	(0.022)
Clusters	722	722	722	722
Time & Labor Market FE	Yes	Yes	Yes	Yes
	Panel B: County Level			
TV signal	-0.013	-0.014	-0.018	0.001
	(0.010)	(0.007)	(0.007)	(0.006)
Clusters	3,111	3,111	3,111	3,111
Time & County FE	Yes	Yes	Yes	Yes

[Note] Dependent variables are summed across county fairs in location m in year t at annual frequency from 1946 to 1957. All variables use the the inverse hyperbolic sine transformation to approximate the log function, while preserving 0s and monetary variables are in 1945 US Dollars. In Panel A the unit of observation m is a local labor market and in Panel B a county. 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 uses 8,664 local labor market observations (7,220 in column 4), while Panel B uses 37,332 county observations (in col 4 31,110). Standard errors, reported in brackets, 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 Fenton and Koenig, 2018.

Table 7: Elasticity of Entertainer Top Pay to Market Reach

	(1)	(2)	(3)
	$Ln(99^{th}\ Percentile\ of\ Entertainer\ Wages)$		
	Panel A: OLS - Cross-section 1939		
ln(Audience size)	0.234	0.023	
	(0.036)	(0.036)	
	Panel B: IV		
ln(Audience size)	0.166	0.149	0.149
	(0.017)	(0.019)	(0.024)
First-stage F-statistic	33.3	25.7	20.0
	Panel C: IV		
ln(Value of market ($))	0.220	0.192	0.198
	(0.028)	(0.022)	(0.036)
First-stage F-statistic	57.10	38.1	28.7
Demographics	–	Yes	–
Local labor market trends	–	–	Yes

[Note] Dependent variable is the entertainer wage at the 99th percentile. Panel A reports coefficients from a cross-sectional regression that uses variation across 573 local labor markets in 1939. Panel B and C show results from an IV regression that uses TV stations as instrument and uses the full panel with 2,148 observations. The corresponding first stage and reduced form results are reported in table 1 and table 5. The first-stage F-statistic is the Kleibergen-Paap F-statistic that allows for non-iid standard errors. Control variables are described in table 1 and market reach measures in table 5. Standard errors are clustered at the local labor market level. Sources: see table 1 and table 5.

Table 8: Effect of Monopsony Power in Labor Markets

	(1)	(2)	(3)
	Entertainer in US Top 1%		
Local TV station (dummy)	5.90	0.753	-0.57
	(3.06)	(1.91)	(0.36)
Multiple local TV station (dummy)		9.07	10.37
		(4.99)	(4.70)
Frozen competitor			1.43
			(2.10)
Clusters	722	722	722
Time-Occupation & Labor Market FE	Yes	Yes	Yes

[Note] Sources and specification as in baseline.

C ONLINE APPENDIX: Extensions

C.1 Equilibrium of the Superstar Model

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. Hence, in the optimization of manager i we can write the production function $Y(S,t)$ as $Y_i(p)$. The firm problem is therefore given by:

$$max_p Y_i(p) - w(p)$$

where $w(p)$ is the wage for a worker at percentile p of the talent distribution. Also, I extend the model and allow for elastic labor supply. This gives rise to a fourth equilibrium object, the participation threshold $\bar{p}$. Condition 1 is a consequence of the single crossing condition $Y_{st} > 0$. For a proof see for example **Sattinger1975** To derive condition 2 I start from the fact that the equilibrium is incentive compatible. Incentive compatibility guarantees that for each firm i the optimal worker p meets:

$$Y_i(p) - w(p) \geq Y_i(p') - w(p') \quad \forall \; p' \epsilon [0,1] \tag{9}$$

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 9 by ϵ and let $\epsilon \rightarrow 0$.

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

$$w'(p) = Y_p(S_i, p) \tag{10}$$

The IC condition can thus be written as a condition on the slope of the wage schedule and proves result 2.

Participation constraints (PC) define the participation threshold $\bar{p}$. They guarantee that both firms and workers are staying in the industry. Denote the reservation wage of workers $w^{res}$and the reservation profits ψ^{res} and hence the PC condition is:

$$Y_i(p) - w(p) \geq \psi^{res} \quad \forall \ p\epsilon[\bar{p}, 1] \tag{11}$$

$$w(p) \geq w^{res} \quad \forall \ p\epsilon[\bar{p}, 1] \tag{12}$$

The marginal participant is indifferent between participating and hence the PC binds with equality: $w(\bar{p}) = w^{res}$ and $Y_i(\bar{p}) - w(\bar{p}) = \psi^{res}$. Individuals with lower levels of skill will work in an outside market where pay is independent of talent and given by w^{res}.

Summing over all firms, we can derive the total revenue in the economy: $S(\pi) = \int_{\bar{p}} -Y(\sigma(\hat{t}), \hat{t})h'(\hat{t})d\hat{t}$. In equilibrium revenues equal total expenditure, denoted by $D(\pi)$, which delivers result 3. Supply is increasing in π (since as $\frac{\partial \bar{p}}{\partial \pi} < 0$), hence there is a unique market clearing price $\hat{\pi}$ as long as demand is downward sloping $D'(\pi) < 0$.

C.2 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 inverse CDF $h_H(t)$ and $h_L(t)$ 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:

$$Y(H, L) = \left[A_H(\sum t^H)^\theta + A_L(\sum t^L)^\theta\right]^{1/\theta}$$

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 π_H and π_L. The price of high talent is given by:

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

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

$$w_{t^H} = \pi_H \cdot t^H$$

Call the inverse CDF of wages p^w_{SBTC} and the probability that a wage is above w_p is:

$$p^w_{SBTC} = Pr(w_{t^H} > w_p) = Pr(t^H > \frac{w_p}{\pi_H}) = h_H(\frac{w_p}{\pi_H})$$

The top tail of the wage distribution follows the same distribution as t^H.[34] With Pareto shape parameter $\tilde{\alpha}$ the talent distribution is $h_H(t) = t^{-1/\tilde{\alpha}}$ and substituting this into the wage distribution yields a wage equation equivalent to the superstar distribution in 4 for the right value of $\tilde{\alpha}$.

C.3 Technical Change and Superstar Effects

This section derives proposition 2.1.

Part a) to see how growth varies across the distribution, differentiate equation 5 wrt to the percentile of the distribution

$$\frac{\partial g_e}{\partial p} = \frac{\tilde{\lambda}\tilde{\xi}}{\lambda\xi}\frac{(\kappa-1)}{\kappa}\frac{\xi}{\alpha}w^{\frac{(\kappa-1)}{\kappa}\frac{\xi}{\alpha}-1}\frac{\partial w}{\partial p} > 0$$

Since wages are increasing along the wage distribution, the expression on the RHS is positive. Hence, the growth rate increase as we move up the distribution, which proves part a). Note that an equivalent result holds for wage growth at the top.

Part b)The top income share is defined as the sum of incomes of individuals above percentile p divided by total income (G):

$$s_p = \int_p^1 w_j dj/G$$

Note that for a Pareto distributed variable the top income share is given by $s_p = (1-p)^{1-\lambda}$, with λ^{-1} the shape parameter of the distribution.[35] Here wages follow a Pareto distribution with shape parameter $\lambda = \frac{\alpha}{\xi}$. The growth in the top income share from superstar effects is therefore given by:

$$g^{s_p} = \frac{s_p^{t+1}}{s_p^t} \approx \frac{(1-p)^{1-\frac{\xi}{\kappa\alpha}}}{(1-p)^{1-\frac{\xi}{\alpha}}} = (1-p)^{-(\kappa-1)\frac{\xi}{\kappa\alpha}}$$

[34] Here we assume that low skill workers do not features in the top tail of the wage distribution.

[35] 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. This result has been used extensively to calculate top income and wealth shares.

The second step uses the property of a Pareto variable, the approximate result indicates that this results can works approximately for a much wider set of distributions. The final equality cancels terms. Top incomes shares are growing and we can see that the growth rate is increasing in p. This implies that the income share of the top 0.1% growths faster than the share that goes to the top 1%, which in turn growths faster than the share of the top 10%. The top 1% takes home a growing fraction of the income among the top 10%.

The core result, that a unit of talent becomes more valuable, 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 differentiating condition 2 with respect to s:

$$w_{ps}(\hat{t}) = Y_{ps}(\hat{t}) + Y_{pp}(\hat{t})\frac{\partial \hat{t}}{\partial s} = \frac{w''(\hat{t})}{\theta'(\hat{t})} > 0 \tag{13}$$

The second equality uses positive assortative matching to invert the assignment function $\hat{t} = \sigma^{-1}(s)$ and differentiates to yield $\frac{\partial \hat{t}}{\partial s} = \frac{1}{\sigma'(\hat{t})}$. 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.

Part c) The falling wage is a result of the growing supply of talent, which reduces π. As a result the Pareto scale parameter in equation 4 falls ($\lambda' < \lambda$) and the wage distribution shifts inward.[36] This level shift occurs across the distribution, among stars the growth in returns from scalability over-compensates for the fall in π, but for non-stars the decline in π dominates. This effect is also reflected in the share of jobs with mid pay. Given the assumption on the demand elasticity $(1 - \varepsilon < (\gamma\phi\pi)^{\kappa-1})$, the first term of equation 5 is smaller than 1(ie $\frac{\tilde{\lambda}\tilde{\xi}}{\lambda\xi} < 1$). As $w \to 1$, the growth rate $g_e = \frac{\tilde{\lambda}\tilde{\xi}}{\lambda\xi} - 1 < 0$ and hence the share of jobs at such pay levels is declining.

Part d) In the model with entry and exit the participation constraint (PC)

[36] 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.

ensures that the marginal participant is indifferent between working and the outside option: $w(\bar{p}) = w^{res}$. Deriving the equilibrium wage from integrating 2, we get $w(p) = \int_{\bar{p}}^{p} Y_t(\sigma(t), t) dh(t)$. And hence: $w(\bar{p}) = Y_t(\sigma(\bar{p}), \bar{p}) = w^{res}$. When Y_t fluctuates changes in $\bar{p}$ ensure that the PC holds. Raising $\bar{p}$ implies that the returns for the marginal worker $Y_t(\sigma(\bar{p}), \bar{p})$ increase, since $Y_{tt} + Y_{ts} > 0$. Hence when falling talent prices ($\tilde{\pi} < \pi$) lead to a decrease in Y_t, equilibrium requires that the participation threshold increases. Periods of technical change therefore lead to higher $\bar{p}$, which confirms statement d).

C.4 Technical Change and SBTC Models

C.4.1 Proportional Top Income Growth

Skill biased technical progress makes high skilled workers more productive ($\tilde{A_H} > A_H$). The wage per talent unit therefore becomes:

$$\tilde{\pi_H} = \tilde{A_H} \left[\frac{\sum \tilde{t^H}}{\tilde{Y}} \right]^{\theta - 1} > \pi_H$$

Next consider wages. The baseline case assumes that labor supply is inelastic, hence the talent distribution ($h^H(t)$) is unchanged. Allowing for a labor supply response complicates notation and generates little additional insight.[37] The wages at p are given by:

$$p^w_{SBTC} = (\tilde{w}_p / \pi')^{-\frac{1}{\alpha}}$$

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

$$g^w_p = \frac{\tilde{w_p}}{w_p} = \frac{\tilde{\pi_H}}{\pi_H} = g^w$$

Wage growth is the same across all percentiles in the top tail. At the top of the distribution technical change leads to a level shift in the wage schedule.

[37] 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 $G_H^{-1}(p)$. The result that follow therefore carry through approximately at the the top of the distribution.

C.4.2 No Fractile Inequality

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

$$g^{s_p} = \frac{s_p^{t+1}}{s_p^t} = \frac{G^t}{G^{t+1}} \frac{\pi_{t+1} \int_p^1 p^{-\alpha} dp}{\pi_t \int_p^1 p^{-\alpha} dp} = \frac{g^\pi}{g^G}$$

The second step uses the definition of top income shares and equation **??**. The final step collects terms and cancels. Top income shares grow as long as the price for talent growths faster than GDP. 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 ratio of the income share that goes to the top 1% and 10% is therefore unaffected by SBD shocks.

C.4.3 No Cannibalisation in SBTC Models

This section proofs that technical progress rules out falling wages in the SBTC model. I study a flexible SBTC model with arbitrary many skill groups 1 ... 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 α_i the associated productivity and θ is the driver of technical change. We allow for exit and therefore impose that no worker type is indispensable in production:

$$\frac{\partial F}{\partial L_i} < \infty \qquad \forall L_i$$

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:

$$\frac{\partial F}{\partial \theta} = \sum L_i \frac{\partial \alpha_i}{\partial \theta} \frac{\partial F}{\partial L_i} > 0 \tag{14}$$

We want to show that this implies that:

$$\frac{\partial \alpha_i}{\partial \theta} \geq 0 \qquad \forall i$$

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 14, 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.4.4 Extending SBTC to match Superstar Effects

By extending the SBTC model we can replicate properties of the superstar effect. It is useful to think what changes to the model are needed. The relation between the the two models is particularly apparent if we take the Rosen, 1981 superstar model where $f(s,t) = s \cdot q(s,t)$, hence output of an actor with quality t and audience s depends on the audience size s and the quality of the output produced q. Market clearing ensures that demand equals supply and for now assume per person spending is inelastic at K, market becomes $\int \pi \cdot f(s,t)h'(t)g'(s)dsdt = K$. Assume $\dot{q}_s < 0, \dot{q}_t > 0$ and $\frac{\partial \dot{q}}{\partial q \partial s} > 0$, where $\dot{q} = log(q)$, as illustration take the case:

$$q_i = t_i e^{1-\delta_t s_i}$$

the quality of output is the discounted talent (t_i) of individual i. The discount depends on the crowdedness of the show s; the more exclusive the show, the higher is the utility from it. Having Madonna play at a private dinner party brings greater utility, compared to listening to the same song on a recording. Comparative advantage $(\frac{\partial \dot{q}}{\partial q \partial s} > 0)$ implies that a world class performer is better able to deal with bigger audiences and thus quality suffers less from crowding $\partial \delta_t / \partial t < 0$. Take the simple case where $\delta_t = t^{-\phi}$, where ϕ is the technology that determines how easy it is to deliver a performance to a large audience. An increase in ϕ, as before, implies it gets easier to scale productions. Wages are given by $w(s,t) = P(q(s,t))s$, with P the price charged for a show of quality q. Profit maximizing implies equilibrium wages are given by:[38]

$$\dot{w}_i = \dot{\pi} + \dot{t}_i \phi$$

[38] with equilibrium $s = -q/q_s$

which features the "talent multiplier effect," where wages are more dispersed than talent, as long as $\phi > 1$. The intercept term captures the market value of a talent unit and does not vary by talent, it therefore is a level shifter in wages. It is also easy to see the superstar effect in the wage distribution by considering the effect of greater scalability. The wage effect of an increase in ϕ is

$$\frac{\partial \dot{w}}{\partial \phi} = \dot{\pi}_\phi + \dot{t}_i$$

the first term is again the same for all i, while the later term indicates that the superstar effect is bigger for more talented actors. The first term captures the cannibalisation effect that hurts less talented actors, as $\pi_\phi < 0$. Finally notice that this result relates to the dispersion in *log* wages. This is an important difference to standard models of skill heterogeneity. When workers are paid in line with their productivity: $w_i = at_i$, then a skill biased shock with an increase in the skill premium ($a \uparrow$) changes the wage distribution, but has only a level effect on the dispersion in *log* wages.

The SBTC model can replicate the superstar effect if we make each skill type unique. Hence, rather than assuming that people have different levels of skill, we assume that each worker is a specific skill group. The crucial difference is that it makes all workers imperfect substitutes and thus allows wages across individuals to differ by more than there skill units. Simultaneously, this makes it feasible for wage gaps to grow differentially for workers with the same skill discrepancy and thus we can make a marginal talent unit more valuable at the top and generate superstar effects. The clue is that imperfect substitutability breaks the law of one price which forced wage differences to be proportional to skill differences. Each individual has it's own productivity term and the wage of i is given by:

$$w_i = A_i \left[\frac{1}{Y}\right]^{\theta - 1}$$

we can thus replicate the superstar effect if a technical change generates $\frac{\partial \dot{A}}{\partial \phi} = \dot{\pi}_\phi + \dot{t}_i$. A model with a continuous distribution of unique talent types replicates, but does not coincide with the superstar model. A key difference is the process that generates wage dispersion. In the SBTC model wage inequality changes from biased productivity shifts, while in the superstar model from shifts in the distribution of customers (or individual specific capital).

There are two unappealing feature of such an extended SBTC model is. First, we have as many productivity terms as workers and thus can explain any kind of wage change, making it a somewhat uninteresting model. Second, to explain falling wages we require technical regress. Somewhat counterintuitively, lower skilled workers lose access to the previous, more productive technology and innovation thus makes them "forget" how to be productive.

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 11 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 14 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 Piketty2003 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 11 confirms this formally. It repeats the previous analysis using other top income measures. Column 1 repeats the baseline estimate. Column 3 uses the top income percentile as defined by Piketty2003.

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 entertainer 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 13 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 14. 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 14. 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 placebo occupations, 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, Larsen, and Palmer, 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, Larsen, and Palmer, 2016 to implement the difference in difference for quantile regressions. The estimated coefficients are plotted in figure 16, 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 15 reports the panel quantile estimates using the Powell, 2016 approach.

D.2 Policy Effects in a Superstar Setting

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 (**Acemoglu2011**; 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 1 of Table 17). 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 (**Kleven2013**; **Moretti2017**; Kleven, Landais, Saez, and Schultz, 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, Saez, and Stantcheva, 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.[39] 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 2 of Table 17 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.

Finally, I consider the possibility that a TV station's 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 3 of Table 17, 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 4), which suggests that pre-TV labor market competitiveness does not greatly influence superstar effects.

D.3 Data construction

D.3.1 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, they are therefore constructed to yield strong within-labor-market commuting and weak across-labor-market commuting.

[39] 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.

- David Dorn provides crosswalks of Census geographic identifiers to commuting zones Autor2013b. I use these crosswalks for the 1950 and 1970 data
- 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.
- 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.3.2 Worker data

Data is provided by the Integrated Public Use Microdata Files (IPUMS, **Ruggles2017**) of the US decennial census from 1930-1970 (excluding Hawaii and Alaska). 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. 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
- To match TV signal exposure to the Census I map county level TV signal information onto geographic units available in the Census. The geographic match uses the boundary shapefiles provided by NHGIS (Manson, Schroeder, Riper, and Ruggles, 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.

D.3.3 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 thus don't fall into the sample of interest.[40] Second, the rate of unemployed is modest compared to

[40]There are a number of cases were the unemployed report an occupation. This occurs if they

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.3.4 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.

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.

- 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.
- 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.3.5 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, Piketty, and Saez, 2011; Atkinson and Piketty, 2010; Blanchet, Fournier, and Piketty, 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[41]:

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

- 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 heteroskedacticity and correlation of errors across observations. Moreover, the log transformation implies that $E(\epsilon_i) = E(log\varepsilon_i) \neq 0$, 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.[42] Polivka, 2001 and Armour, Burkhauser, and Larrimore, 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 than the bottom of the distribution. Even at the top of the distribution changing Pareto coefficients may be required to fit the data (Blanchet, Fournier, and Piketty, 2017). Misspecification is particularly problematic for the more efficient estimators (Finkelstein, Tucker, and Alan Veeh, 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, $rank_j$ the rank in the wage distribution (1 being the top), w_j wage at rank j and ω the smallest

[41] Here $\beta = ln(\underline{income}) - ln(\underline{rank})$ where lower bars represent the lower bound of the interval considered

[42] Since the covariance structure of order statistics is known, GLS yields the same result

wage in the sample:

- MLE: $\hat{\beta}^{MLE} = \frac{1}{n}\sum_{j=1}^{n} log(w_j/\omega)$
- MLE (top code adjusted): $\hat{\beta}^{MLE_{TC}} = \frac{T}{n}\sum_{j=T}^{n} log(w_j/\omega) + T * log(w^{TC}/\omega)$
- OLS: $log(w_j) = \delta - \beta^{OLS} * ln(\frac{rank_j}{n+1}) + \epsilon_j$
- Close to cut-off: $\hat{\beta^A} = (\sum_{j=1}^{3} \frac{ln(w_j/w_{j-1})}{ln(rank_j/rank_{j-1})})^{-1}$
- 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 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 16 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.3.6 Data on Market Reach of Entertainment Shows

- Data on potential show audiences is collected from the "Julius Cahn-Gus Hill theatrical guide." For each local labor market I compute the potential maximum audience. For physical venues this is the seating capacity of the largest venue.

- Show revenues in theatres are the price of tickets multiplied by the audience. I use the average price if multiple ticket prices are reported. For TV shows I collect price data from rate cards. Such cards specify the price for sponsorship of a show at a local station, which allows me to compute the price charged for a TV show. From the price per show I can compute a price per TV viewer, analogue to a ticket price, which quantifies the marginal return to reaching one more customer. 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 use them to estimate a demand elasticity for TV audiences, taking the supply of TV hours as given. The demand curve for a TV viewer is estimated as: $ln(price) = 4.051 + -0.460 * ln(TV households)$. The negative elasticity indicates that, as expected, the marginal value of reaching a household is declining. The negative demand elasticity in turn implies that TV station revenues do not increase 1:1 with audience, the revenue elasticity is 0.54.

- The potential audience of TV shows is the number of TV households that can watch a local TV station. This is computed using information on TV signal catchment areas (from Fenton and Koenig, 2018) and TV ownership records from the Census.

D.3.7 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 as 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.3.8 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.[43] 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.3.9 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.

[43] Such tax tables have been used by Kuznets and Piketty to construct time series of top income shares for the US population.

References

Armour, Philip, Richard Burkhauser, and Jeff Larrimore (2015). "Using the Pareto Distribution to Improve Estimates of Topcoded Earnings". In: *Economic Inquiry*.

Atkinson, Anthony B and Thomas Piketty (2010). *Top Incomes - A Global Perspective*.

Atkinson, Anthony B, Thomas Piketty, and Emmanuel Saez (2011). "Top Incomes in the Long Run of History". In: *Journal of Economic Literature* 49.1, pp. 3–71.

Blanchet, Thomas (2016). "Wealth inequality in Europe and in the United States: estimations from surveys, national accounts and wealth rankings". In: June.

Blanchet, Thomas, Juliette Fournier, and Thomas Piketty (2017). "Generalized Pareto Curves: Theory and Applications". In: *WID.world Working Paper* 2017/3.March.

Chetverikov, Denis, Bradley Larsen, and Christopher Palmer (2016). "IV Quantile Regression for Group-Level Treatments, With an Application to the Distributional Effects of Trade". In: *Econometrica* 84.2, pp. 809–833.

Feenberg, Daniel R. and James M. Poterba (1993). *Income Inequality and the Incomes of Very High-Income Taxpayers: Evidence from Tax Returns*. Vol. 7. January, pp. 145–177.

Fenton, George and Felix Koenig (2018). "The Labor Supply Response to Innovation in Entertainment: Evidence from TV". In: *Working paper*.

Finkelstein, Mark, Howard G. Tucker, and Jerry Alan Veeh (2006). "Pareto Tail Index Estimation Revisited". In: *North American Actuarial Journal* 10.1, pp. 1–10.

Katz, Lawrence F. and Kevin M. Murphy (1992). "Changes in Relative Wages, 1963-1987: Supply and Demand Factors". In: *The Quarterly Journal of Economics* 107.1, pp. 35–78.

Kleven, Henrik Jacobsen et al. (2013). "Taxation and International Migration of Superstars: Evidence from European Football Market". In: *American Economic Review* 103.5, pp. 1892–1924.

Kuznets, Simon and Elizabeth Jenks (1953). *Shares of Upper Income Groups in Savings*. Vol. I, pp. 171–218.

Manson, Steven et al. (2017). *IPUMS NHGIS: Version 12.0*. Tech. rep. Minneapolis: University of Minnesota.

Penniman, Clara and Walter Heller (1959). *State income tax administration*. University of Michican: Public Administration Service, p. 280.

Piketty, Thomas and Emmanuel Saez (2003). "Income Inequality in the United States, 1913-1998". In: *The Quarterly Journal of Economics* 68.February, pp. 1–39.

Piketty, Thomas, Emmanuel Saez, and Stefanie Stantcheva (2014). "Optimal Taxation of Top Labour Incomes: A Tale of Three Elasticities". In: *American Economic Journal: Economic Policy* 6.1, pp. 230–271.

Polivka, Anne Elise (2001). "Using Earnings Data from the Monthly Current Population Survey".

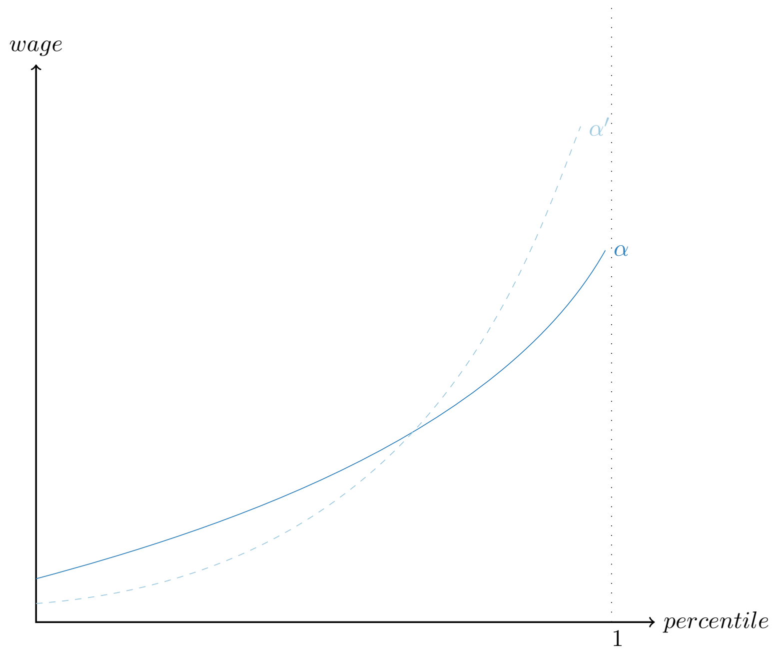
Powell, David (2016). "Quantile Regression with Nonadditive Fixed Effects". In: *RAND Working Paper* April, pp. 1–28.

Rosen, Sherwin (1981). "The Economics of Superstars". In: *American Economic Review* 71.5, pp. 845–858.

Scheuer, Florian and Iván Werning (2017). "The Taxation of Superstars". In: *Quarterly Journal of Economics* 132(1), pp. 211–270.

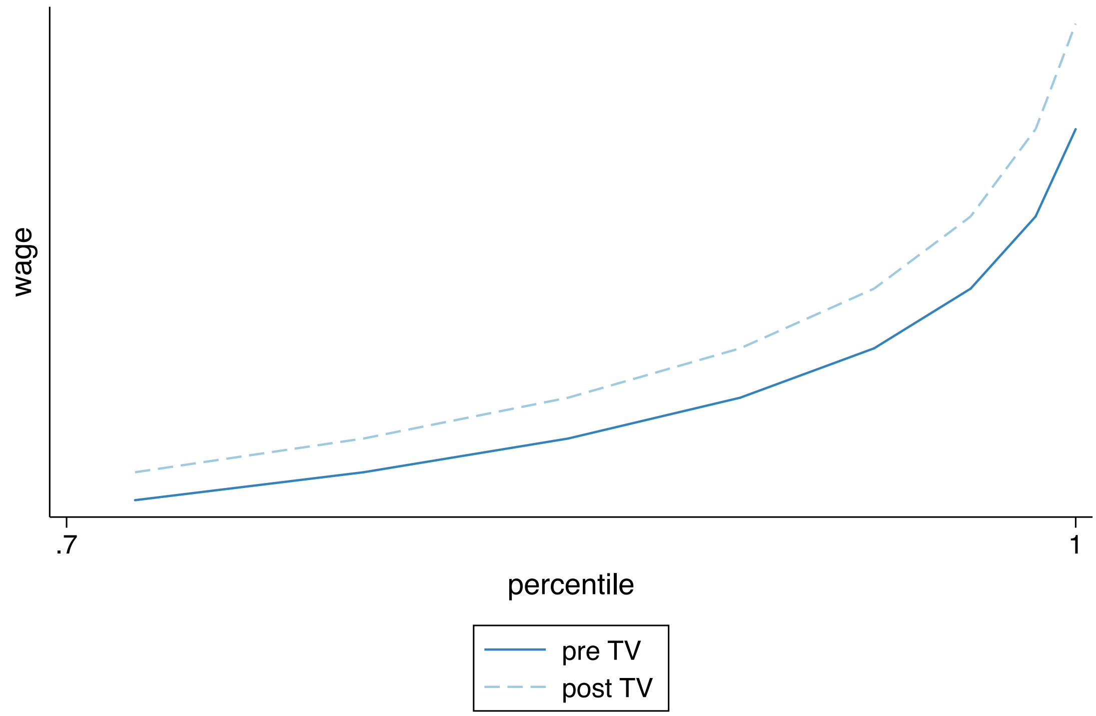
D.4 APPENDIX: TABLES & FIGURES

Figure 9: Superstar Wage Distribution



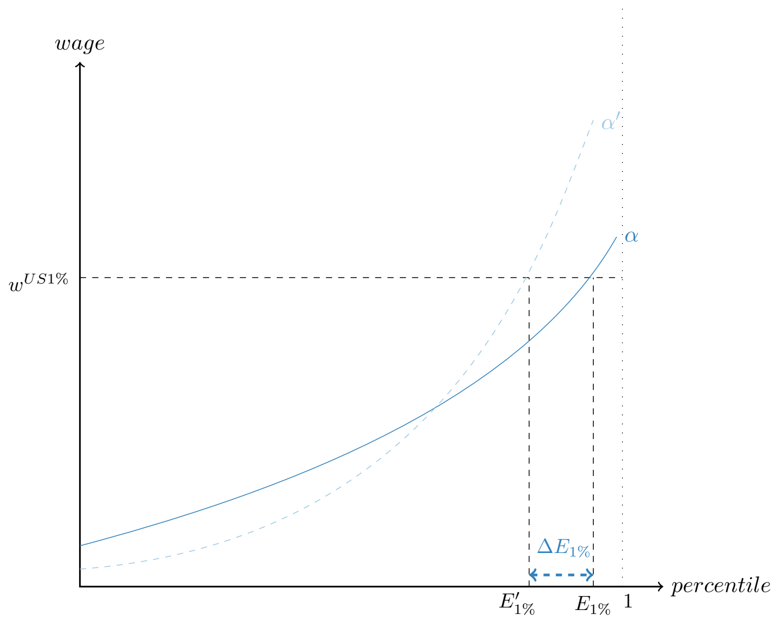
Note: Wages based on a superstar model ($w_p = \pi \cdot \kappa \cdot (1-p)^{-(\alpha\gamma-\beta)}$). α is the shape parameter of the market size distribution ($\alpha' > \alpha$). The percentiles shown are the upper tail of the wage distribution. With exit they correspond to the percentiles in the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

Figure 10: Effect of Technical Change on Wage Distribution - Skill Biased Demand Model



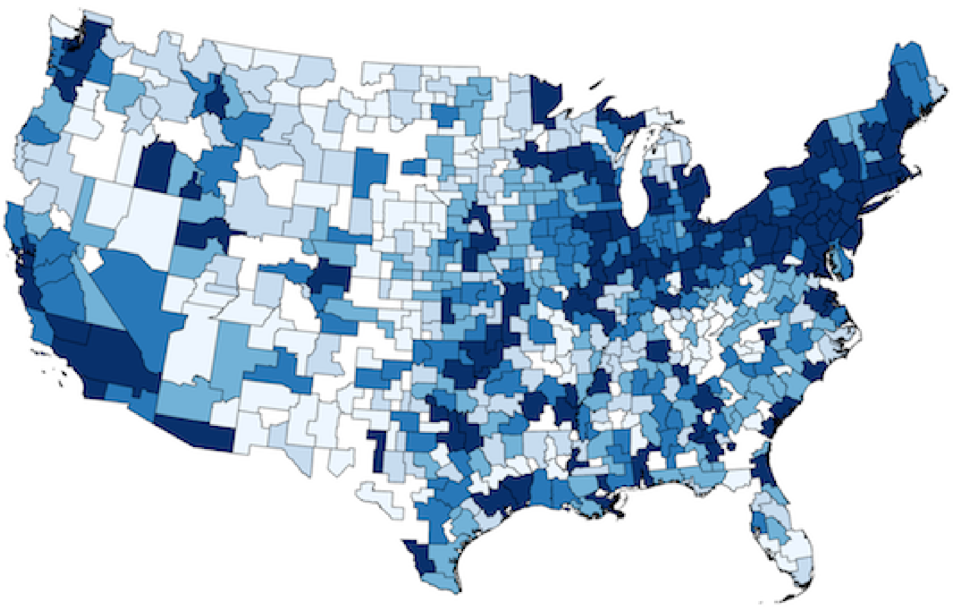
[Note] The figure shows the wage distribution above the 70th percentile. The talent distribution has been chosen to match the 1940 wage distribution. The change in the skill premium matches the growth in the share of top earners.

Figure 11: Superstar Effect on Top Earner



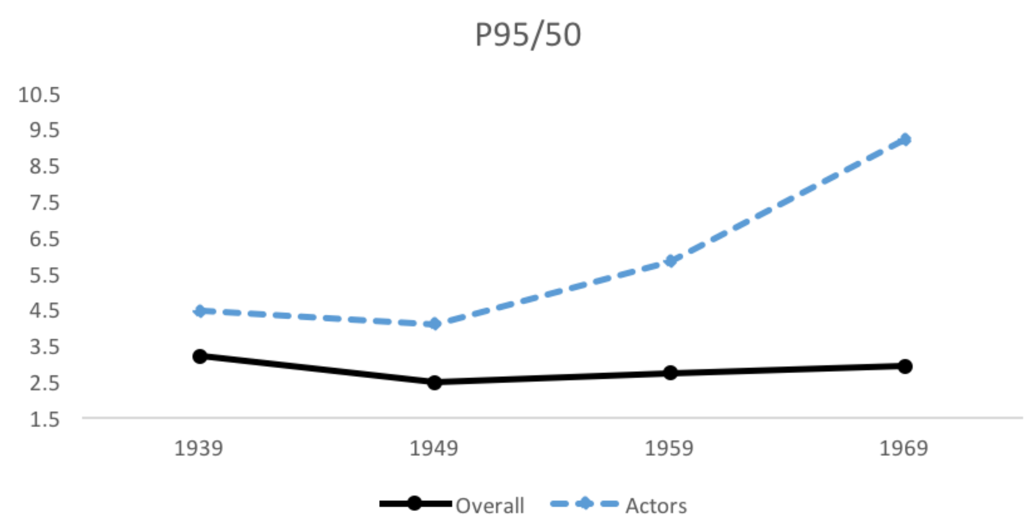
[Note] Details as in figure 9. $w^{US1\%}$ is a wage threshold that defines a top earner, e.g. the national top percentile. $E_{1\%}$ and $E'_{1\%}$ are the share of entertainers above the threshold. $\Delta E_{1\%}$ is the change in top earners when market size becomes more dispersed (move from α to α').

Figure 12: Theatre Seating Capacity



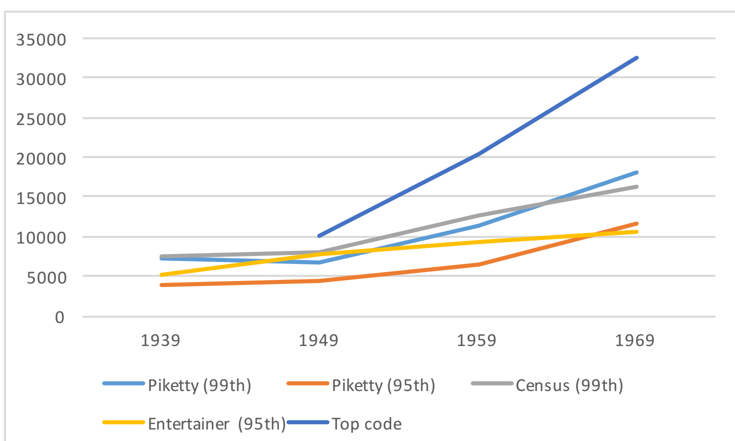
[Note] Performance venues are the venues listed in Julius Cahn-Gus Hill's 1921 theatrical guide. Size refers to the average seating capacity of the largest venues in the commuting zone.

Figure 13: P95-P50 Gap



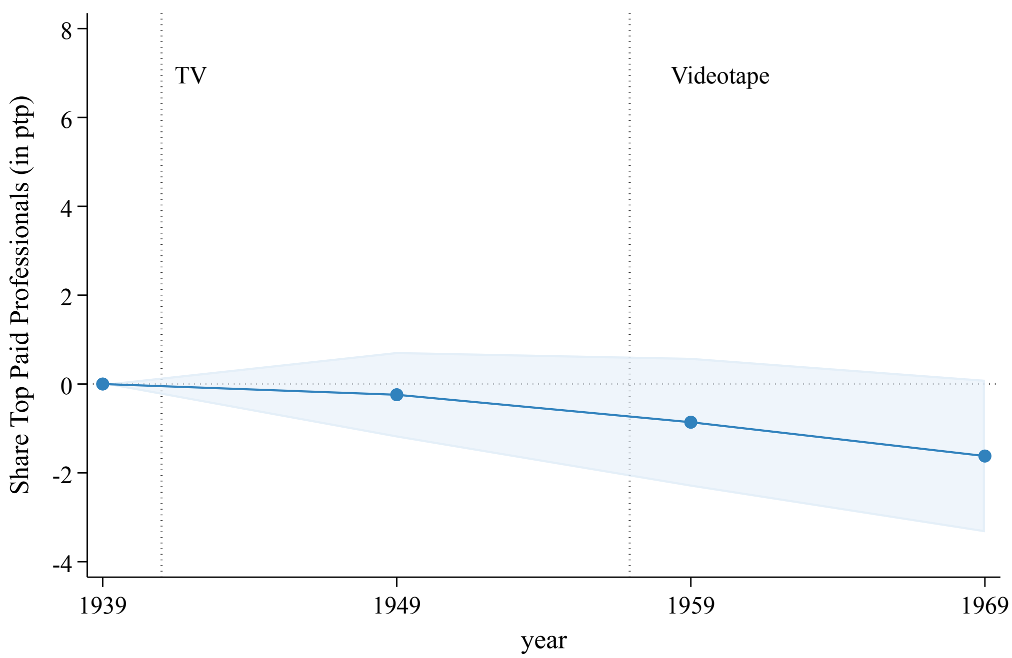
[Note] Figure reports the ratio of wages at the 95th and median. Percentiles are from the wage distribution reported in the US decennial Census for the lower 48 states.

Figure 14: Top Income Percentile Values



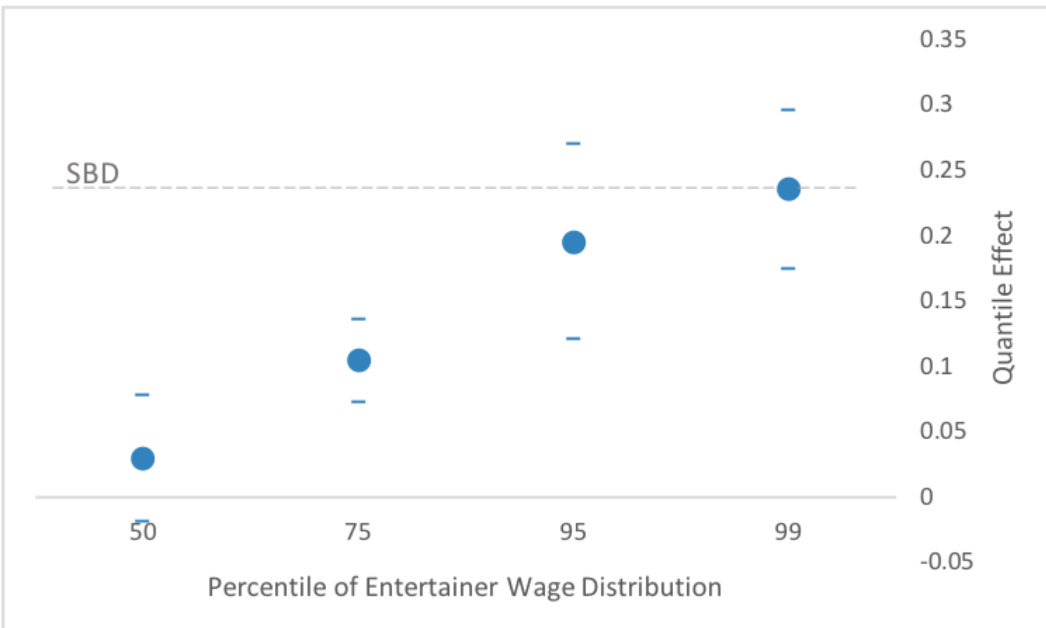
[Note] The Figure shows the top code cut-off in the US Census data and top percentiles of the wage distribution in the Census years. The name in the legend refers to the sourcee of the wage distribution: Census refers to percentiles in the Census data wage distribution, Entertainer to percentile in the distribution of entertainer wages in the Census, Piketty to the data reported in the World Top Income Database, top code is the top code in the IPUMS Census data – there is no top code for the 1939 full count Census data. The number in the bracket in the legend indicatees the percentile of the distribution that is shown.

Figure 15: Dynamic Treatment Effect of TV stations - Placebo Occupations



[Note] The figure shows regression coefficients from the dynamic difference in difference regression for placebo occupations. Reported are the coefficients on local TV antennas and 95% confidence bands are shown. Standard errors are clustered at the local labor market level.

Figure 16: 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, Larsen, and Palmer, 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 9: Effect of TV on Top Earner - Placebo Occupations

	(1)	(2)	(3)
		Panel A: ln(Wage at 99^{th} Percentile)	
Local TV station	0.023	0.019	0.016
	(0.004)	(0.003)	(0.005)
Outcome mean	9.08	9.08	9.08
Effect size	2.3%	1.9%	1.6%
		Panel B: Share of Occupation in US Top 1% (ptp)	
Local TV station	0.21	0.66	1.09
	(0.52)	(0.89)	(0.52)
Outcome mean	5.55	5.55	5.55
Effect size	4%	12%	20%
		Panel C: Local Population Share in US Top 1% (in 10,000)	
Local TV station	0.438	0.524	0.865
	(0.221)	(0.234)	(0.319)
Outcome mean	10.86	10.86	10.86
Effect size	4%	5%	8%
Cluster	722	722	722
Demographics	–	Yes	–
Local labor market trends	–	–	Yes

[Note] Each cell is the regression coefficient of a separate regression. Panel A uses a quantile regression for within group treatment Chetverikov, Larsen, and Palmer, 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 occupation-local labor market - year 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 devided by the population in a local labor market. Entertainer are Actors, Athletes, Dancers, Entertainers Not Elswhere 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 10: Effect of TV on Top Earner - Alternative Top Income Measures

	(1)	(2)	(3)
	Panel A: Count Entertainer in US top 1%		
Local TV station	30.91	32.09	19.31
	(8.92)	(9.92)	(8.31)
Outcome mean	15.53	15.53	15.53
	Panel B: Share Entertainer in US top 1% (denominator fixed)		
Local TV station	6.51	6.73	9.21
	(1.90)	(1.89)	(3.44)
Outcome mean	6.39	6.39	6.39
	Panel C: Share Entertainer in US top 1%		
Local TV station	0.178	0.193	0.194
	(0.025)	(0.038)	(0.063)
Outcome mean	0.28	0.28	0.28
Cluster	722	722	722
Demographics	–	Yes	–
Local labor market trends	–	–	Yes

[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 11: Alternative Top Income Measures

	(1)	(2)	(3)	(4)	(5)
	Share in US top 1%		Count top 1%	Share in top 5%	
Local TV station	90.19	132.5	30.91	31.64	120.0
	(26.25)	(35.92)	(8.92)	(16.36)	(47.85)
threshold	Census	Piketty & Saez	Census	Entertainer	Entertainer (1940)
mean outcome	94.27	109.09	18.39	150.02	372.10
% growth	96%	121%	168%	21%	32%

[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 12: Effect of TV on Top Earner - Micro Data

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

[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 3,438,002 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 13: Effect of TV on Top Earner - State Level

	Share in Top 1%		
	(1)	(2)	(3)
Local TV station (1940)			-9.62
			(5.95)
Local TV station (1950)	20.94	20.18	-2.98
	(8.09)	(7.36)	(1.79)
Local TV station (1960)			-9.95
			(6.17)
Local TV station (1970)			-13.33
			(8.07)
Years	1940-1970	1916-1970	1916-1970
Observations	912	1008	1008

[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 14: Earning Effect - triple diff

	Share in Top 1%		
	(1)	(2)	(3)
TV × Placebo Occupation	-0.41		
	(0.47)		
TV × Performance Entertainer	4.87	4.87	4.17
	(2.16)	(2.16)	(1.57)
TV × Interactive Leisure		-3.40	
		(1.29)	
TV × Drink & Dine		-3.80	
		(1.84)	
TV × Professional Services		5.23	
		(4.86)	
TV × Medics		-3.24	
		(1.52)	
TV × Engineer		-1.12	
		(1.23)	
TV × Manager		3.55	
		(2.21)	
Location & Occupation-Year FE	Yes	Yes	–
Pairwise Interaction: Location, Year, Occupation FE	–	–	Yes

[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 15: Quantile Effect of TV

	Wage Percentiles			
	99th	95th	75th	50th
Local TV station	260.3	85.00	22.33	19.13
	(92.23)	(3412.5)	(445.3)	(101.2)

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

Table 16: Small Sample Performance of Pareto Shape Parameter Estimators

Estimator	sample 10%	local 5% sample
True	0.460	0.460
OLS	0.558	0.715
MLE	0.617	0.629
MLE (top code)	0.640	0.618
Close to cut-off	0.478	0.480

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 ($\alpha < 1$)

Table 17: Policy Effects in a Superstar Setting

	(1)	(2)	(3)	(4)
Local TV station	4.83	4.59	4.25	4.25
	(4.56)	(1.77)	(2.25)	(2.25)
Local TV station ×	-1.42			
% with high-school degree	(12.23)			
Local TV station ×		0.10		
high tax state		(1.65)		
Local TV station ×			-0.00	
theatre count			(0.02)	
Local TV station ×				-0.00
population density				(0.02)

[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.