Superstar Earners and Market Size: Evidence from the Entertainment Industry*

Felix Koenig, LSE

May 2018

Abstract

This paper provides a first quasi-experimental test for the superstar theory. The superstar theory predicts that wage growth is concentrated at the top during periods of technical change, with particularly stark growth for the highest percentiles of the wage distribution. To test the model, this paper exploits technical change in entertainment coming from the quasi-random introduction of television filming across local labor markets. A government run licensing procedure and a regulator shut down lead to plausibly exogenous variation in television access and allows me to test the key model predictions. The impact of technical change on the wage distribution is closely aligned with the predictions of the superstar model providing strong evidence for the superstar effect. My results show that the superstar effect can generate vastly skewed income growth. The income share of the top 1% of entertainers grows by 30% as market size doubles. At this rate superstar effects could explain about half of the growth in US top income shares.

Keywords: Superstar Effect, Inequality, Top Income, Skill Biased Demand
JEL classification: J31, J23, J24, M52, O33, D31

*E-mail: f.koenig@lse.ac.uk. Address: Department of Economics, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, U.K. Acknowledgements: I am grateful to many people, in particular David Autor and Steve Pischke for detailed comments on this draft.

1 Introduction

Top incomes have grown rapidly in many countries. In the US the share of national income going to the top 1% of the income distribution has more than doubled over the past 50 years. Such trends have occured in many sectors and affects most OECD countries. However, little is known about the economic forces that drive the rise in top incomes. Many economists argue that technical change can give rise to so called "superstar effects." It's best known prediction is that the value of talent rises with scale of operation. A good idea, for instance, is more valuable if it can be used on a bigger scale. On the contrary a mistake is particularly costly if it affects large operations. Employers may thus be willing to pay a lot more for a small gain in talent. This leads to the key prediction of the superstar model: If it becomes feasible to serve bigger markets, income will become more concentrated at the top. At the same time ordinary workers suffer from the growing market penetration of stars. Recent technological innovations plausibly made it easier to operate on a large scale. Technical innovation may thus explain the rise in top incomes. Yet there is little evidence that superstar effects operate in practice.

This paper tests the superstar theory and provide a first quasi-experimental evaluation of its core predictions. To the best of my knowledge there is no evidence for the superstar effect. A key limitation for such tests is that top income growth is not a unique feature of superstar models. The paper makes three contributions. First, I show what makes the superstar model different from conventional models of inequality. Second, I test the superstar model using quasi-random variation from the entertainment sector. And third, I estimate how much superstar effects contributed to the changes in the US top 1% income share.

The superstar model has unique predictions about the returns to talent, however talent is hard to measure. To overcome this problem this paper focuses on a different comparative static, the change of the income distribution during periods of technical change. I show that the superstar model generates a particular form of skewed income growth that is characteristic for the model. This effect is distinct from the impact of conventional skill biased demand shocks and independent of the underlying distribution of talent. Periods of technical change will allow us to test the superstar model.

Technical change is rarely observed and tends to occur for endogenous reasons, estimating the effect of technical change is therefore challenging. A popular approach is to compare labor market trends across sectors or tasks that are differentially exposed to technical change (e.g. routine vs non-routine tasks). A limitation of this approach is that the effect of the technical change can not be distinguished from broader time trends that are occurring simultaneously. A second approach looks at differential adaption of technology. Critics have pointed out that this will likely conflate the effect of technology with the endogenous factors that caused the differential adaption in the first place (see DiNardo and Pischke [1997] for an example).

To overcome the endogenity challenge this paper makes use of the government regulated introduction of a new technology: the construction of television in the middle of the 20st century. I use the government restricted, staggered roll-out of television across local labor markets to estimate the causal effect of the technology on the labor market for entertainers. Early television shows were mainly produced locally. This gives rise to a location specific change in entertainment production. When a TV station was set up local entertainers could appear on TV and had a dramaticly bigger audience than previously. The feasible scale of an entertainment shows almost trippled. The shift in the production technology coincides with a striking change in the wage distribution of entertainers. Figure 1 shows the sharp rise in inequality in entertainment. The extremes of the distribution grew, while mid paying jobs disappeared. To test wether the change in the wage distribution is indeed the result of technical change we can exploit the local variation during the roll-out period.

Two features of the roll-out allow to check if the introduction of television is exogenous to local labor market conditions. First, in 1948 processing of all pending licenses was stopped, regardless of the characteristics of the local labor market. The shut down of the regulator was the result of an error in the assignment model that led to signal interference between existing stations. Information on the pending licenses allows me to identify which areas were affected. These areas are excluded from the technology for plausibly exogenous reasons. We can use "stations that did not happen" to test whether local labor markets change differentially around the time of (planned) TV introduction. Second, the invention of the videotape made TV shows a transportable good. Shows could be recorded on a videotape in a central location and distributed to stations across the country. This led to a concentration of production in very few markets and the demise of local production. Local TV licenses issued after the invention of the videotape should have no effect on local entertainer stars. We can use this variation to test whether the introduction of a television is spuriously related to local labor market changes. Moreover, the superstar effect of stations that previously produced locally should disappear as they stop relying on local talent.

Production and consumption of entertainment is unusually well measured and documented, which makes the sector an appealing set up to study the effect of technology. Despite the availability of such data, the records had so far not been digitized systematically. This paper brings together a large volume of newly collected data on entertainment in the US. It links archival data on the production technology in entertainment, performances and entertainment consumption. This data includes information on the universe of broadcasting licenses, locations and audience sizes of TV stations and historic capacity of over 3,000 performance venues as well as local attendance and spending at over 4,000 county fairs. This data is complimented by administrative records on the licensing process of TV stations. I combine the data with labor market data from the microdata files of the US Census. The Census covers the universe of the US population until 1940 and samples in the following decades. Detailed information on wages, occupation and location identify entertainers and whether they are affected by television. A second advantage of the setting is that entertainment is a locally produced and non-tradable during the period studied. This allows us to compare wage distributions across seperate local labor markets. A final advantage of the entertainment setting is that the introduction of television leads to a sizable shock to the scalability of production. This major change gives us the best chance to detect and precisely estimate the effects of interest.

I find substantial evidence for the superstar effect. Television leads to a sharp rise in inequality in entertainment with particularly large gains for the top of the wage distribution, moreover income growth escalates the further we move up the wage distribution. The introduction of a TV channel doubles the local share of entertainers who make it into the top 1% of the US wage distribution. The income share of the top 0.1% of entertainers grows substantially faster than the top 1%, which again growths faster than the share of the top 10%. Television therefore increases the gap between the 90th and the 99th percentile of the wage distribution in entertainment and leads to greater dispersion of incomes in the tail. Moreover, in line with the superstar model the impact of the technical change on the wage distribution is U-shaped. Large income gains for top entertainers come at the expense of ordinary entertainers. I find that the introduction of local TV leads to a decline of mid paid entertainment jobs, while the low paid segment thrives. Demand is shifting from mediocre local entertainment to a smaller group of regional star entertainers. The rise of entertainment stars is accompanied by a significant drop in entertainment jobs. The introduction of a TV channel reduces local employment in entertainment by about 10%.

Can market forces generate extreme concentration of incomes at the top? This debate is one of the longest running debates in economics. Marshall observed that "A business man of average ability and average good fortune gets now a lower rate of profits [...] than at any previous time, while the operations, in which a man exceptionally favored by genius and good luck can take part, are so extensive as to enable him to amass a large fortune with a rapidity hitherto unknown." [Marshall, 1890]. Tinbergen formalized the idea of superstar effects in an assignment model. He shows that competitive markets can generate substantial income concentration at the top [Tinbergen, 1956]. Building on this result Rosen endogeneises consumer demand and shows that all income goes to a single superstar as marginal cost goes to 0, while income is convex in talent in non degenerate cases [Rosen, 1981]. Since this seminal paper the theory has been used to account for a number of other stylised facts. Increases in CEO pay have been linked to superstar effects. A superstar model can generate most of the observed rise in executive pay for plausible parameter values [Gabaix and Landier, 2008, Terviö, 2008]. Dispersion of pay within firms can be rationalized by a version of the superstar model that takes hierarchical firm structures into account [Garicano and Hubbard, 2009, Garicano, 2000]. Recently, Autor et al. [2017] argue that superstar effects at the firm level can explain the observed falling labor share. Despite the wide applicability of the model the empirical evidence for the effect is thin, the handbook on income inequality concludes that the superstar model has "yet to receive much careful empirical scrutiny to determine its broader relevance for understanding wage structure changes" [Katz and Autor, 1999]. This paper aims to fill this gap by offering a quasi-experimental test of the core prediction of the model. Skill biased technical change is the current dominant model for the impact of technology on the labor market (see Acemoglu and Autor [2011] for a summary). This model matches the trend in the US skill premium remarkably well [Katz and Murphy, 1992]. This paper shows that this class of skill demand models would struggle to generate substantial change in top income inequality. The conventional skill demand model features two groups of workers low (L) and high (H) skilled. Within these groups workers are perfect substitutes. There are only two wage levels, therefore the simple model is silent on top income inequality. The model can be extended to feature workers with different amounts of H (or L). With the appropriate assumption on the distribution of skills this extended model can produce any distribution of wages, making it hard to reject the model. To get around this problem, I focus on the dynamic prediction of the model. Here the skill demand model is far less flexible. I show that the skill demand model only generates limited changes in top income inequality. Perfect substitutability of workers within skill groups constraints such changes. A worker with twice the talent to the rest of the skill group can be replaced by two ordinary workers, this in turn implies that the wage for the talented worker will always be twice the ordinary wage, independent of changes in the skill premium. This contrasts sharply with a superstar model where a worker has a unique type of talent that cannot be easily replaced by other workers. This imperfect substitutability inherent to the superstar model will lead to more pronounced changes in top income inequality, pivoting the wage profile. Most of the income gains will go to the superstars, with growth rates declining as we move downward in the distribution and eventually turning negative. Unlike a positive demand shift the superstar effect will result in falling wages for ordinary workers.

The magnitude of superstar effects and its role in explaining growth of top incomes is controversial. Kaplan and Rauh show that top incomes are rising across a broad set of occupations in the US. Based on this they conclude that superstar effects likely play an important role in explaining top income growth [Kaplan and Rauh, 2013]. Piketty, however, cautions that top income growth has not occurred uniformly across countries. Based on this he disputes that superstar effects have played a major role [Piketty, 2014]. The literature has estimated elasticities of top pay to market size to quantify the importance of superstar effects. Most of these elasticity estimates are based on results from the literature on CEO pay and correlate CEO pay with measures of firm value. The estimated elasticity of top income to changes in market size fluctuates between 1/3 and 1 in recent decades[Gabaix and Landier, 2008, Gabaix et al., 2014, Murphy, 1999] and 0.1 earlier in the 20th century [Frydman and Saks, 2010]. At the higher end these estimates imply that top pay rises one to one with market size. I derive the elasticity for the entertainment sector using the quasi-experimental variation described above. This allows and calculate the predicted rise in top incomes from superstar effects. I find that the elasticity is around 0.1. In an empirical calibration I show that if we apply these estimates to the US macro economy the superstar effect explains about 25% of the growth in the top 1% income share in the US.

The remainder of the paper will present the superstar model in section 2, outline the empirical strategy in section 3, present the data in section 4, results of the test of the superstar model in section 5 and finally estimate the magnitude of superstar effects in section 6, section 7 will conclude.

2 The Superstar Model

This section discusses how the wage distribution responds to technical change in a superstar model. The headline prediction of the superstar model is that technical progress leads to income growth that is skewed towards the top. The model offers further predictions covering the entire income distribution. This will give rise to a range of tests that can be taken to the data. The power of the superstar model to generate top income growth has received much attention. There has been less on focus on whether other models could generate similar predictions. I argue that a suitably extended model of skill biased demand (SBD) shifts can also produce income concentration at the top. Similarly a model of wage bargaining will generate top income concentration if bargaining power is shifting towards top earners. The second part of the section discusses predictions that allow to differentiate superstar effects from these models. These predictions arise from the core assumptions of each model and do not require a strong stands on the exact specification of the model. I propose a number of tests that allow to rule out a wide range of alternative models.

Superstar models come in one of two flavors, "differential rent models" (as in Rosen [1981], Garicano and Rossi-Hansberg [2004]) or "assignment models" (Gabaix and Landier [2008], Terviö [2008], Edmans and Gabaix [2016]). The two types yield the same predictions for the wage distribution but differ in the derivation. The main difference between the two types is an assumptions on the timing of decisions. In the differential rents model span of control and worker talent are chosen simultaneously, while in the assignment models differences on the firm side are taken as given at the time of choosing talent. In many ways these assumptions are similar to a model that treats capital as fixed or flexible at the time of choosing labor. The differential rent model is analogous to a model where capital and labor are chose simultaneously, while capital is treated as fixed at the time of choosing labor inputs in the assignment model. Empirical predictions will be easier to illustrate in an assignment model. This approach treats differences in employment opportunities as fixed and analyze an exogenous change in those characteristics. Each firm is endowed with immovable productivity characteristics. They are commonly thought of as market size or span of control (S_i). Treating S_i as given will allow us to analyze the response of the wage distribution to a shift in S_i.

The superstar model features workers with heterogeneous talent. Talent is distributed according to some distribution function with percentile $p = F(T)$.[1] Each worker is matched with a job i with span of control S_i. In the context of entertainment, the productivity characteristics of an employer can be thought of as the audience capacity of a stage. For simplicity I assume that at each point in time only one individual can appear on each stage. This assumption hard codes the imperfect substitution of quantity and quality into the model. Firms cannot employ additional workers to compensate for the lower talent. Other work discusses how the model extends matching many workers to a given market (see Sattinger [1993]) The output of a worker depends on the market size S and is given by $Y(S,p)$ with $Y_S > 0$, $Y_p > 0$, $Y_{pp} \leq 0$ and $Y_{Sp} > 0$ where subscripts denote partial derivatives. The last condition ensures that talent becomes more valuable with market size. This condition will generate positive assortative matching in equilibrium. The optimization problem for a firm with characteristics S_i is given by:

[1] We assume that the CDF is continuous and differentiable. This will guarantee competitive like behavior of the market, despite the fact that each talent level is only supplied by a single worker.

$$max_p Y(S_i, p) - w(p)$$

where $w(p)$ is the wage for a worker at percentile p of the talent distribution. The equilibrium assignment must be incentive compatible and meet participation constraints. Incentive compatibility (IC) ensures that in equilibrium each firm prefers hiring it's match to any other match. For each firm i the IC condition guarantees that the optimal worker p meets:

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

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

$$Y(S_i, p) - w(p) \geq \pi^{res} \quad \forall p \epsilon [0, 1] \tag{2}$$

$$w(p) \geq w^{res} \quad \forall p \epsilon [\bar{p}, 1] \tag{3}$$

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

The number of IC constraints can be reduced substantially for these kind of incentive compatibility problems. If the IC holds for the adjacent p' all the other ICs will hold as well. We can therefore focus on the percentiles just above and below p. The IC for the adjacent $p' = p + \epsilon$ can be further simplified if Y is differentiable in p. Divide equation 1 by ϵ and let $\epsilon \to 0$.

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

$$w'(p) = Y_p(S_i, p) \tag{4}$$

The IC condition can thus be written as a condition on the slope of the wage schedule. This function pins down the difference in wages when moving along the talent distribution. The superstar model therefore offers a tight prediction about the shape of the wage profile.

To fully characterize the equilibrium we require a second condition that determines which worker gets assigned to which firm. Denote the assignment by $S_i = \theta(p)$. As hinted to above the superstar model generates positive assortative matching. This is a result of the single crossing condition $Y_{Sp} > 0$. An extra unit of talent is worth most on the largest stage and hence the greatest talent will work on the biggest stage. For notational convenience let i represent the percentile of the stage in the distribution of stage sizes. Positive assortative matching implies $i = p$.

We can use these equilibrium conditions to analyze the forces that generate wage dispersion in the superstar model. To see what is driving changes in the talent premium, let us assume that production is multiplicative by $Y(S_i, p) = \pi \cdot S_i^{\gamma} \cdot p$, with π the price for a unit of talent.[2] An alternative way of thinking of the productivity of a firm is the marginal revenue product per talent unit ($\pi \cdot S_i^{\gamma}$) at firm i. Productivity thus depends on the characteristics of the firm (S_i) as well as the prevailing price of skill (π). Along with the literature i will refer to S_i as market size, however this is a somewhat misleading label. S_i is an employer specific primitive that captures firm specific attributes that increase the productivity of a talent unit. Unlike talent these attributes cannot be moved across firms. Arguably we can capture such firm attributes by looking at a measure of primitive firm size. Note though, that this is not the same as the observed firm value. The value of the firm is given by Y and also reflects the employee talent and the price of skill. In the empirical section I will be able to distinguish a change in S_i from the endogenous firm value Y.

We can now define the wage distribution. Substituting the production function into equation 4, the slope of the wage schedule becomes:

$$w'(p) = \pi \cdot T'(p) \cdot S_p^{\gamma} \tag{5}$$

Wages at percentile p are found by integrating equation 5. Wages are pinned down up to a constant that represents the reservation wage of the least talented worker that participates b. For simplicity I will normalize $b = 0$. Wages are given by:

$$w(p) = \int_{\bar{p}}^{p} \pi \cdot T'(j) \cdot S_j^{\gamma} dj \tag{6}$$

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

$$D(\pi) = \int_{\bar{p}}^{1} T(j) \cdot S_j^{\gamma} dj \tag{7}$$

since the right hand side is increasing in π, a downward sloping demand curve will ensure

[2] A multiplicative production function will make the exposition easier, but the results go through more broadly.

that there exists a market clearing skill price. Growth in S makes talent more abundant and leads to a falling price of skill. Growing scale of operation at the top will therefore reduce $\pi$and decrease wages for all other entertainers. We assume that the only supply response is on the participation margin, $\bar{p}$ increases when πfalls. The model has however been extended to include an hours response Scheuer and Werning [2017].

To analyze wage changes we need more information on the functional form of the firm size and talent distribution. Gabaix and Landier [2008] show that we can solve for top incomes, given mild assumptions on the talent distribution. They noted that for "regular" continuous distributions the tail of the distribution can be approximated by: $T'(p) = -Bp^{\beta-1}$. Further they assume that the market size primitive is Pareto distributed with distribution $F_s(S_p) = 1 - 1/AS_p^{-1/\alpha}$, with α the shape parameter of the Pareto distribution. A bigger value of α implies a fatter tail of the distribution, with larger markets. Assuming a Pareto size distribution has a number of advantages. The Pareto distribution fits the tail of many distributions remarkably well, a result that has become known as "Zipf's law". For firm size this has been documented in Axtell [2001] and Fujiwara et al. [2004]. Second, the Pareto destribution allows for a tractable solution to the model. An alternative to making an assumption on the market size distribution would be to model it. This is the path pursued by "differential rent models." Geerolf takes this approach and shows that under plausible assumptions the market size follows a Pareto distribution (see Geerolf [2016]). The assumption therefore appears well grounded. Using the two distributional restrictions we can solve for the wage distribution at the top. After substituting the distributional equations into equation 6, we find the distribution of top wages:

$$Pr(w > w_p) = \frac{\alpha\gamma - \beta}{N} w_p^{\frac{-1}{\alpha\gamma-\beta}} \tag{8}$$

where $N = \pi A^{\gamma} B$. Re-arranging yields the wage schedule for top incomes:

$$w_p = \frac{N}{\alpha\gamma - \beta}(1-p)^{-(\alpha\gamma-\beta)} \tag{9}$$

Figure 2 plots the wage schedule for a case with superstar effects ($\gamma > (1+\beta)/\alpha$). The economic substance of the superstar assumption is that decreasing returns are small enough to avoid that wages flatten out at the top. This holds if γ is large enough.[3] In that case wages increase as we move up the top tail of the distribution and do so exponentially. Top paid individuals thus make substantially more than slightly inferior workers.[4]

[3]In the general model the assumption required for superstar effects is $-Y_{pp}/\frac{\partial S_i}{\partial p} > Y_{pS}$.

[4]We saw that a fragmented market structure will lead to less top income inequality. As it becomes feasible to serve bigger markets, the wage-talent profile will become steeper. We can show this formally by evaluating equation 4 at the equilibrium values and differentiate with respect to S:

$$w_{pS}(p^*) = Y_{pS}(p^*) + Y_{pp}(p^*)\frac{\partial p^*}{\partial S} = \frac{w''(p^*)}{\theta'(p)} > 0 \tag{10}$$

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. This effect isn't second order.

Next, consider the effect of technology on the wage distribution. The driving force for wage changes are shifts in marginal revenue product. Above we saw that the productivity depends on the market size and the skill price. Technical change that make it feasible to reach bigger markets will generate growing superstar effects. A tractable way of modeling such a change is an increase in the shape parameter. Maintaining the Pareto distribution allows us to illustrate the impact of technical change in a simple closed form solution. An alternative but notationally more cumbersome approach allows firm size to change exclusively at the top. The stylized results are however the same and I therefore focus on a shift in α. Define the new shape parameter $\alpha' = \alpha * s$ with $s > 1$. By substituting α' into equation 9 we can see that the wage schedule becomes:

$$w_p^{t+1} = \frac{N^{t+1}}{(s * \alpha\gamma - \beta)}(1-p)^{-(s*\alpha\gamma-\beta)} = w_p^t d(1-p)^{(s-1)\alpha\gamma} \tag{11}$$

superscripts $t+1$ and t indicate wages before and after the market size change. The resulting change in the wage schedule is illustrated in Figure 2. There is two effects, for one, top entertainers can sell more units of skill through the new technology. In the wage equation this effect is captured by the growth in the exponent. As p gets close to 1 the effect of the exponent is magnified. The top of the distribution therefore sees the biggest wage growth, returns to being a superstar have risen.

At the same time, the greater abundance of skill has reduced the skill price ($\pi^t > \pi^{t+1}$).[5] The revenue product of any given venue declined. If the audience capacity of an entertainers' job is unchanged, he will therefore see productivity decrease. The wage for any given venue declines. The downward shift is the result of the falling skill price and greater dispersion in market size. The second effect is a result of the simplifying assumption that maintained a Pareto distribution. Growing α fattens the tail at both ends of the distribution. Changes at the bottom feed through to a lower level of compensation throughout the wage distribution and amplify the effect of a falling skill premium. The wages decrease is given by factor $d \equiv \frac{\pi^{t+1}}{\pi^t}\frac{(\alpha\gamma-\beta)}{(s*\alpha\gamma-\beta)} < 1$. Taking these effects together the wage schedule has shifted downwards and pivoted it upward. Workers at the top are therefore benefitting from the technology at the expense of the rest. An immediate implication of this change in the wage schedule is that incomes have become more concentrated at the top. Within the top tail, the last fractiles will see particularly large gains in their income share.

2.1 Superstar Model vs Alternative Models

Next consider the kind of wage change generated by a skill biased demand shift. Skill biased demand shifts (SBD) have been widely applied to study the US wage distribution

We can sign the equation without appealing to the envelope theorem as long as the assignment function is invertible. A unit of talent thus becomes more valuable as S increases.

[5]The intuition for this result is that the average market size increases with α. Formally, we can look at the RHS of equation 7. It is now given by $\int_{p^{res}}^{1} \lambda p^{\beta-\alpha} dp = \lambda[1-(p^{res})^{\beta-\alpha+1}]/(\beta-\alpha+1)$ which is increasing in α(and λ is a constant). The skill price πhas to fall to bring the market into equilibrium.

and estimate the impact of technical change on the wage distribution (Acemoglu and Autor [2011], Akerman et al. [2013], Katz and Murphy [1992]). In it's simplest form the model features only two types of skill, a low and a high skill groups. Such a simple two skill group model is silent on top income dispersion. A more interesting extension of the model allows for workers with different quantities of skills. Workers with greater skill will earn higher wages and the model thus features income dispersion at the top. And SBD will generate rising top income inequality.

Since we are interested in the top tail of wages consider an economy with only one skill group. This allows us to study SBD in the same framework we used above to study superstar effects. Think of the quantity of skill as the amount of talent (T_P) a worker at percentile p possesses.Wages are still given by equation 6. A skill biased demand shift increases demand for skill $D(\pi)$, consider an increase in the demand for skill $D'(\pi) > D(\pi)$. From equation 7 it follows that the price for skill πincreases to π'. This leads to a first distinctive prediction. Unlike in the superstar model the skill price is rising in the SBD case. As a result we should observe employment *growth* in the SBD framework, while in the superstar model we would observe employment losses.

The increase in the skill premium also generates differential wage gains across the distribution, with the largest gains at the top. To see this differentiate the wage equation 9 with respect to π. The derivative is increasing in (1-p) and wage gains are thus largest at the top. The prediction of the SBD model differ however from the results of a superstar model. To see this consider the case where the model matches the initial wage distribution illustrated in figure 1 and the increase in the skill premium π'/π matches the growth in top incomes. The change in wages from a SBD is illustrated in figure 4. Wage rises at the top, yet the implications for the other percentiles differ from the superstar predictions. In the SBD model we observe increases at all percentiles.

Differences between SBD shifts and superstar effects can be tested. The key property that distinguish SBD changes from superstar effects is the perfect substitutability of talent within a skill group. The skill premium π therefore is constant across all p. The law of one price dictates that there is a single market clearing price for talent. In a superstar model on the other hand quality and quantity aren't perfect substitutes. A more talented worker cannot be replaced by two workers with the same aggregate amount of talent. In a superstar model wages can therefore differ by more than the gap in talent.

The difference between the two models has strong implications for there respective predictions on top income inequality. The effect of a SBD shift can be seen from equation 6. A positive demand shift will shift all wages up. A SBD shift is therefore a tide that lifts all boats. Moreover, wage growth rate will be the same at all percentiles. A person who initially earns twice as much, will do so again after the skill premium rises. The SBD model can generate bigger gains at the top, but does not generate skewed wage *growth*. To illustrate this point, I will abstract from exit and entry and assume talent remains unchanged.[6] Wage growth follows from equation 11. The wage growth in SBD and superstar models (SM) are

[6]This makes little difference for the top of the distribution. Entry has negligible effects on wages and the results thus would carry through approximately. At the bottom of the distribution entry and exit would matter.

given respectively by

$$g_{w_p}^{SBD} = \frac{\pi' * \int_{\bar{p}}^{p} T'(j) \cdot S_j^{\gamma} dj}{\pi * \int_{\bar{p}}^{p} T'(j) \cdot S_j^{\gamma} dj} = g$$

$$g_{w_p}^{SM} = d(1-p)^{(s-1)\alpha\gamma}$$

Wage growth in the SBD model is independent of percentile p. In the superstar model wage growth varies with p, and is particularly strong at the top. Contrast this with the superstar model, here wages grow strongest at the top. We can thus distinguish a superstar model from a SBD framework by testing whether wage growth rates are the same across the top.[7]

A final test looks at income dispersion *within* the top tail, where SBD shocks produce very limited change. To see this consider the share of income going to a top fractile. This metric has been used extensively to document growing inequality at the very top ?Piketty and Saez [2003]. Over the last decades the US has seen large increases in the share of income going to the top 1%, 0.01% and 0.001% of the distribution. The largest growth occurred among the top 0.001%.

This latter fact will not occur in an SBD model. The SBD model yields growth in top income share at percentile p (g^{s_p}) of:

$$g^{s_p} = \frac{s_p^{t+1}}{s_p^t} = \frac{\int_p^{\infty} w_i^{t+1}/Y^{t+1} di}{\int_p^{\infty} w_i^t / Y^t di} = \frac{Y^t}{Y^{t+1}} \frac{\pi' \int_p^{\infty} \int_{\bar{p}}^{p} T'(j) \cdot S_j^{\gamma} dj}{\pi \int_p^{\infty} \int_{\bar{p}}^{p} T'(j) \cdot S_j^{\gamma} dj} = \frac{g}{g^Y} = g^s$$

The second equality uses the definition of top income shares, with Y total income in the economy. Next I use the fact wage equation 6. The next steps collect terms and cancel. 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. This is a result of the perfect substitutability of talent in models of SBD. The law of one price prescribes that top wages at all percentiles grow at the same rate, leading to constant top income shares in a SBD model.

Now consider models of non-competitive pay setting. Here wages are not pinned down by worker productivity but set through a wage bargaining process that splits surplus among workers. To the extend that exogenous changes in technology are orthogonal to bargaining power we will be able to separately identify the superstar effect from the effect of bargaining power shifts. To further lend credibility to the idea that changing bargaining power is not driving the result we can test if wage changes arise absent changes in productivity. The rent capture or norm hypothesis is that changes in wages are the result of shifts in wage bargaining, while relative productivity is unchanged. In the superstar model on the other hand pay is linked to the marginal revenue product of workers. Shifts in marginal revenue product are the result of greater demand for top talents, making star workers more productive. We can distinguish the superstar effect by testing if demand does indeed shift

[7] The wage results for the SM are derived for the top tail. Tests will therefore focus on the upper halve of the distribution.

towards star workers.

3 Empirical Strategy

An ideal test would compare the income distribution in the same labor market with different production technologies. While in practice we can't observe literally the same market in two states of the world, we will be able to use quasi-random variation in the availability of the technology to get close to such a test. The baseline analysis uses a difference in difference strategy. Consider the the set of US local labor markets $m = 1, ...M$. The simple (naive) OLS regression is given by:

$$Y_{mt} = \alpha_m + \delta_t + \gamma X_{mt} + \beta S_{mt} + \epsilon_{mt}$$

where Y_{mpt} is the outcome for market m in year t (e.g. the share of entertainers with wages in the top 1% of the US wage distribution), α_m and δ_t are labor market and year fixed effects, X_{mt} a vector of time varying labor market characteristics and S_{mt} is the audience size that an entertainer can reach while working in labor market m in year t. The coefficient of interest is β that captures the response of the wage distribution to market size changes. We can trace how changes effect different parts of the distribution by running this regression for different parts of the wage distribution.

The regression considers different regions as separate local labor markets. In practice this may not hold exactly as workers may move between labor markets. It is therefore worth considering how violations of this assumption may affect the result. Take the extreme case where workers are perfectly mobile. Here the stylized results of the superstar model remain the same but at the top they would be amplified by inward movement of talent. The observed effect than captures the wage adjustment to the relative scarcity of talent, but also reflects adjustments in the local ability distribution. In practice mobility may not be a big concern. In the full mobility benchmark case, the most talented entertainers start out in the labor market that offers highest rewards for top talent initially. These are places with the largest performance audiences. It turns out that these locations are also the places that are predominantly affected by TV, we thus wouldn't observe substantial mobility. Even in the full mobility benchmark case mobility should not substantially affect the results. Moreover, workers are less mobile than the full-mobility benchmark suggests, at least in the short run Blanchard and Katz [1992], Autor and Hanson [2013].

3.1 Testing the Core Predictions of the Superstar Model

Superstar effects have differential impacts on different parts of the distribution. Recall from above that we ought to observe growing returns at the top and falling wages for ordinary workers (see 2). We need to turn this into testable predictions. The full wage distribution of local entertainer labor markets is not observed. Top coding and data scarcity make it infeasible to estimating the shape of the full distribution for all local labor market before and after TV non parametrically. We would have to estimate over 3,000 distributions. The empirical analysis takes therefore two alternative approaches. We can capture the essence of the shift in the wage distribution with a number of non-parametric tests. Four predictions capture the core wage shift implied by the superstar model:

1. More workers in the affected occupation become top earners
2. Fewer workers earn mid-income
3. A larger low pay sector emerges
4. Fewer people are employed

To illustrate the intuition for these tests consider the first test, "more workers in the affected occupations become top earners." We can test this prediction by looking at the share of entertainers who are from a treated labor market and make it into the top of the overall US wage distribution. The test is illustrated in figure 3. The share of entertainers who cross the top-earner threshold ($w^{US1\%}$) has risen by $\Delta E_{1\%}$. This test is feasible for any data with top-coding above the top income threshold. The test is similar to a quantile regression in panel data which analyses the wage threshold that divides workers into a group of top earners and non-top earners. My specification captures the same idea, but can be implemented with a simple OLS estimator. Similar outcome variables can be constructed for the other three predictions.

The second approach makes a parametric assumption on the shape of the wage distribution. The shape of the tail of the wage distribution has received much attention in the literature. In Kuznets' pioneering study in 1953 he finds that the US wage distribution can be well approximated by a Pareto distribution. An overwhelming number of studies have since replicated this finding for a variety of time periods and countries (see Atkinson2011a for an overview). Moreover, given the model assumption above, the wage distribution will follow a Pareto distribution exactly. In a Pareto distribution top income dispersion is characterized by a single parameter. The appealing feature of the Pareto distribution is that we can estimate this parameter in top coded data and we can estimate it independent of the size of markets and samples.

3.2 Staggered TV Roll-Out

Since S_{mpt} might be endogenous, I instrument for the available audience size. I use the presence of a local TV station as instrument. To fix ideas, consider first the simple difference in difference case. Here there are only two scenarios: either a local labor market has a local TV station or not. The dummy $D_{mt}^{TV} = 1$ if a TV station broadcasts from the local labor market. The framework can be easily extended to accommodate more nuanced measures of TV filming activity. The first stage can than be written as:

$$S_{mt} = \alpha_m^F + \delta_t^F + \gamma^F X_{mt} + \beta^F D_{mt}^{TV} + \epsilon_{mt}^F$$

where the fixed effects are defined analogously to the previous equation. The specification allows for differential effects of TV on different parts of the distribution. We might for example expect that TV effects a mediocre entertainer differently to one at the top of the distribution. Historical sources suggest that this was indeed the case. TIME magazine writes about the landmark change in entertainment in a cover story in 1949:

"As the clock nears 8 along the Eastern Seaboard on Tuesday night, a strange new phenomenon takes place in U.S. urban life. Business falls off in many a nightclub, theater-ticket sales are light, neighborhood movie audiences thin. [...] For the next hour, wherever a signal from an NBC television transmitter can be picked out of the air, a large part of the population has its eyes fixed on a TV screen" (TIME - May 1949) Television appears to concentrate demand on star entertainers.

The above regressions treat the introduction of a TV station as a shock to the *local* entertainer labor market. From a modern prospective this may seem puzzling. Today's TV stations predominantly broadcast content produced elsewhere. During the early period of TV this was different. Early TV shows were predominantly filmed in the vicinity of broadcasting antennas and broadcast live. A number of constraints made local production very common. For one, storing and transporting shows was costly and led to inadequate image quality.[8] Second, the infrastructure to broadcast live from elsewhere was not yet in place. Sterne [1999] gives a detailed account of the pioneering work on building a national TV network and the major technical obstacles. Finally, the regulator imposed restrictions on the location of TV studios. The annual report outlines the position of the regulator "the rules require that the main studio be located in the principal community served" (FCC annual report 195). Non-local shows were thus a poor substitute for local productions. As a result early TV stations introduced a demand shift for local star entertainers.

To identify the effect of television the analysis exploits regional differences in the location and timing of TV entertainment production. Television was introduced in a slow and staggered fashion across US local labor markets. The launch date of the first Commercial Television stations was July 1, 1941.[9] The regulator drew up a road map for the roll-out and specified local media markets for which licenses could be obtained. World War II interrupted the production of receiver sets and television remained a niche technology for the first 5 years. After the war the technology started to spread rapidly. By 1949 the number of active stations had reached 124. Figure8 illustrates which local labor markets had been treated at this point. The majority of the country however remained without access to television, only by 1960 was the entire country covered by television signal.

An valuable asset for the analysis is that TV stations could not spread freely. A licensing process determined where the technology could be set up. The aim of the licensing process was to avoid interfere with existing infrastructure. Broadcasting took place over

[8]Non-local content had to be put on film, shipped to other stations and finally a mini film screening had to be broadcast live. This technology, known as “kinescope”, resulted in poor image quality and was therefore unpopular. Moreover, it was costly. There are notable exceptions. A handful of stations, mainly along the east-coast, experimented with various forms of interconnection (microwave relays, stratospheric broadcasting and coaxial cable connections).

[9]Exploratory broadcasting experiments, technical showcases and experimental broadcasting had taken place in a few locations and on road sows since the 1920s and had familiarized the population with the new technology. In four cities experimental broadcasters where later turned into commercial television channels. Prior to the launch of commercial television the private ownership of TV sets was however minimal.

airwaves and the signal could potentially interfere with military communication, air-traffic control, maritime communication or other TV stations. To avoid interference the regulator determined locations and catchment areas that were considered safe for use. This gives rise to plausibly exogenous variation in the location and audience size of a TV station. The arising regional differences in TV access span a remarkable 15 years. Since labor markets get access to the and gives ample variation to test the effect of television. The variation can be analyzed using the following reduced form regression:

$$Y_{mt} = \alpha_m^R + \delta_t^R + \gamma^R X_{mt} + \beta^R TV_{mt} + \epsilon_{mt}^R \tag{12}$$

The coefficient β^R captures the effect of TV on the wage distribution. Equation 12 forms the baseline of the analysis. To test the predictions of the superstar model we will want to evaluate β^R. As before the regression can be run for different parts of the wage distribution to analyze differential effects across the distribution. TV_{mt} is a measure for the exposure to local TV production in market m at time t, this will be either the number of stations active in the locally, the audience served via TV from the local market or a dummy for local TV access. As before the regression controls for labor market and time fixed effects. These fixed effects absorb time invariant differences across labor markets. The fact that labor markets that receive TV are different from the control group is thus not per-se a problem for identification. The identifying assumption of the difference in difference model is that the timing of a TV station is unrelated to confounding changes in the local labor market. The strong involvement of government in the roll-out makes this assumption more plausible. The licensing procedure specified a deadline for the start date of TV stations. Failure to comply could result in a withdrawal of the license. This left little room to deviate from government dictated roll-out schedule.

3.3 Variation from Regulator Freeze

We can add further credibility to the estimation by exploiting plausibly random variation that arises from a regulator shut-down and technical innovation that made local productions redundant. Both these shocks exclude local entertainers from the TV treatment. Consider first the regulator shut-down. The licensing process is overseen by the Federal Communication Commission (FCC). Based on a signal propagation model FCC had delineated media markets where stations were guaranteed inference free broadcasting. It turned out that the signal propagation model was erroneous and the designated media markets led to signal interference between stations. As interference issues became clear in 1948 the FCC stopped the licensing of any new stations and started the re-design of the assignment process. Figure 7 shows the number of approved applications over time. The shut down in the approval of applications starting in October 1948 stands out notably. The sudden shut-down of licensing became known as "the freeze." Manipulation of application dates can be ruled out as the blocked applications date back substantially before the freeze of application was discussed. Stations already licensed continued to operate. I digitize detailed data on the licensing process. This data includes information on pending applications and allows me to identify local labor markets that were excluded from TV by the FCC review. Figure 9 shows which local labor markets are affected by the freeze. During the freeze period the FCC undertook extensive field studies and expert hearings to improve the scientific standard of their signal model. Licensing did not resume until 1952.[10] The "freeze" of applications thus delays the onset of TV nearly 4 years in many markets. This will give rise to a placebo test. This test uses the same specification as in 12, comparing only untreated areas and replacing TV_{mt} with an equivalent measure for placebo stations ($TV_{mt}^{placebo}$):

$$Y_{mt} = \alpha_m^P + \delta_t^P + \gamma^P X_{mt} + \beta^P TV_{mt}^{placebo} + \epsilon_{mt}^P \tag{13}$$

Looking at stations that were banned from starting operation allows us to test the identifying assumption that license timing is exogenous to local labor market conditions. This tests the hypothesis that "Stations that didn't happen" due to the FCC "freeze" have no effect on local entertainer wages ($\beta_p^P = 0$).

A second quasi-experiment uses the demise of local TV production. The invention of the Ampex video tape made it possible to store and transport TV productions cheaply, this transformed the TV production industry.[11] The introduction of the videotape finally made productions from outside the labor market a close substitute for local live shows. This led to the demise of local TV productions and the concentration of TV show production. The video tape proved an instant hit with TV stations. When the product was presented at the National convention of broadcasters in 1956, over 70 videotape recorders were ordered immediately by TV stations across the country. The same year CBS started to use the technology and the other networks followed suit the following year. TV production became concentrate in two hubs, New York and Los Angeles. Local entertainers in those centers can now reach even bigger audiences, however the choice of these two places is likely endogenous. The two places were likely chosen because production was particularly attractive there. To avoid an endogeneity issue I will use a pre-determined proxy for differences in local production cost. This proxy is based on the location of movie productions in the period prior to TV. The measure is thus orthogonal to contemporaneous changes in those labor markets. The rise of the two production hubs resulted in the demise of local TV shows. This provides a powerful test for the analysis. We should observe that the effect of TV station licenses disappears after the adaption of the videotape. We can therefore test if differences between control and treatment areas are constant by comparing those areas without treatment before and after the period of local TV. The test thus goes beyond the conventional parallel trend check and relies both on the pre- and post-treatment period.

[10] Initially the freeze was only expected to last about a year. However, additional technical developments prolonged the freeze period. Beside reconsidering the assignment of existing frequencies, the FCC started to experiment with making additional frequency bands available to television. Moreover, the FCC wanted to ensure that the new system was compatible with the arising transmission of colored images. It thus bundled the testing and processing of these issues.

[11] The World Intellectual Property Organization describes the innovation in here: www.wipo.int/wipo_magazine/en/2006/06/article_0003.html

3.4 Variation in TV Signal

To analyze the effect of television on the skill price it will not be sufficient to look at the local labor market where TV is produced. TV signal travels beyond the local labor market and affects the supply of talent in neighboring labor markets.[12] In areas that receive TV signal local entertainer face competition from TV shows. Labor markets exposed to TV signal from outside their own market allow us to identify the effect of TV on the price of talent. To estimate this I measure where people can watch TV. The data on signal comes from an irregular terrain model (ITM). This model calculates the propagation of airwaves from information on technical antenna features & information on topography.[13] Depending on the terrain and the technical features of the antenna the signal reach can vary from a few miles to around 100 miles. Hilly areas, for instance, reduce the catchment area of a TV station. Figure 13 illustrates areas reached by television signal in 1950 in blue. While areas that would have been served by stations that were held back by the regulator "freeze" are shown in red. Signal is difficult to target and many areas that weren't aimed for became "treated" nevertheless. The set of areas that experience greater competition from technology are thus diverse. Many areas at the periphery of the signal catchment area have very similar characteristics to areas that are just outside the reach of the antenna. Since signal is difficult to target, treatment at the fringe is mainly due to topographic features and thus plausibly exogenous. We can use this variation to test how much expenditure on local entertainment declines and how much employment lost due to the introduction of television. Estimates of the change in expenditure allow us to identify the shift in productivity induced by TV. We can than observe how these productivity chances translate into labor market outcomes by looking at changes in employment. The specification of this test is equivalent to equation 12, except that TV_{mt} is replaced by a dummy for signal access in local labor market m at time t. As above we can probe the validity of the identifying assumption by running placebo checks with frozen stations.

4 Data

The analysis combines data from archival sources. The data are hand collected and turned into a dataset in a large scale digitization process. The data covers the production and consumption of entertainment in the 1940s and 1950s. It combines information on consumer spending, local TV stations, their stars, physical performance venues and TV signal propagation, as well as labor market information from the Census and tax records on entertainers.

Census Data

The Census collects information on the US population every decade. The data cover entertainers from the entire country and thus allow me to analyze changes in local labor markets for entertainers. This spans over 700 local labor markets. I use Census data from 1920-1970

[12] In such neighboring labor markets entertainment from outside the labor market will be consumed and the demand for talent thus doesn't equal the local supply for talent. This is analogous to a trade inflow.

[13] the relevant parameters are height, power, frequency

as provided in the the Integrated Public Use Microdata Files (IPUMS Ruggles et al. [2017]). For 1920-1940 the data covers the full population of US residents and for later years a representative sample of the population.[14] The period covers the entire roll out period and substantial pre and post periods. The Census data on wages, occupation and employment is available consistently for individuals over the age of 15. I therefore restrict my sample to that age group.

I followAutor and Dorn [2013] and define a labor markets as urban centers together with their respective commuters belt.[15] I extend the period to link labor markets with Census data going back to 1920. For each Census I use the smallest available location information and map it into a commuting zone. This makes it possible to analyse consistent labor markets spanning the mainland US throughout time. There are 722 commuting zones covering the mainland USA.[16] On average a commuting zone has about 400,000 inhabitants and ca. 500 workers in entertainment occupations.

The core of the analysis focuses on occupations that could appear on television. I identify workers in entertainment using three digit occupation information in the Census. The relevant occupations in the census are: Actors, athletes, dancers, musicians and entertainers not elsewhere classified. The last group is relevant because it includes most circus and vaudeville actors, one of the most important forms of entertainment at the time.[17]

In many settings the reclassification of occupations over time poses a problem. Entertainment occupations however are well established and there is little change to their definition throughout the sample period.[18]There are a few exceptions, most relevant for the above groups is that the Athlete category is discontinued in 1970. To account for such time shifts in the occupation definition, the regressions will control for occupation specific time effects.

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

Part of the analysis will introduce additional occupations as control groups. I will analyze whether traditionally high paid occupations respond to TV too. The relevant occupation

[14]Changes in variable definitions prevent an extension of the time period further. From 1980 onwards the Census uses different occupation groups that make comparisons with previous periods difficult. Prior to 1930 a significant change in the definition of employed workers occurs and makes comparisons of employment hard.

[15]These areas are defined using strong within labor market commuting and weak across labor market commuting.

[16]These regions are consistently defined over time. Dorn provides crosswalks for 1950 and 1970. I build additional crosswalks for the remaining years.

[17]The original string occupation title is available in the 1940 Census and confirms this. The category includes acrobats, clowns, animal trainers etc. Unfortunately the original job-title is not available for all sample years and thus doesn't allow me to divide entertainers not elsewhere classified further.

[18]The 1950 occupation classification system was used in the 1940-1970 censi. For the years 1920 and 1930 IPUMS has mapped the original answers into the 1950 categories.

groups are finance, accounting and law, medics, engineers and managers.[19]

For each of the labor markets I calculate outcome variables at the year - occupation level. The outcome variables are: employment in entertainment occupations, share top earners and share at other percentiles bins of the national wage distribution.

TV data

For each labor market I compute two metrics of exposure to TV: television filming and television broadcasting. The first captures the change in the production technology and records where TV shows are produced. Second, I build a variable of local TV signal, which measures where local entertainers face competition from television.

Television filming

Data on Television facilities come from the "Annual Television Factbooks." The Factbooks record the address, technical equipment, launch date and the assigned channel and call letter for each TV station.[20]I geocode the location of the studios and match them to the local labor market. This allows me to track the roll-out of TV across local labor markets. The location of TV studios is one of the main sources of variation in the analysis. Figure 9 shows where licenses were granted in 1949. [21]To trace the roll-out of TV production I match each station to its local labor markets and trace the number of active stations. Moreover, I identify local labor markets that are affected by the freeze in licenses. An area is coded as treated if the station was operating at the beginning of the year, I assume that all stations are filming locally at that time. There are a handful of exceptions, a few stations operated a local network. These interconnected stations could relay local shows to nearby stations through upgraded phone lines (run by AT&T) or microwave relay technology (run by Bell). Interconnections was rarely feasible because the technical infrastructure was still in its infancy. In my main specifications I code all members of such networks as treated. This approach avoids potential endogenous selection of filming locations within the network.[22] Television production shifts away from local station with the invention of the videotape. Production will take place where conditions for filming are most favorable. Such marginal cost of production are determined by endogenous factors such as local wages negotiations and tax rates as well as fixed location characteristics (e.g. sunshine hours, availability of equipment and expertise and local scenery). To avoid that changes in local labor markets drive the location decision I construct a measure of production cost that is pre-determined. Movie filming did not suffer from the location constraint, the location of cinema filming in the 1920s therefore captures pre-determined local production cost. The data on the

[19]The appendix gives further details on the occupation codes for these groups.

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

[21]The year 1949 is of particular interest because TV shows are produced locally at the time and it is a year when wage data is available in the Census.

[22]Robustness checks explore alternative treatments. As expected, within those networks effects on top incomes show up in the labor market where filming is mostly located.

location of film shoots comes from the "Internet and Movie Database" (ImDB). ImDB is a widely used platform (self-proclaimed #1 world wide) for information on movies and holds metadata on over 4 million movies. In 1920 around 200 movies were produced in the US. For each labor market I compute the share of movies produced in this market in 1920.

TV Licensing

I collect data on the licensing process of television stations. The data captures which places are about to receive a TV license. I collect this data at the onset of the license freeze and identify places that are affected by the freeze. The data is based on the weekly bulletins by Federal Communication Commission (FCC), which are summarized annually in the "Television Factbook." This data is supplemented by information on the roll-out formula used by the regulator. The roll-out rule prioritizes areas with the largest population, at least 40 miles from the nearest station. I digitize the order in which frozen places will be processed after the freeze. The ranking in the priority list of the regulator is the running variable in the regression discontinuity introduced by the license freeze.

Television broadcasting

TV airwaves travel beyond the local labor market where shows are filmed. I use information on TV signal to calculate audience size of TV shows. Signal reach of TV stations comes from Fenton & Koenig who re-constructs historic catchment areas based on historic technical antenna features. It uses an irregular terrain model (ITM) to calculate signal propagation. In this model signal reach depends on the technical features of an antenna (channel, frequency, height etc.) and on terrain that blocks airwave travel (e.g. mountains). For the details on the data construction see Fenton & Koenig. Based on the signal data I compute the potential audience size of a TV stations. I combine the information on the catchment area of each TV station with Census data on household location and TV ownership. The match is performed by mapping information on TV signal into geographic units available in the Census. The geographic match uses the boundary shape files provided by NHGIS (Manson et al. [2017]). I than identify how many TV households fall in the catchment area of each TV station. This allows me to construct a measure of potential audience size. The median TV station could reach about 75,000 households. Even the smallest TV audience exceeded the show audience of local venues substantially. [23]

[23]Television data used in this study is distinct to the data used in earlier studies of TV. Foremost, this study focuses on the production of entertainment while previous studies focused on the effect of consuming TV Durante et al. [2015], Gentzkow and Shapiro [2008], Gentzkow [2006], Chong and La Ferrara [2009], Enikolopov et al. [2011]. The data on signal transmission also differs from previous studies of the US. Gentzkow [2006] for instance defines the catchment area of a TV station by it's current media market. This study instead uses the data collected by Fenton & Koenig that re-constructs historic catchment areas based on historic technical antenna features.

Theatre data

Audience information from the pre-TV era comes from a historic companion book for the entertainment profession, the 1921 "Julius Cahn-Gus Hill theatrical guide." I hand collect data on over 3,000 performance venues, which cover more than 80% of US local labor markets. I collect data on the seating capacity of these venues. On average a performance venue has 872 seats, but capacity varies between a few hundred seats to several thousand. The most iconic performance venue was the New York Hippodrome which was hailed as the "world's largest theater." The seating capacity of over 5,000 seats was substantially bigger than most such venues. On average the capacity of the largest theatre in a labor market had 1165 seats.

Taxable Income

I digitize information on earnings of entertainers from historic tax tables published by the Internal Revenue Service (IRS). These tables list how many people received an income in each income bracket. Such tables have been used by Kuznets and Piketty to construct time series of top income shares for the US work-force. In 1916 the IRS publishes income table broken down by state and occupation. I digitize these tables for entertainment occupations. This data paints a detailed picture of the top tail of the income distribution in entertainment in each state. Unfortunately, a break-down by occupation and state is only available in 1916. For later years I rely on the US Census to measure incomes.

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. 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.

County Fairs

The "Cavalcade of Fairs" contains detailed information on county fairs and is published annual as a supplement to the Billboard magazine. This data allows me to measure spending and attendance at local entertainment outlets. Fairs provide a range of amusement activities, usually consisting of a Carnival with rides, food stalls and activities and a grand stand show with performances of local dance squats, music groups, sport competitions and similar highlights. I collect information on the ticket sales and revenues for over 4,000 fairs spanning 13 years (1945-1957) and the majority of US labor markets. For each year and local labor market I construct the total spending and attendance at local fairs. I record spending in two categories: fair ticket, show entrance (e.g. grand stand), carnival[24] (e.g. fair rides, merchandise and food). Demand changes can be analyzed separately for leisure activities that are differentially close substitutes for television.

5 Empirical Results

During the period of of TV roll-out the wage distribution in entertainment changed dramatically. Entertainment was an outlier during this period. For the economy as a whole top incomes were stable or even declining (This has also been documented in Atkinson and Piketty [2006]). Figure 6 plots wages at the 95th percentile relative to median wages. The P95/P50 ratio moved little for the economy as a whole. For actors on the other hand, wages at the 95th percentile grew substantially faster than median pay. The 95-50 ratio for actors grew by more than 80% during the period of TV roll out. Just as striking is the fanning out of wages within entertainment. The wage distribution for entertainers before and after TV is shown in figure 1. Three facts stand out, first, in line with superstar effects, a substantially larger share of entertainers is earning high wages. Second, many of the middle paid jobs have disappeared. Third, the income distribution has become more polarized with a greater share of entertainers earning low wages. There are two reasons why demand shifts are unlikely to explain this pattern. One, conventional demand shifts would lift wages across the board and thus would not produce falling wages at the bottom and growing wages at the top. Second, the described pattern holds for *log* wages. In models of skill biased demand wage changes are driven by the skill premium. The skill premium enters wages multiplicatively and thus for log wages results in a parallel outward shift. The log wage distribution would experience a right shift at all percentiles. Skill biased demand shifts are therefore unlikely driving the observed shift in the aggregate wage distribution. The superstar effect however is a plausible candidate. To estimate the effect of the superstar effect, we will now trace the impact of TV by comparing local labor markets. First we analyze the core predictions of the superstar model for the wage distribution. Second, we look at models that can generate similar predictions and test those models against the superstar effect.

5.1 Effect on Top Earners

Sharp income growth at the top of the distribution is the most prominent prediction of the superstar model. To estimate this effect we need to take a potential pitfalls into account. Falling employment could bias the analysis of the wage distribution. The model predicts that employment of non-stars decreases as they suffer from the reduced price for talent. If low paid entertainers exit, top wage percentiles move up even if wages are unchanged.

[24]This category is unavailable in 1953 and 1955

Take the case where the bottom 10% of the distribution stops working. Mechanically all percentiles move to higher wages and we would find growing wage percentiles without any change to wages. To avoid this bias, the analysis focuses on a different metric for top incomes. It tests how many entertainers in a local labor market are among the highest paid individuals in the country. The baseline analyses focuses on two alternative measures. One, the share of entertainers whose wage falls in the top 1% of the US wage distribution ($D^{US1\%} = 1$) and who reside in local labor market m. I denote this employment share in the top 1% by $e_{m,t}^{1\%}$:

$$e_{m,t}^{1\%} = \frac{\sum_i E_{i,m,t} \cdot D^{US1\%}}{\sum_m \sum_i E_{i,m,t}}$$

where $E_{i,m,t}$ takes value one if an individual i works in entertainment and works in makret m at time t. Note that the denominator is the same across labor markets and thus amounts to a normalization.[25] This avoids that fluctuations in local labor market employment drive the findings (Appendix **??** shows that the effects are indeed driven by the numerator).[26] A further benefit is that the measure it is unaffected by top coding, as long as the top code binds above the 99th percentile of US wages. During the sample years this is the case and the top code therefore does not pose a problem. The top code does however restrict us from analyzing even higher percentiles, such as the top 0.1%. Second, the share of entertainers from labor market m in the US top 1%. Call the entertainer fraction in the top 1% of the US wage distribution $Top_{m,t}^{1\%}$:

$$Top_{m,t}^{1\%} = \frac{\sum_i E_{i,m,t} \cdot D^{US1\%}}{0.01 \cdot EMP_t}$$

with EMP_t total employment.[27] The results hold for a wide range of alternative top income metrics (see Appendix **??** for robustness checks).

The baseline identification strategy uses the local TV roll-out described in section 3. This compares changes in the entertainer wage distribution across the 722 local labor markets over the 4 decades covered by the census. We will first focus on the reduced form effect of TV by estimating a version of equation 12. The treatment is $TV_{m,t}$ which counts the number of TV stations producing local shows in m at time t. To control for potential time fluctuations in the occupation definition I will run the regression at the occupation, labor market, year level and introduce occupation-year fixed effects. The regression equation is therefore given by:

$$y_{o,m,t} = \theta_{o,t} + \alpha_m + \beta \cdot TV_{m,t} + \gamma \cdot X_{m,t} + \varepsilon_{o,m,t}$$

The standard errors $\epsilon_{m,o,t}$ are clustered at the local labor market level so that running

[25] Since these fractions are small, the variable report shares per 100,000 entertainer.

[26] The superstar model predicts that the number of local entertainers declines, hence looking at the top percentile *within* entertainment would be upward bias.

[27] For readability fractions are multiplied 1,000.

the analysis at the occupation - labor market level therefore will not artificially lower my standard errors. $\theta_{o,t}, \alpha_m$ respectively capture region and occupation-time specific effects. The vector of controls $X_{m,t}$ controls for time varying characteristics in local labor markets. This includes a pre-determined measure of marginal cost of producing TV shows, interacted with a dummy for the time period of national TV (after 1956). This controls for the concentration of TV production after the invention of the videotape. The coefficient β captures the effect of the local TV show production. We now turn to testing the superstar model prediction β>0.

Before the introduction of TV the average of $e_{m,t}^{1\%}$ is 0.094 percentage points. This implies that in labor markets that become ultimately treated about 3% of entertainers are in the top 1% of the US wage distribution. Entertainment therefore had a sizable high-income sector even before TV. Conversely, the fraction of the US 1% highest earners that are entertainer is small. This is actually true for any industry, as a single industry share in overall employment is modest. The mean of $Top_{m,t}^{1\%}$ is 0.0004 percentage points. Prior to TV all entertainers in treated labor markets combined make up about 0.15% of US top earners. In the analysis I will multiply both variables by 1,000 for better readability, hence their pre-TV mean is respectively 94 and 0.4.

The introduction of a local TV station resulted in a marked increase top earning entertainers. Table 1 reports the results. The share of entertainers in the top 1% nearly doubles, with a point estimate of 90. The effect is highly statistically significant and also economically large. The share of entertainers in the top 1% goes from about 3% to over 10%.[28] Similarly, the fraction of the top 1% who are entertainers increases. The results are in line with the previous estimates, the share nearly doubles if a new TV station goes live. Top pay in the entertainment clearly outgrew the rest of the economy. Entertainers make up about 0.4% of the highest percent earners, up from 0.15%.[29] The fractions are small, but the growth rate an impressive 170%. These results confirm that local change in the production technology lead to vast gains at the top of the entertainer wage distribution, as predicted by the superstar model.

The identifying assumption is that that the introduction of TV is unrelated to time varying changes that affect top incomes in entertainment. The licensing driven roll-out process makes this plausible. To the extend that TV timing is beyond the control of market actors it should be largely exogenous to local changes in labor market conditions. To probe the validity of this assumption we can run the above analysis controlling for time varying characteristics in the local labor market. This sheds light on whether differential trends are driving the results. Including such controls has very little effect on the estimates (column 2 & 5). The treatment thus appears to be orthogonal to time varying characteristics in the local labor market. This confirms that the licensing process is not very responsive to changes in local labor market conditions. An alternative explanation for the stable treatment effect is that the controls are not picking up the relevant changes in local labor market

[28]The growth rate per TV station is (90+94)/94 and there are 1.78 stations per labor market.
[29]The growth rate per TV station is (0.389+0.4)/0.4 and there are 1.78 stations per labor market.

conditions. Time varying unobservable factors might be driving the observed results. To check if this is the case I allow for differential linear trends for each labor market. This is a very demanding specification as we add more than 700 additional regressors. The standard errors increase accordingly. If differential trends are driving the results we would expect the effects to disappear in this specification. The point estimates remain remarkably stable and significant despite the increased standard errors (column 3 & 6). We can thus rule out that different linear trends in treated and untreated labor markets are driving the results. This of course doesn't rule out that local unobserved characteristics change suddenly around the time of TV introduction. It may for instance be the case that new licenses coincide with an economic upswing in the local labor market that changes the local return to skill. We can go further and test whether the introduction of television coincide such location specific shocks to top pay. 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 panel B. The effect of TV is insignificant across all specifications. Moreover, the point estimate is very close to zero (panel B). The small point estimates are particularly remarkable since we are now looking at top-paid occupations which have a much higher baseline value. The implied growth rates are negligible. The baseline value imply that the placebo group initially makes up about 45% of top earners. TV leads to an insignificant growth rate of about 1.5%.[30] The difference to entertainment is therefore also economically meaningful. The effect on placebo occupations is about $\frac{1}{100}^{th}$ of the effect on entertainment. We can thus be confident that unobserved changes in the local skill premium are not the driver of these results.

Next, let us zoom in on the labor markets were the freeze in licensing delayed the launch date of TV. Recall that a large number of applications for TV stations were put on hold in 1948. The time pattern of approvals is shown in figure XYZ and shows the sharp drop in approvals. We can use this variation to add further credibility to our estimates. By comparing the change in labor markets affected by the freeze to the change in untreated labor markets we can test whether entertainer labor markets experience sudden changes in top incomes around the time of the planned TV introduction. This goes beyond the standard pre-trend check to test the identifying assumption of the difference in difference setting. It is also arguably more convincing than a test based on placebo occupations, as we can test for local shocks looking at the same occupations. The test is implemented with the equation 13. Estimating the equation we find that placebo stations ("stations that didn't happen") have no effect. The point estimate is a precise zero. Figure 10 illustrates the time path of the coefficients on "stations that didn't happen".[31] Placebo and untreated labor markets

[30] The aggregate growth rate is given by $g = \frac{0.45*(12.97+0.115*1.78)/12.97}{0.45}$

[31] At first glance the standard errors might look small here. They are indeed small, however note that treatment and placebo figures are plotted on the same scale for ease of comparison. Standard errors are

follow the same time path throughout the entire period. We can rule out that areas targeted for TV experienced any meaningful shock around the time of planned TV introduction. We can use this result to put a lower bound on the effect of TV. TV resulted in at least 86% growth in the fraction of top earners among entertainers.

A further useful safeguard is to ensure that there are no pre-trends. The idea of this test is to rule out that different trends in treatment and control group lead to a spurious results. Above we showed that trends are not driving the result, yet the a visual illustration of the time pattern of the effect is useful. We observe treatment and control areas without treatment both before and after the removal of local TV. If trends are the same absent of treatment we would expect that after the removal of the treatment the differences between treatment and control group revert to the pre treatment differences.[32] We thus can go go beyond the standard pre-trend check and provide a more powerful test that doesn't rely exclusively on the pre-period to extrapolate trends. To confirm parallel trends we would like to see that the treatment effect arises when local TV productions are introduced and disappears when they are removed.

Figure 11 the coefficient of a dynamic version of the difference in difference regression. It plots the difference in treatment and control areas before, during and after local TV. The number of top earners in treated areas rises when television production is introduced and the effect fades when local television productions start to disappear. By 1969 the differences between treatment and control group returned to the pre treatment level. Differences between treatment and control group thus are the same before and after treatment removal. This suggests that the common trend assumption holds.

For peace of mind I provide the conventional pre-trend checks as well. The results confirm that pre-trends are not driving the effects identified here. They are reported in the Appendix **??**. Appendix **??** presents further placebo checks, using leisure occupations that do not appear on television and triple difference analysis. All suggest that the findings are capturing the effect of television on entertainers.

5.2 Income Dispersion

The superstar model offers predictions for other parts of the distribution as well. We saw in section 2 that the superstar effect leads to income concentration at the top. The effect of a technical change is strongest on the top of the distribution where it leads to wage gains. This effect becomes smaller as we look move down in the wage distribution. For mediocre workers the effect changes sign. Such workers are only affected through the falling price for talent. They therefore see their income decline.[33] We therefore ought to observe a declining share of entertainers with such wages. The share of entertainers earning very low wages should increase.

around 1.

[32]This requires that the temporary increase in available market size has no lasting effect.

[33]Note that this is at odds with a conventional model of skill biased demand shifts. Positive demand shifts would lead to rising wages across the board.

The superstar model therefore offers a rich set of predictions that we can test in the data. A non-parametric way of implementing this test is to estimate the effect of television on different parts of the distribution in separate regressions. This amounts to running the baseline equation 12 multiple times, altering the dependent variable. We thus impose no restriction on the pattern of the treatment effect across the distribution. In this sense this is a completely unrestricted model and thus gives us the biggest chance of detecting deviations from the predictions of the superstar model. The dependent variable ($e_{m,t}^{p,p'}$) is defined analogous to above, where $p$and p' represent the upper and lower bound of the percentile bin under consideration, $e_{m,t}^{95,99}$ thus is the share of entertainers in m that earn wages between the 95th and 99th percentile of the US wage distribution.

The results for the various regressions are plotted in figure 12. TV leads to an increase in $e_{m,t}^{95,99}$ by about 12%. This is substantially smaller than the effect on the top 1%. The equivalent growth rate for the top 1% was 100%. The declining pattern is confirmed by looking at the next income bins. The effect of TV on the 90^{th}-95^{th} bin is insignificant, with a negative point estimate. For the 75^{th}-90^{th} bin the effect continuous to decline and becomes significantly negative. The effect remains similar for the following bin, while it starts rising again for the bottom part of the distribution. Below the median TV doesn't seem to have a significant impact. We observe a modest rise in the share of entertainers with wages at the very bottom of the distribution. Taking these effects together, the impact of TV follows a U-shape. The response of the wage distribution follows the prediction of the superstar model remarkably closely. The effect is biggest at the very top and declines continuously as we move down the distribution, turning negative below the 90th percentile. Moreover, we see a growth in entertainers with very low paid jobs. Particularly striking is that the non-parametric estimates match both the predicted sign, as well as the relative magnitude of the superstar model prediction. The resulting pattern thus provides strong evidence in favor of superstar effects.

5.3 Effect on Employment

A final distinctive prediction of a superstar effect is the change in employment. In a superstar setting we should see concentration of demand. When TV becomes available customers substituted away from local entertainment. As a result demand for local entertainers declines and employment should fall. In the model this is the result of a declining skill price which pushes some wages below the reservation wage. The employment effects are particularly interesting because the prediction contrast sharply with models of skill biased demand shocks. A positive demand shock would increase employment.

Places that are exposed to TV signal are not necessarily the same as labor markets that produce TV shows. Television signal travels beyond the labor market were shows are produced. I identify areas that are exposed to competition from TV shows. Here we should see the impact of declining skill prices. The treatment definition thus differ from the previous sections. It now measures whether TV can be watched in the local labor market m at time t. Figure 13 shows areas with TV signal in 1950 in blue.

The first set of results reports effects on employment in performance entertainment in local labor markets. The estimates are based on the difference in difference specification described in 3. The outcome variable is the inverse hyperbolic sine of employment. This transformation allows to interpret the effects as percentage changes, while handling a cell size of zero. To save on notation I will refer to it loosely as the log of employment. Employment data is available in the Census for additional years. A consistent definition is available since 1930. We can therefore extend the sample period by a decade. The introduction of TV signal leads to sizable employment loss. Table 2 reports the results. Employment in performance entertainment declines by around 13% when TV is available in the local labor market. As before columns 2 adds time varying demographics and column 3 controls for local labor market specific trends. The effects remain very close to the baseline. Differential local time trends are therefore unlikely to drive the results. The observed employment loss suggests that exposure to competition from TV has substantial effects on the local skill premium.

Next, consider checks for the common trend assumption. In contrast to local TV studios, TV signal is not removed in later years. We therefore have to use the pre-treatment period to assess the common trend assumption. With the extended sample it is feasible to look at traditional pre-trends. Column 4 runs the event study. Treatment and control group look similar in the run up to the treatment. The result confirms that there are no significant pre-trends, the coefficient on the lead of the treatment is insignificant and small in magnitude. Panel B repeats the same analysis for the time period where wage data is available. The result are very similar.

Next we turn to the variation introduced by the licensing freeze. Areas that would have received TV signal from "station that didn't happen" are marked in Figure 13. Panel C exploits this variation. The panel reports the employment effect of TV signal that did not happen because of the freeze. As expected such signal has no effect. There is therefore no evidence for spurious effects in areas that are about to be licensed. The above results appear to pick up the causal effect of TV.

5.4 Skill Biased Demand and Pay Setting Norms

Alternative models can potentially produce similar shifts in the wage distribution to the superstar model. As discussed above Skill Biased Demand shifts and changing pay setting norms have considerable flexibility and could potentially match the core predictions of the superstar model. As we saw in section 2 these models differ from superstar effects in other dimensions. To distinguish superstar effects from these models we therefore turn to these predictions.

We already saw that television led to sizable employment losses, which is at odds with a positive demand shift. Next I will test if the pattern of wage growth across the distribution by using quantile regressions. A shortcoming of the quantile regression is that the estimates are sensitive to 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 SM 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. The quantile regressions mimic the difference in difference logic by controlling for group and time fixed effect following the procedure in Chetverikov et al. [2016]. The estimated coefficients are plotted in figure A, alongside the prediction of the SBD model. Television has the biggest effect at the top of the distribution and notably smaller effects at the lower percentiles. This result is in line with the superstar model but contradicts a model of SBD. We can test this formally by estimating whether the percentile ratios are unaffected by the introduction of television. This will test whether the visible differences in figure A are statistically meaningful. I construct percentile ratios and regress them on the introduction of television. Television has a significant effect on all percentile ratios. The SBD hypothesis that these ratios are unaffected is rejected at the 95% confidence level for any ratio (see Table 3).

A further prediction that distinguishes superstar and demand effects is the impact on dispersion of income within the top tail. We saw above that the SBD model predicts that top income shares at all percentiles grow at the same rate. To test this prediction I compute the share of income captured respectively by the top 10%, top 1% and top 0.1% of entertainers in a given local labor market -year. If the top tail of the distribution is not observed I use Pareto interpolation to estimate top income shares. This follows a large literature that uses Pareto interpolation on limited data, e.g. from tax tables [Kuznets and Jenks, 1953, Feenberg and Poterba, 1993, Piketty and Saez, 2003, Piketty and Zucman, 2014]. The procedure is described in Appendix **??**.

We can now analyze how top income shares in entertainment change during the roll-out of TV. These regressions run the baseline difference in difference at the local labor market - year level, using local top income shares ($S_{p\%}$) as outcome variable. The results are reported in table 4. The income share of the top 1% of entertainers growths by 3.7 percentage points. That is a large increase on the pre-TV share of 3.8%, the income share of the top 1% almost doubles. Looking at even higher fractals of the income distribution reveals that most of the gains accrue to the very top. The top 0.1% of entertainers see their income share rises by 2.4%. More than halve of the rise of the income share of the top 1% therefore come from the top 0.1%. The income share of the top 10% of entertainers also growths by 6 percentage points.

We can now test formally whether the observed change in the income distribution could arise from a SBD shock. Recall that in a SBD model all top income shares grow at the same rate. The data is clearly at odds with this prediction. While the share of income going to the top 1% growths by 100%, the equivalent share for the top 0.1% growths by 300%, while the top 10% share only growths by only 30%. To estimate standard errors, I run a regression of the ratio of top income shares on the introduction of local TV. The hypothesis that the top 1% growths at the same rate as either the top 10% or the 0.1% is rejected. With over 99% confidence we can discard the null that television doesn't change the relative dispersion of income in the top tail. The growth rate of income is highly skewed towards the top tail, giving rise to growing dispersion within the top tail. This is at odds with models of skill biased demand where wages grow proportionally across all percentiles, but the finding is in line with the superstar model where wage growth is skewed towards the top of the distribution.

Finally, I study whether changes in pay setting norms can explain the results. In a model of norm driven pay setting wages are divorced from productivity and shifts in bargaining power drive income concentration at the top. The identification strategy above controls for time fixed effects and thus rules out that trends in norms are driving the findings. Technical change may still affect bargaining directly in the affected local labor markets. For example by breaking established pay-setting institutions or introducing a new "corporate culture". This would result in growing top pay, while leaving the underlying productivity of workers unchanged. To test whether wage changes reflect changes in productivity I collect data on consumption expenditure on entertainment. Consumption patterns will trace the marginal revenue product of different groups of workers. This will establish whether the effects are driven by a shift in demand and an associated decline in productivity. To the best of my knowledge there is no dataset that traces sufficiently disaggregated data to look at the demand within the entertainment category at the local level. I digitize detailed local spending data from records of county fairs. Data on expenditure on county fairs is of course only a fraction of overall entertainment spending. However, this data allows me to build a measure of demand for entertainment at the local labor market level at annual freuqency from 1946 to 1957. During this period there exists substantial regional variation in the exposure to TV. The data allows to analyse inter Census years and covers a period where we have information on the number of channels available locally. We can thus extend the sample and use the richer variation in TV exposure to measure the effects of TV.

The impact of TV on spending on local county fairs is reported in Table 5. The introduction of television signal reduces ticket revenues of county fairs by about 6% (column 1). The effects are imprecisely estimated as receipts are only reported for a subset of all fairs. We have more data on fair attendance. Most of the revenue drop is accounted for by a decline in the number of fair visits (column 2). The number of visits declines by about 9%. The demand shift away from local entertainment decreases the marginal revenue product of local county fair entertainers. In line with a superstar model, wage changes are a result of changes to productivity.

It is unclear whether the total demand for local entertainment declines similarly to the demand for county fairs. The employment loss for local entertainers is however remarkably close to the estimated decline in demand for local fair entertainment. One interpretation of this finding is that employment is reduced to cut costs in light of the revenue shortfall. This would suggest that local entertainer wages are relatively rigid and wages at the bottom of the distribution absorb little of the shock. In line with the idea of local wage rigidities, county fair tickets prices also seem to respond little to the demand shock. Most of the shock is thus reflected in job losses.

To further add credibility to the demand estimates, I distinguish between the demand for different types of entertainment at county fairs. The effect of television should be strongest for entertainment that is a close substitute for television. Indeed in our employment regressions above we found that job losses were concentrated on entertainers who are very substitutable by TV. To estimate the analogous demand swing I collect data on receipts at two extremes of substitutability. On the one hand revenues from grand stand shows. These are performance entertainment similar to much of the entertainment shown on TV at the time. It included vaudeville acts, thrill shows, dance groups and beauty pagans. On the other hand I collect data on receipts from traditional carnival activity. These include sale of candy and fairy rides. Such activities are less substitutable by television. Columns 3 and 4 report the results for both. Show receipts indeed see the largest decline, while carnival receipts are rising slightly. Power is limited but the evidence is very much in line with a superstar story. People are responding to television by reducing their demand for local performance entertainment, while demand for other types of local entertainment holds up reasonably. Overall the shift in entertainment demand is in line with the observed changes in the labor market. The measured wage changes thus appear to reflect shifts in productivity of workers. My results therefore lend support to models of competitive wage setting in this industry. At the same time, I find evidence for wage rigidities at the bottom of the pay distirbution.

5.5 Channels

The above analyses the total effect of production scale on the income distribution. To understand the mechanism that drive this effect it will be useful to analyze different channels.

A first possible driver for the rise in top incomes could be migration. Higher returns to talent might attract more high skilled individuals to the labor market. Such migration could potentially explain part of the rise in top incomes in the affected areas. On the other hand, TV stations were set up in markets with the largest theaters and performance venues. While TV changes the size of production scale it doesn't change the ordering of markets. TV stations may therefore not change where top workers would want to locate. Moreover, TV only provided a temporary change to local production scalability. If entertainers were anticipating that the change is temporary the incentive to relocate is further dampened. We can test empirically which of these factors dominate.

The Census records information on migration which can be used to test the effect of television on migration patterns. Unfortunately the migration question does not distinguish between moves within and across labor markets, moreover the question changes between years. The census asks whether a person has moved in the last X years, but X differs between Census years. Year fixed effects may help to absorb some of this variation across Census years. Migration rates are fairly high. The probability of having moved house within the last 5 years is around 60%. Many of these moves are likely within the same labor market. The effect of television on the probability of moving is however a precise zero. Column 1 - 4 of table 6 show that changes in moving rates are below 1%. Despite the substantial noise in the outcome variable the confidence intervals rule out large migration responses. The small role for migration suggests that the inequality effects are mainly driven by changes in returns to skill.

A related question is whether migration results in spill-overs to the control group. This would occur if there was migration from the control areas to the treated areas. Given the modest migration response such effects are likely small. To analyses the impact quantitatively, we can use heterogeneity in the cost of relocating. Moves between neighboring markets is arguably easiest and we should see most of the spill-overs take place between neighboring markets. By excluding control areas that neighbor treated labor markets from the analysis we will get a sense for the importance of spill-over effects are. The results of such a regression are very similar to the baseline, see column 5 of table 6. Relocation between neighboring markets does not play a major role.

6 Magnitude of Superstar Effects

The test for superstar effects confirms that labor markets respond in line with the model predictions. We don't however know yet whether superstar effects are large or small. In particular, we would like to know how much top income growth superstar effects can explain. To get a sense of magnitude we need to express the results in terms of elasticities. An elasticity with respect to market size gives the top income response to a given shock. To estimate such an elasticity we need to know how much TV shifted market size. I therefore run the baseline specification with market size as outcome variable. The market size variable measures potential audience of a single performance in a given labor market. The data on audience comes from the newly collected sources described above.

Table 7 reports the results of a regression of the log of audience size on television. The arrival of television leads to 140 log points growth in audience size. Converting the log points this implies a market size growth by factor 4 to 5. Through television entertainers could therefore reach a substantially larger market.

We can compute the ratio of the estimates of TV on top earners and on market size to get at the elasticity. This is an IV estimate for the elasticity of top earners to audience size, using the timing of TV arrival as instrument. The implied elasticity is around 0.3. A 10% growth in the available market size thus results in a 3% increase in the number of top earners.

6.1 Elasticity of Top Incomes

Let's next compare the estimates with the previous literature. A pervasive finding is that the elasticity of pay to firm value for these top earners is around 1/3 in cross-sectional data of CEO pay (see Murphy [1999] for a summary). Gabaix and Landier [2008] replicate this finding and argue that in the time series the elasticity should be larger, as growth in average market size amplifies top pay. In line with this result they find a time series elasticity closer to 1. Frydman and Saks [2010] expand the time period back to 1936 and dispute that the time series elasticity is as large. They find that over the longer period the elasticity is closer to 0.1. The range of suggested estimates is therefore large, suggesting potentially vast differences in the importance of superstar effects for top income growth.

All these studies are based on correlations between CEO pay and market size. It seems plausible that a link between the two variables comes about because a good CEO is worth more to a large firm. However, market value and pay could also be linked for other reasons. My estimates allow to identify the causal role of market size. Comparing my estimates to the previous literature can thus help us understand potential biases. A caveat is that superstar effects may operate differently in entertainer and CEO labor markets. The next section will explore in detail how superstar effects vary across different institutional settings. It finds that such differences are likely modest. A comparison of two settings thus is likely informative.

My estimates in the previous section are not directly comparable to the previous literature. Recall that analyzing top earner employment instead of wages allowed us to bypass the top-coding problem. The CEO literature looks at a wage elasticity, while the estimate discussed so far instead analyze the elasticity of the number of top earners. These two estimates are linked. The link is very simple if the wage distribution is Pareto and the superstar effect is order preserving. This is a useful benchmark and we can relax those assumption somewhat below.

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

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

Where we define G(x) as the share of individuals above x. Consider a small increase in the number of top earners. If the order of individuals in the distribution has remained the same we can re-write the expression.[34] The top earners in period 1 are the top earner of period 0 plus individuals that were previously just below the top earner threshold. Let's denote the lowest period 0 wage of a period 1 top earner by $\tilde{w}$. The number of new top earners thus becomes:

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

It follows:

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

where the last equality holds for small changes in w.

If the shape of the CDF is known this equation allows to translate a change in employ-

[34] Rosens' model of superstar effect is an example where this assumption holds. This is however a strong assumption that can be relaxed with additional data on the position of TV-stars in the pre-TV wage distribution. This is to follow.

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

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

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

This gives us a simple expression for the link between the elasticity of number of top earners and the elasticity of pay of top earners to market size. I estimate the α parameter on the pre-TV distribution using the full count, non top-coded Census of 1940. I experiment with a number of estimation strategies with similar results.[35] Independent of the approach the estimated Pareto coefficient is close to but bigger than 3. To err on the conservative side, I will use a value of 3 for the analysis.

In the analysis above we found $\varepsilon_{TE,m} = 0.3$. Using the relation derived here we can translate this into an elasticity of income. The implied elasticity of top wages to market size is $\varepsilon_{w,m} \approx 0.1$. A doubling in market size will thus raise top wages around 10%.

The wage elasticity would be bigger if we relaxed the assumption that the effects are order preserving. Without such homogeneous treatment effects individuals from further down in the wage distribution could become earning superstars. This would require a large wage rise for these people and thus potentially increase the estimated elasticity. To assess how much this matters in practice, we need to know where in the income distribution local TV stars came from. Figure 15 plots this. The figure matches local TV stars to their pre-TV earnings in the 1939 Census.[36] It plots the position of these stars in the pre-TV income distribution after correcting the wage for age, gender and education effects. The figure makes clear that most of the star entertainers were earning high incomes in the pre TV period.The order preserving assumption thus looks like a reasonable approximation. Allowing for heterogeneous treatment is unlikely to change the conclusions substantially.

As discussed above most of the previous literature looks at the correlation of executive pay and firm size. There the cross-sectional relation is 0.33 ($\hat{\beta}^{OLS} = \frac{1}{3}$). The elasticity estimate here is thus only a third of these estimates. My estimates are closest to time series estimates for the middle of the 20th century (Frydman and Saks [2010]).

The difference may be explained by differences in the two settings. It seems however also plausible that the OLS estimates from the CEO literature have overstate the importance of

[35]The baseline results use Kuznets' approach to estimate the Pareto parameter. This approach uses the fact that average income above a threshold is proportional to the threshold. With α the coefficient of proportionality. I also run the Atkinson & Piketty approach and use different threshold values. All estimates are above 3, with most between 3.02 and 3.16.

[36]The information on TV stars comes from the 1949 "TV and Radio Annual." The magazine publishes a who is who of the industry. Biographical information is used to link those individuals to their 1940 Census records. The information is thus based on the subset of TV stars who are listed in the who is who and can be found in the 1940 Census.

superstar effects relative to the causal effect. OLS regressions of CEO pay and firm size don't control for the potential endogenous change in market size. It seems likely that omitted factors, such as effort or endogenous market size change, are positively correlated with market size growth and wages. This would lead to an upward bias of the OLS coefficient. The widespread rise in performance pay throughout the 80s likely worsened the bias Hall and Liebman [1998]. This plausibly explains why estimates from the earlier period are closer to the causal effect of market size.

If the estimates found in this study are closer to the causal impact of superstar effects, superstar effects explain less of the rapid growth in top incomes than many economist speculated.

6.2 Market Structure and Superstar Effects

An important question is how the findings here translate to other settings. To get a sense of this I will analyze how superstar effects vary with the institutional setting. An example of institutional factors that will affect superstar effects are taxes. Scheuer and Werning [2017] show that labor supply elasticities are larger in markets with superstar effects. Settings with large tax rates therefore ought to lead to smaller superstar effects. Similarly, imperfect competition in the labor market would likely affect the magnitude of superstar effects. The above benchmark model is perfectly competitive and no rents occur because all distributions are smooth. If employers have market power, they will capture some of the surplus generated by the greater scalability of production.

The empirical setting offers rich heterogeneity that allows us to analyze how much superstar effects across different institutional settings. First consider the case of imperfect competition. If there is a single employer that has access to the new production technology we would expect that the returns of the technology are captured by the employer. Wages should thus not respond in that case. In contrast, if there is competition for top workers wages will be bid up and workers will capture the benefits of the technology.

The licensing process limits employer entry into labor markets. This generates differences in the competitive structure of labor markets. There are a number of labor markets with a single TV station, while in other markets there are several TV stations. To test how much monopsony power affects the results I allow differential effects in markets with a single TV station. Column 4 of Table 8 allows for a differential effect for markets with monopsony TV employers. Splitting the treatment group does increase the standard errors. Nevertheless, the differences between monopsonistic and competitive labor markets is significant. Top earners gain little in markets with a monopsony employer, while the gains are large when there is more than one employer.

Alternatively, the pre-TV labor market may have been imperfectly competitive. The entry of a TV station may break up anti competitive structures. To investigate this possibility I allow the effect of TV to differ across labor markets with different amount of pre-TV employers. The result is reported in column 2 of Table 8. The effect 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. An alternative measure is the population density. Fixed costs might lead to natural monopolies in less densely populated areas. Population density may thus be an alternative proxy for competition in the labor market prior to TV. In practice there is again no effect. This confirms that imperfect competition in the pre TV labor market doesn't matter greatly.

Taxes are hypothesized to reduce the magnitude of superstar effects. The idea is that part of the superstar effect is that star workers increase their work effort in response to higher returns to their skill. Moreover, in a superstar setting taxes can distort the optimal assignment of workers to markets. Taxes thus introduce an added distortion. It is however unclear how important these effects are in practice. It may for instance be the case that for earning superstars the income effect dominates the substitution effect. In that case workers would not increase their labor supply in response to the change in returns to skill. The effect of taxes on superstar effects is thus an empirical question.

The Empirical literature on taxes and superstars has mainly focused on migration. Stars move significantly more across state or country boundaries in response to tax incentives (Kleven et al. [2014], Moretti and Wilson [2017], Kleven et al. [2013]). The behavioral response of superstars may however be substantially different. The share of movers is very small and the associated distortion from migration might be dwarfed by labor supply changes of the stayers. Moreover, tax rates may have different effects in superstar markets to other markets. Scheuer and Werning [2017] documents that the distortion effect of taxes gets amplified by superstar effects. The empirical literature hasn't used any variation in superstar effects to test this implication. We therefore don't have any evidence how superstar effects and tax rates interact. Along the same lines we don't know how much taxes reduce the growing superstar effects.

Top marginal tax rates differ across states as states levy additional income taxes. This variation in tax rates allows me to test whether superstars adjust their labor supply decision in response to differential tax rates. Data on historic tax rates of individual state level is not centrally collected. I take the data from information collected by the historic study of Penniman & Heller [Penniman and Heller, 1959]. They collect detailed information on income tax legislation across US states during the sample period.[37] Using this information I construct a dummy variable that takes the value one for high tax states. These are states with tax rates above the median.[38] I ran a regression that allows for differential superstar

[37]As far as possible I use information on tax rates in 1945. This predates most of the roll-out of TV and avoids potential endogeneity concerns. Most of the data is collected in 1957 but tax reforms are noted. If no reform is reported I use the 1957 tax rate. I exclude Delaware were a substantial reforms took place between 1945 and 1957.

[38]I use a binary variable because marginal tax rates are difficult to interpret in this context. Deductibility

effects in high tax states. This estimates combine the effect of out-migration and reduced labor supply by stayers. Column 4 shows that there is no significant difference between high and low tax states. While the standard errors are again 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 this superstar market.

7 Conclusion

Vast changes in top incomes occurred over the last decades but little is known about the causes for this. Technological innovation, foremost in communication technologies have made it easier to operate over distance. Companies have gone from small local enterprises to global operations. These changes have been associated with growth in top incomes. This paper is the first to provide causal evidence on the effect of growing scalability of production on wages.

The paper analyses superstar effects in the entertainment industry. The staggered introduction of television has led to a substantial growth in the market reach of entertainers. A single performance has been able to increase his audience by factor 4 with the help of the new technology. This generates the kind of variation that is alleged to lead to superstar effects. Regulatory and technological problems led to quasi-random variation in the roll-out of the technology and allows me to identify the effect. I find that a doubling in the market size leads to doubling in the number of top earners in entertainment. Moreover, wage dispersion among top earners rises. This is consistent with the predictions of the superstar model.

I also find evidence that a superstar technology creates losers. This model prediction contrasts sharply with other forms of skill biased technical change. The ability to offer one's product simultaneously to a big audience reduces the return for mediocre workers and pushes some out of the market. I find evidence for both of these effects. Competition from the new technology leads to a reduction in employment of ca. 10%. The increase in production scalability has therefore profound effects on inequality both at the top and bottom of the distribution.

To assess the magnitude of superstar effects, the paper provides elasticities of top incomes with respect to market size. The estimates imply that the share of income going to the top 1% increases by 30% if market size doubles. The greater concentration of income is driven by income growth at the top and income losses for mediocre workers. Top wages grow around 10% in response to a doubling in market size. Compared to estimates from the CEO literature this elasticity is at the lower side. The elasticity of CEO pay to market size is close to 1 in recent decades. A plausible reason for the difference is that executive pay is linked to firm size through performance packages. The correlation of pay and firm vlaue may therefore be the result of incentive contracts rather than superstar effects. Performance pay has become more common in recent decades, and the CEO literature does indeed find pay

rules generate a wedge between MTR and headline rates. This is less of a problem for a comparison of high and low tax states to the extend that deductibility rules don't change whether a state is a low or high tax state.

elasticities in line with my results for earlier periods. If the estimates found here are indeed closer to the causal impact of superstar effects, such effects are smaller than previously thought. Nevertheless, the superstar effect could still acounts for at least halve of the US growth in top incomes.

References

Daron Acemoglu and David H. Autor. Skills, tasks and technologies: Implications for employment and earnings. *Handbook of Labor Economics*, 4(PART B):1043–1171, 2011. ISSN 15734463. doi: 10.1016/S0169-7218(11)02410-5. URL `http://dx.doi.org/10.1016/S0169-7218(11)02410-5`.

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

Anthony B Atkinson and Thomas Piketty. The Distribution of Top Income in the UK 1908 - 2000. In *Top Incomes over the 20th Century*. 2006.

David Autor, Lawrence F. Katz, David Dorn, Christina Patterson, and John Van Reenen. The Fall of the Labor Share and the Rise of Superstar Firms. *NBER Working Paper Series*, (23396), 2017. doi: 10.3386/w23396.

David H. Autor and David Dorn. The Growth of Low Skill Service Jobs and the Polarization of the U . S . Labor Market. *American Economic Review*, 103(5):1553–97, 2013. ISSN 0002-8282. doi: 10.1257/aer.103.5.1553.

David H. Autor and Gordon H. Hanson. Untangling Trade and Technology: Evidence from Local Labor Markets. 125(7329):1–24, 2013. ISSN 00130133. doi: 10.1111/ecoj.12245.

R. L. Axtell. Zipf distribution of U.S. firm sizes. *Science*, 293(5536):1818–1820, 2001. ISSN 00368075. doi: 10.1126/science.1062081.

Olivier Blanchard and Lawrence F. Katz. Regional evolutions. *Brookings Papers on Economic Activity*, pages 1–75, 1992. ISSN 00072303. doi: 10.2307/2534556. URL `http://search.ebscohost.com/login.aspx?direct=true{\&}db=buh{\&}AN=9301100352{\&}site=ehost-livehttp://www.jstor.org/stable/2534556`.

Denis Chetverikov, Bradley Larsen, and Christopher Palmer. IV Quantile Regression for Group-Level Treatments, With an Application to the Distributional Effects of Trade. *Econometrica*, 84(2):809–833, 2016. ISSN 0012-9682. doi: 10.3982/ECTA12121. URL `https://www.econometricsociety.org/doi/10.3982/ECTA12121`.

Alberto Chong and Eliana La Ferrara. Television and Divorce: Evidence From Brazilian Novelas. *Journal of the European Economic Association*, 7(2-3):458–468, 2009. ISSN 1542-4766. doi: 10.1162/jeea.2009.7.2-3.458. URL `http://dx.doi.org/10.1162/JEEA.2009.7.2-3.458`.

John DiNardo and Joern-Steffen Pischke. The returns to computer use revisited: have pencils changed the wage structure too? *Quarterly Journal of Economics*, 112(1):291–303, 1997. ISSN 00335533. doi: 10.1162/003355397555190. URL `http://eprints.lse.ac.uk/6152/`.

Ruben Durante, Paolo Pinotti, and Andrea Tesei. The Political Legacy of Entertainment TV. *Working Paper*, (April), 2015.

Alex Edmans and Xavier Gabaix. Executive compensation. *Journal of Economic Literature*, 54: 150, 2016. URL `http://search.ebscohost.com/login.aspx?direct=true{\&}db=bth{\&}AN=10230632{\&}site=ehost-live`.

Ruben Enikolopov, Maria Petrova, and Ekaterina Zhuravskaya. Media and political persuasion: Evidence from Russia. *American Economic Review*, 101(7):3253–3285, 2011. ISSN 00028282. doi: 10.1257/aer.101.7.3253.

Daniel R. Feenberg and James M. Poterba. *Income Inequality and the Incomes of Very High-Income Taxpayers: Evidence from Tax Returns*, volume 7. 1993. ISBN 0262161354. doi: 10.1086/tpe.7.20060632. URL `http://www.journals.uchicago.edu/doi/10.1086/tpe.7.20060632`.

Carola Frydman and Raven E. Saks. Executive compensation: A new view from a long-term perspective, 1936-2005. *Review of Financial Studies*, 23(5):2099–2138, 2010. ISSN 08939454. doi: 10.1093/rfs/hhp120.

Yoshi Fujiwara, Corrado Di Guilmi, Hideaki Aoyama, Mauro Gallegati, and Wataru Souma. Do Pareto-Zipf and Gibrat laws hold true? An analysis with European firms. *Physica A: Statistical Mechanics and its Applications*, 335(1-2):197–216, 2004. ISSN 03784371. doi: 10.1016/j.physa.2003.12.015.

Xavier Gabaix and Augustin Landier. Why Has CEO Pay Increased So Much? *Quarterly Journal of Economics*, 123(1):49–100, 2008. ISSN 0033-5533. doi: 10.1162/qjec.2008.123.1.49.

Xavier Gabaix, Augustin Landier, and Julien Sauvagnat. CEO pay and firm size: An update after the crisis. *Economic Journal*, 124(574):40–59, 2014. ISSN 00130133. doi: 10.1111/ecoj.12084.

Luis Garicano. Hierarchies and the organization of knowledge in production. *Journal of political Economy*, 108(5):874–904, 2000. ISSN 00223808. doi: 10.1086/317671. URL `http://www.journals.uchicago.edu/toc/jpe/current`.

Luis Garicano and Thomas N. Hubbard. Specialization, firms, and markets: The division of labor within and between law firms. *Journal of Law, Economics, and Organization*, 25(2):339–371, 2009. ISSN 87566222. doi: 10.1093/jleo/ewn003.

Luis Garicano and Esteban Rossi-Hansberg. Inequality and the organization of knowledge. *American Economic Review P&P*, 94(2):197–202, 2004. ISSN 00028282. doi: 10.1257/0002828041302037.

François Geerolf. A Theory of Pareto Distributions. *Working Paper*, (May):1–49, 2016.

Matthew Gentzkow. Television and voter turnout. *The Quarterly Journal of Economics*, 121 (August):931–971, 2006. ISSN 0033-5533. doi: 10.1162/qjec.121.3.931. URL `http://qje.oxfordjournals.org/lookup/doi/10.1162/qjec.121.3.931`.

Matthew Gentzkow and Jesse Shapiro. Introduction of Television to the United States Media Market, 1946-1960, 2008.

Brian Hall and Jeffrey Liebman. Are CEOs Really paid like bureaucrats. *Quarterly journal of economics*, XCIV(May):437–452, 1998.

Steven N. Kaplan and Joshua Rauh. It's the Market: The Broad-Based Rise in the Return to Top Talent. *Journal of Economic Perspectives*, 27(3):35–56, 2013. ISSN 0895-3309. doi: 10.1257/jep.27.3.35. URL `http://pubs.aeaweb.org/doi/abs/10.1257/jep.27.3.35`.

Lawrence F. Katz and David H. Autor. Changes in the Wage Structure and Earning Inequality. *Handbook of Labor Economics*, 3, 1999.

Lawrence F. Katz and Kevin M. Murphy. Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *The Quarterly Journal of Economics*, 107(1):35–78, 1992. ISSN 0033-5533. doi: 10.2307/2118323. URL `http://www.jstor.org/stable/2118323{\%}5Cnhttp://www.jstor.org/stable/pdfplus/2118323.pdf?acceptTC=true`.

Henrik Jacobsen Kleven, Camille Landais, Emmanuel Saez, and Esben Anton Schultz. Taxation and International Migration of Superstars: Evidence from European Football Market. *American Economic Review*, 103(5):1892–1924, 2013. ISSN 0002-8282. doi: 10.1257/aer.103.5.1892. URL http://www.nber.org/papers/w18885.

Henrik Jacobsen Kleven, Camille Landais, Emmanuel Saez, and Esben Anton Schultz. Taxation and International Migration of Top Earners: Evidence from the Foreigner Tax Scheme in Denmark. *The Quarterly Journal of Economics*, 129(1):333–378, 2014. URL http://www.nber.org/papers/w18885.

Simon Kuznets and Elizabeth Jenks. *Shares of Upper Income Groups in Savings*, volume I. 1953. ISBN 087014054X. doi: 10.2307/2343040. URL http://www.nber.org/books/kuzn53-1.

Steven Manson, Jonathan Schroeder, David Van Riper, and Steven Ruggles. IPUMS National Historical Geographic Information System: Version 12.0. Technical report, University of Minnesota, Minneapolis, 2017.

Alfred Marshall. Principles of Economics. In *Book VI*, page § 11. London: Macmillan and Co., 1890.

Enrico Moretti and Daniel Wilson. The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists. *American Economic Review*, 107(7):1858–1903, 2017. ISSN 0002-8282. doi: 10.3386/w21120.

Kevin J Murphy. Executive compensation. In *Handbook of Labor Economics, Vol. 3*, number June, pages 2485–2563. 1999. ISBN 9780444501882. doi: 10.1016/S1573-4463(99)30024-9.

Clara Penniman and Walter Heller. *State income tax administration*. Public Administration Service, University of Michican, 1959.

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

Thomas Piketty and Emmanuel Saez. Income Inequality in the United States, 1913-1998. *The Quarterly Journal of Economics*, 68(February):1–39, 2003. ISSN 0033-5533. doi: 10.1162/00335530360535135. URL http://piketty.pse.ens.fr/fichiers/AngristKrueger1991.pdf.

Thomas Piketty and Gabriel Zucman. Capital is Back: Wealth-Income Rations in Rich Countries 1700-2010. *Quarterly Journal of Economics*, 129(3):1255–1310, 2014. ISSN 0033-5533. doi: 10.1093/qje/qju018.Advance.

Sherwin Rosen. The Economics of Superstars. *American Economic Review*, 71(5):845–858, 1981. ISSN 00028282. doi: 10.2307/1803469. URL http://search.ebscohost.com/login.aspx?direct=true{\&}db=buh{\&}AN=4507864.

Steven Ruggles, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. Integrated Public Use Microdata Series: Version 7.0. Technical report, University of Minnesota, Minneapolis, 2017.

Michael Sattinger. Assignment Models of the Distribution of Earnings. *Journal of Economic Literature*, 31(2):831–880, 1993. ISSN 00220515.

Florian Scheuer and Iván Werning. The Taxation of Superstars. *Quarterly Journal of Economics*, 132(1):211–270, 2017. doi: 10.3386/w21323.

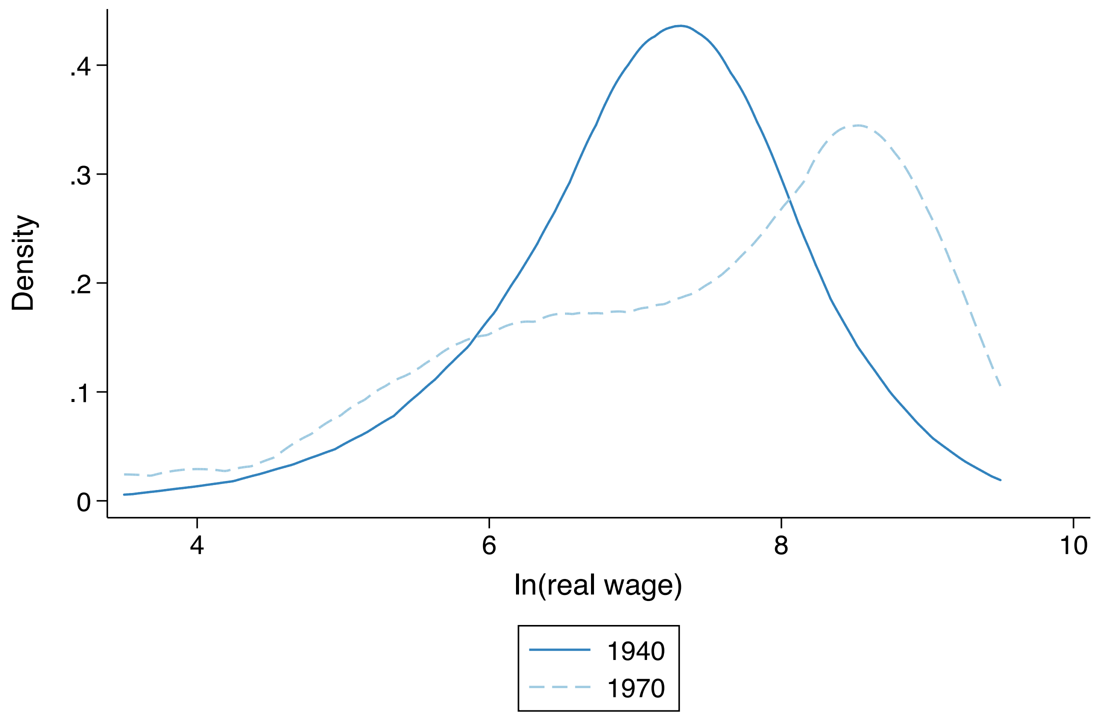
Jonathan Sterne. Television under construction: American television and the problem of distribution, 1926-62. *Media, Culture & Society*, 21(199905):503–530, 1999. ISSN 0163-4437. doi: 10.1177/016344399021004004.

Marko Terviö. The difference that CEOs make: An assignment model approach. *American Economic Review*, 98(3):642–668, 2008. ISSN 00028282. doi: 10.1257/aer.98.3.642.

Jan Tinbergen. On the Theory of Income Distribution. *Weltwirtschaftliches Archiv*, 77(2):155–175, 1956. ISSN 00432636. doi: 10.1111/j.1536-7150.1957.tb00186.x. URL `http://www.jstor.org/stable/40435398`.

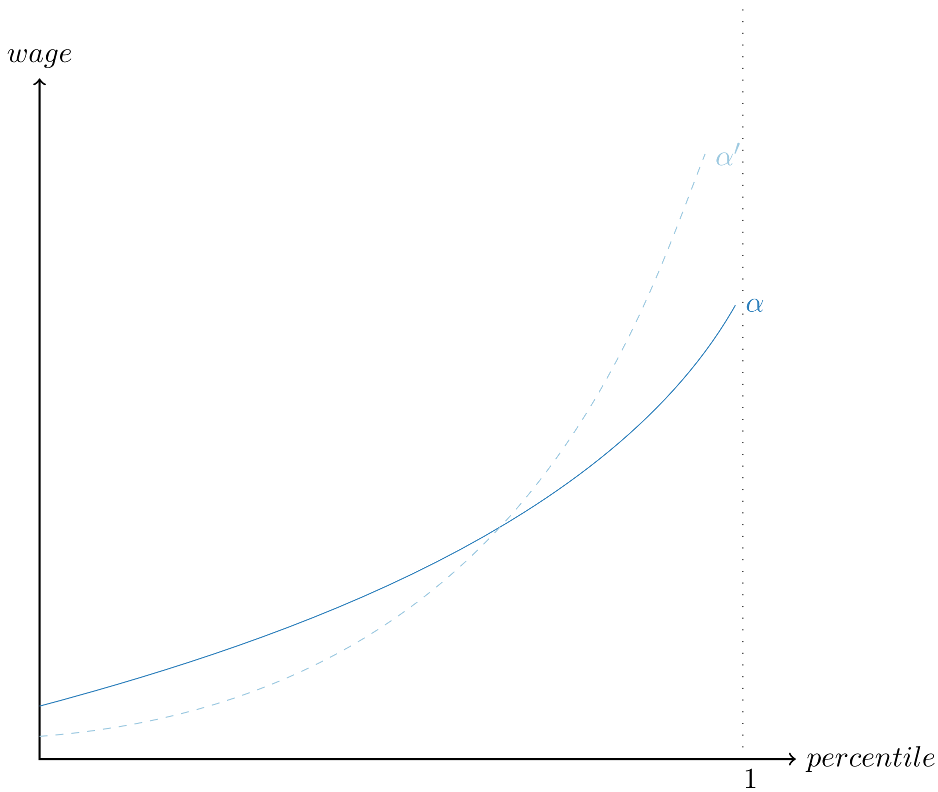
A Figures

Figure 1: Wage Distribution 1940 and 1970



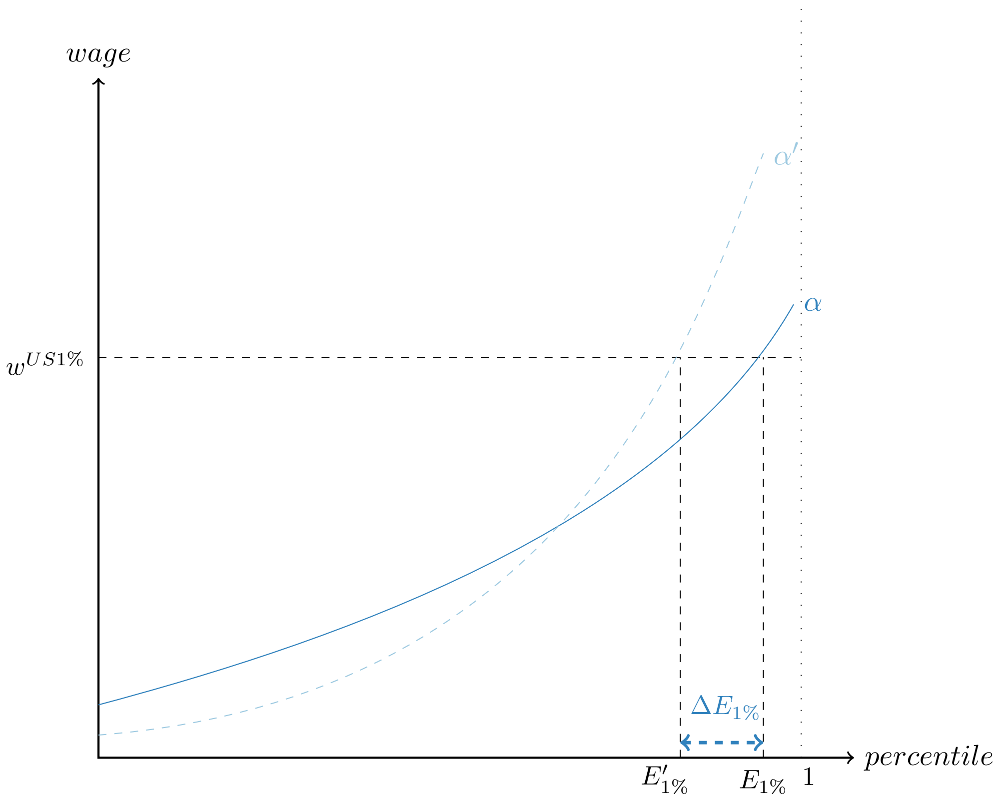
[Note] Log real wage distribution for performance entertainers from the lower 48 states of the US from the 1940 and 1970 Census. Real wages in 1950 USD using Census sample weights. Density is estimated using the Epanechnikov smoothing kernel with a bandwidth of 0.4. Common top code applied at $85,000.

Figure 2: Superstar Effect on Wages



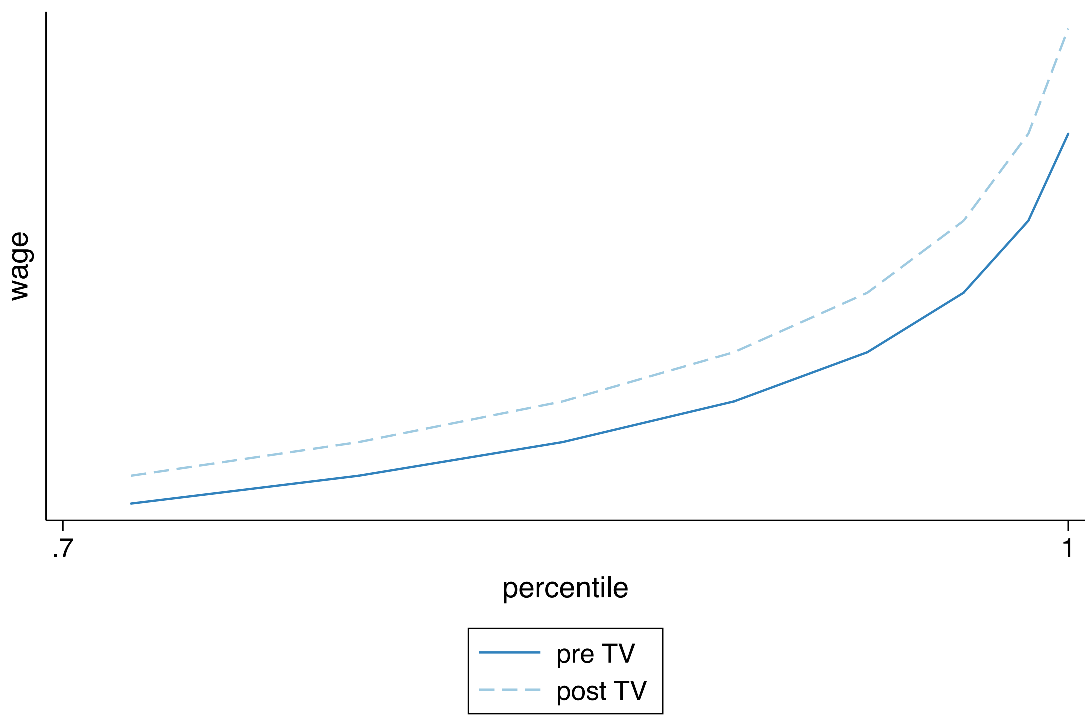
Note: Wages based on a superstar model ($w_p = \pi \cdot \kappa \cdot (1-p)^{-(\alpha\gamma-\beta)}$). α is the shape parameter of the market size distribution ($\alpha' > \alpha$). The percentiles shown are the upper tail of the wage distribution. With exit they correspond to the percentiles in the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

Figure 3: Superstar Effect on Top Earner



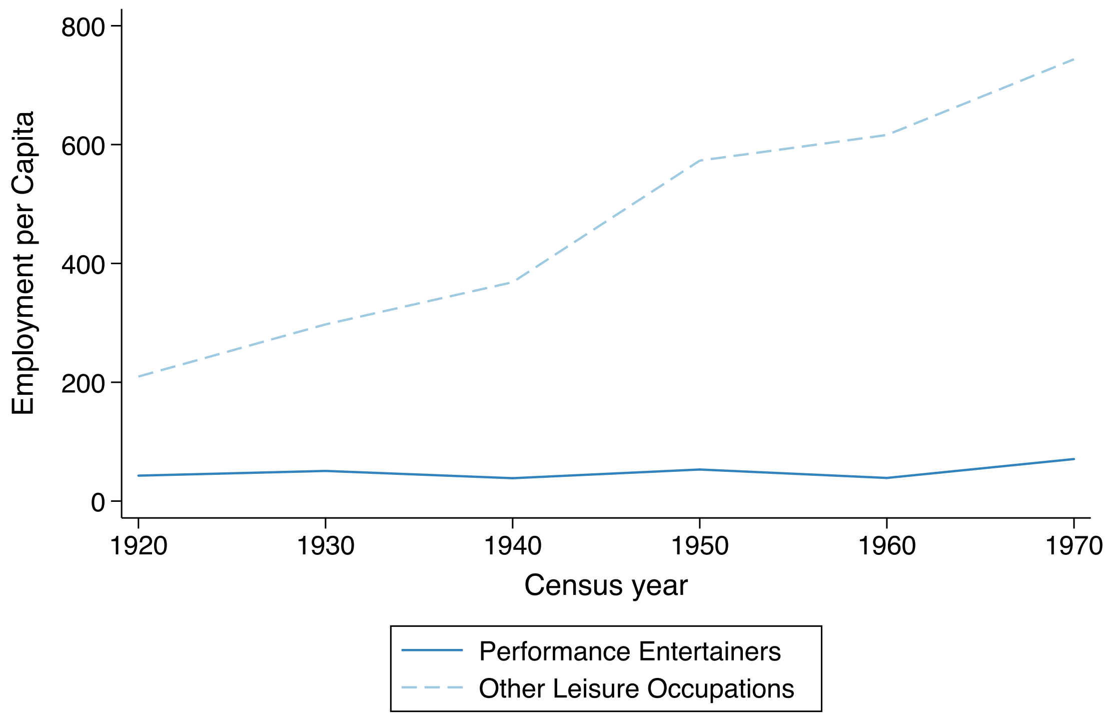
[Note] Details as in figure 2. $w^{US1\%}$ is a wage threshold that defines a top earner, e.g. the national top percentile. $E_{1\%}$ and $E'_{1\%}$ are the share of entertainers above the threshold. $\Delta E_{1\%}$ is the change in top earners when market size becomes more dispersed (move from α to α').

Figure 4: Skill Biased Demand Wage Distribution

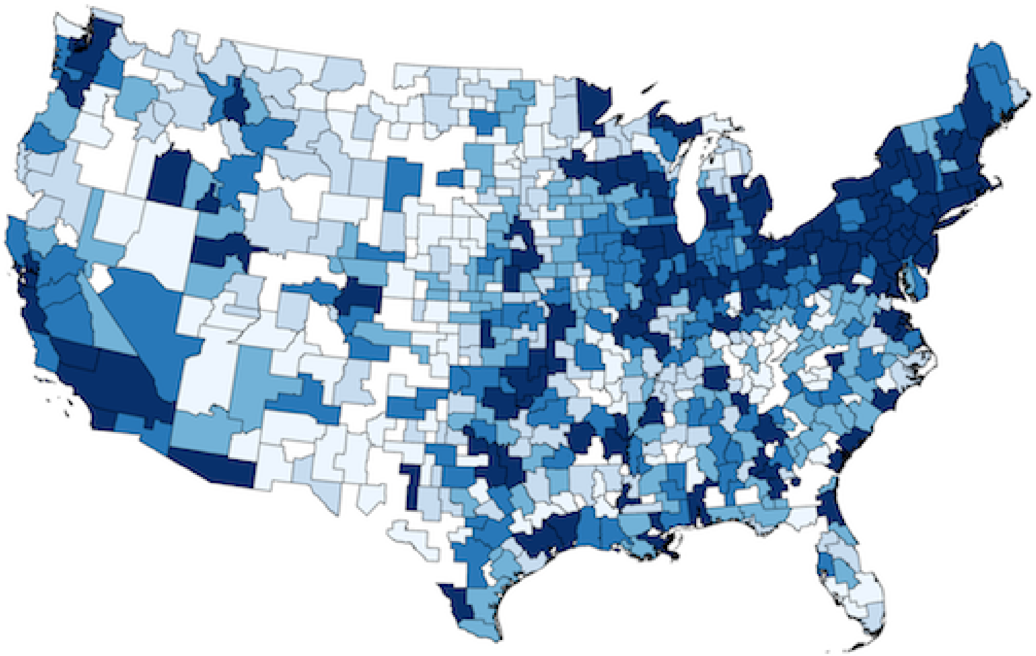


[Note] The figure shows the wage distribution above the 70th percentile. The talent distribution has been chosen to match the 1940 wage distribution. The change in the skill premium matches the growth in top earners.

Figure 5: Entertainment Employment per Capita

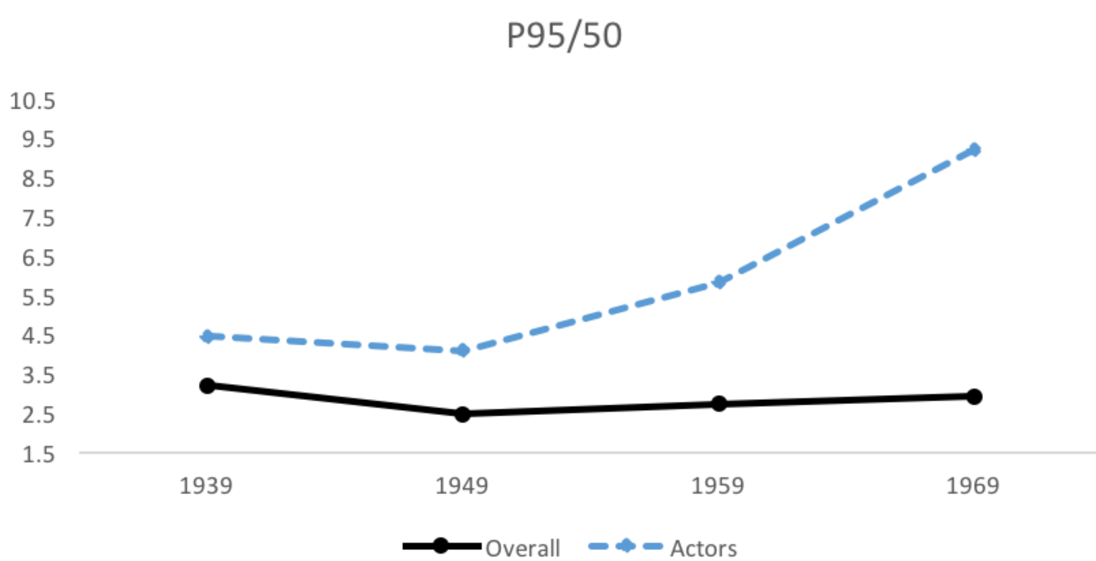


[Note] The mean for performance entertainers is 49 and for other leisure occupations 468. Data is from the US Census and covers the mainland US. Employment is measured per 100,000 inhabitants. For consistency with early Census vintages the employment measure includes the unemployed when they report an occupation. Performance entertainment is defined in the main text. Other Entertainment includes “drink and dine” occupations and “interactive leisure” professionals.



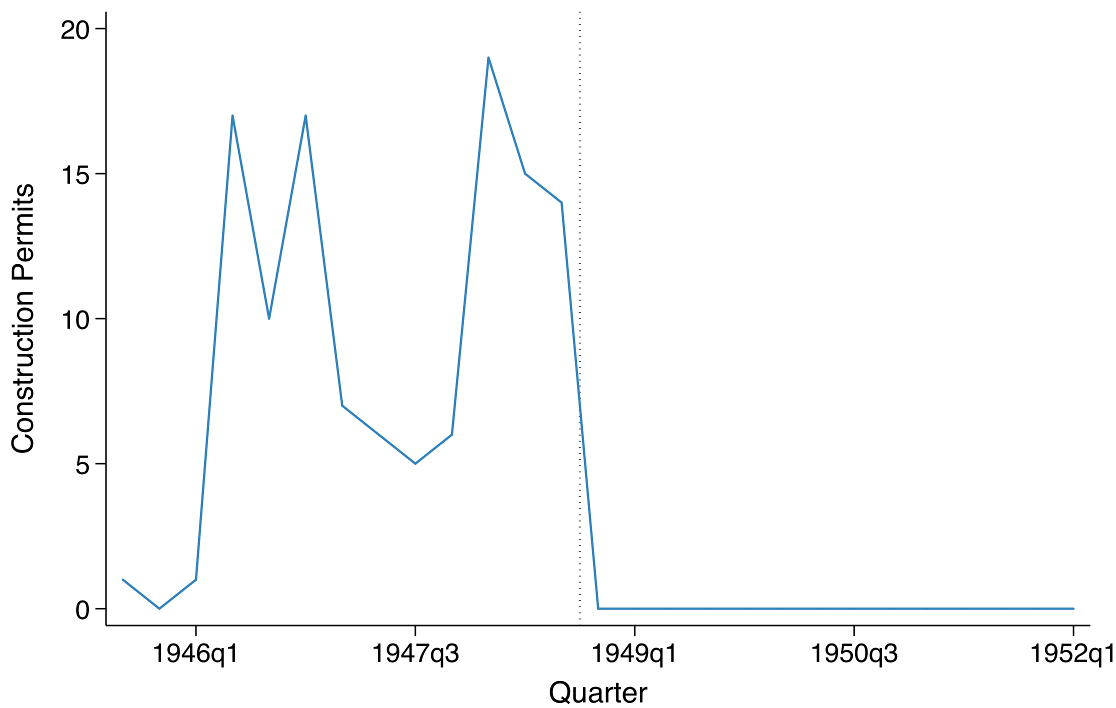
[Note] Performance venues are the venues listed in Julius Cahn-Gus Hill's 1921 theatrical guide. Size refers to the average seating capacity of the largest two venues in the commuting zone.

Figure 6: P95-P50 Gap



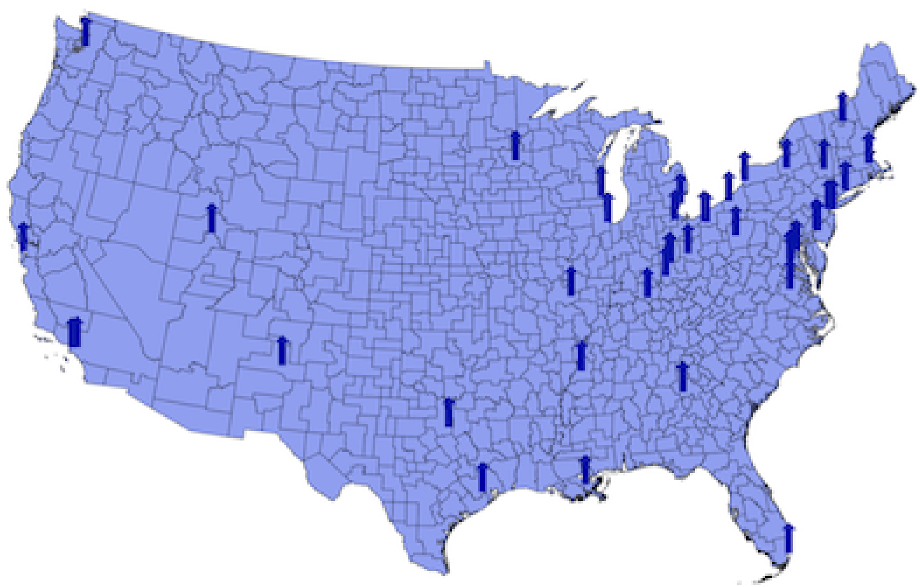
[Note] Figure reports the ratio of wages at the 95th and median. Percentiles are from the wage distribution reported in the US decennial Census for the lower 48 states.

Figure 7: Number of TV Licenses Granted



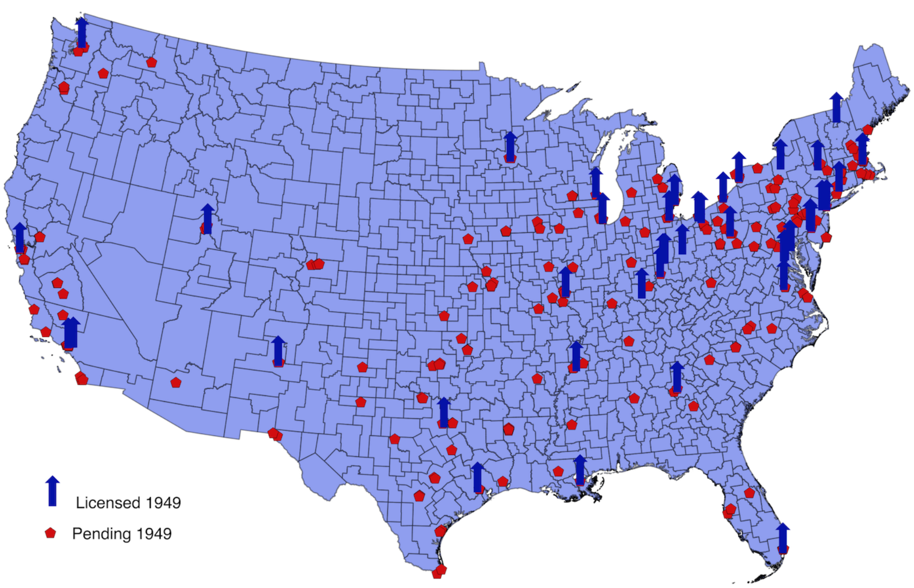
[Note] Data from TV Digest reports on FCC licensing activity. Missing construction permit dates are inferred from date of operation start.

Figure 8: TV Production 1949



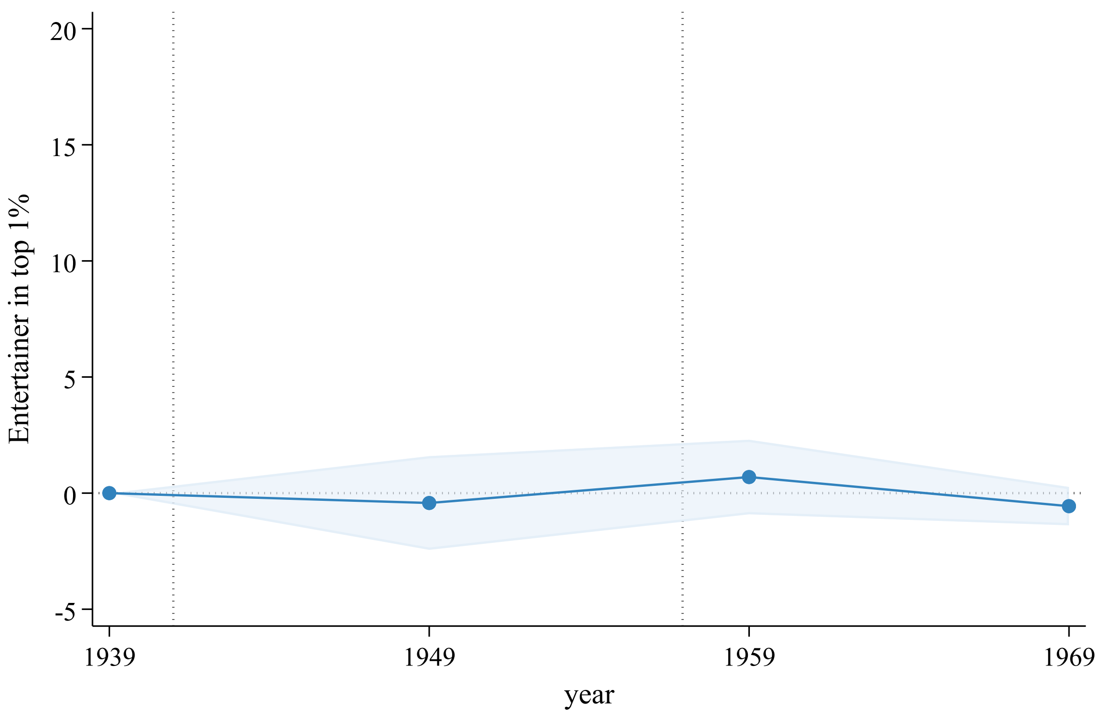
[Note] Arrows indicate an active broadcasting stations in 1949. Data is collected from Television Factbook. Location derived from geocoding addresses.

Figure 9: Granted and Frozen TV Licenses



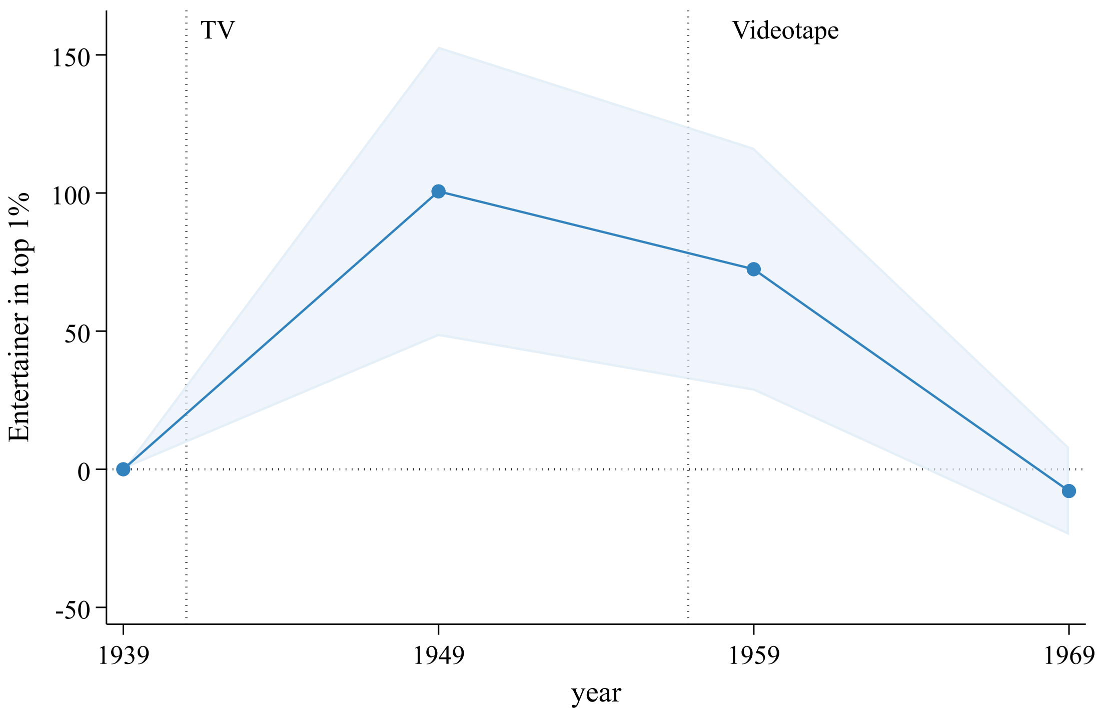
[Note] Arrows indicate an active broadcasting stations in 1949. Dots a pending application affected by the "freeze." Data is based on FCC reports as reported in Television Factbooks and TV Digest. Location is derived from geocoding addresses.

Figure 10: Placebo and Control Group Difference over Time



[Note] Plots time varying effect of local 1949 TV stations, β_t based on a difference in difference regression. The first vertical line represents the introduction of TV and the 2nd the invention of the videotape. The area shaded in light blue marks the 95% confidence interval. Standard errors are clustered at the local labor market level.

Figure 11: Dynamic Treatment Effect

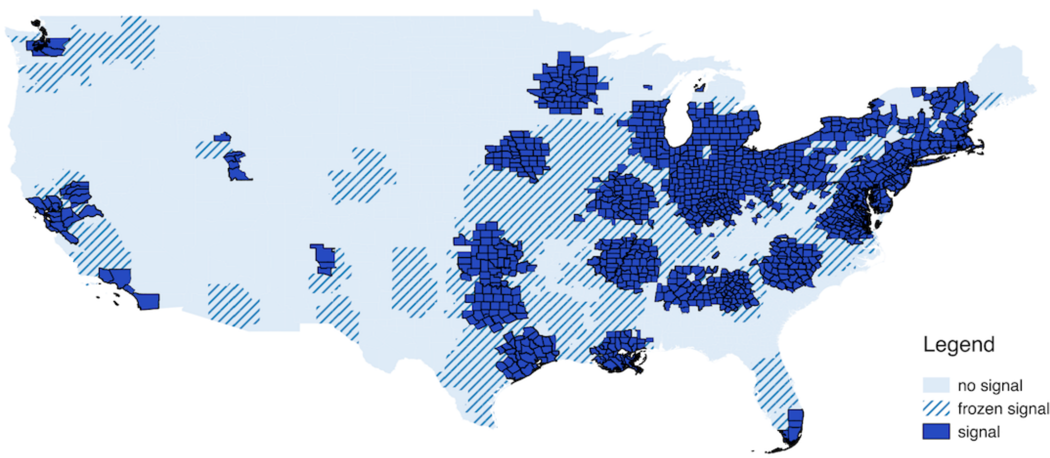


[Note] Plots the coefficient on $TV_{m,t}$ from a dynamic difference in difference regression. Vertical lines mark the beginngin of local TV ("TV") and the end of local TV ("Videotape"). The area shaded in light blue marks the 95% confidence interval. Standard errors are clustered at the local labor market level.

Figure 12: Effect of TV Station on Entertainment Wage Percentiles

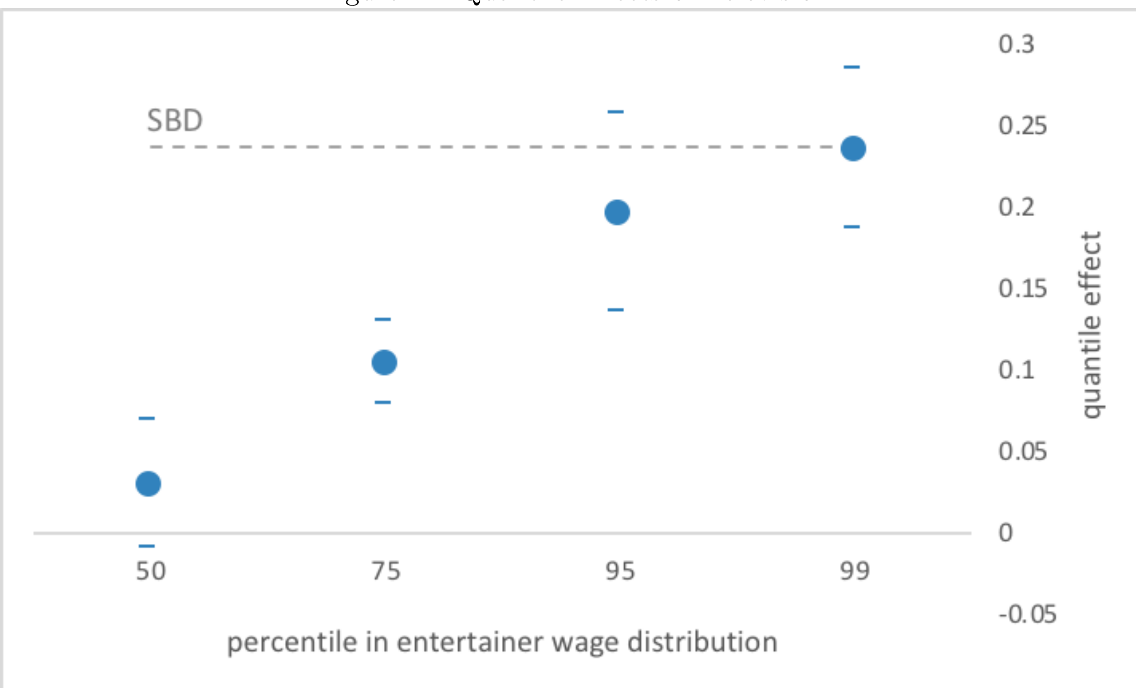
[Note] Each dot is the point estimate of a separate difference in difference regression. It shows the effect of adding a TV station to a local labor market on the change in entertainer employment in a given wage range (in percent). Percentile bins are defined in the overall US wage distribution. Standard errors are clustered at the local labor market level and 95% confidence intervals reported.

Figure 13: TV Signal of Licensed and Frozen Stations



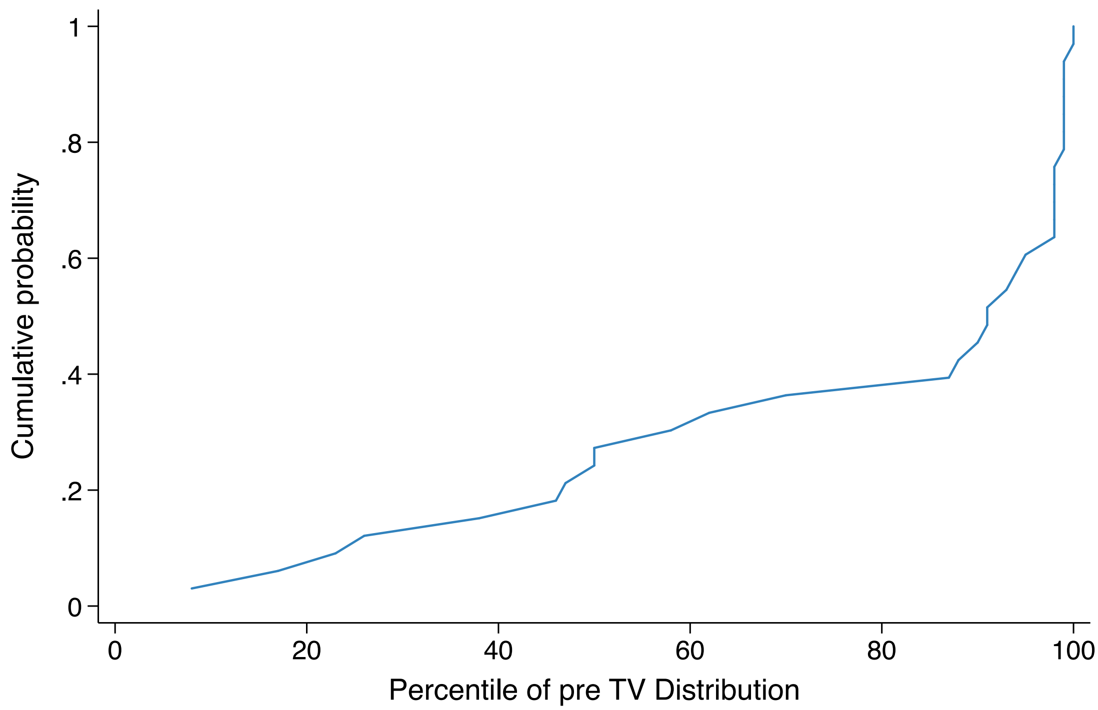
[Note] Signal coverage is calculated using an Irregular Terrain Model (ITM). Technical station data from FCC files is fed into the model (as reported in TV Digest and Television yearbooks). Signal is defined by signal threshold of -50 of coverage at 90% of the time at 90% of receivers at the county centroid.

Figure 14: Quantile Effects of Television



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

Figure 15: Wage Percentiles of Future TV Stars in 1939



[Note] TV stars are defined in the "who is who of TV" in "Radio Annual, Television Yearbook 1950." These individuals are linked to their 1939 Census records. 1939 wages are corrected for age, education and gender. The position in the distribution is calculated based on the residual of a regression of log wages on a cubic in age, 12 education dummies and a sex indicator.

Table 1: Effect of TV on Top Earner

	Share in Top 1%			Share of Top 1% from Entertainment		
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Entertainer			
Local TV station	90.19	93.16	127.6	0.389	0.325	0.301
	(26.25)	(26.16)	(47.68)	(0.160)	(0.157)	(0.182)
			Panel B: High Wage Professions			
Local TV station	0.962	8.596	13.83	0.115	0.023	0.084
	(5.879)	(13.78)	(6.632)	(0.422)	(0.365)	(0.439)
Demographics	–	Yes	–	–	Yes	–
Local labor market trends	–	–	Yes	–	–	Yes

[Note] Each cell is the regression coefficient of a difference in difference regression. The sample covers 1940-1970. All specifications control for local labor market fixed effects, time - occupation fixed effects and local production cost of filming in years after 1956. Demographics are median age, % female, % black, population density and trends for urban areas. Employment in top 1% is the number of workers with wages in the top 1% of the overall wage distribution, normalized by total employment in the occupation and multiplied by 100,000. Share of top 1% is the fraction of the US top 1% who works in the respective occupation, multiplied by 100,000. Mean of the dependend variables in panel A is 94 and 0.4, for panel B 97 and 12.97. Observations are 13,718 in panel A and 62,042 in panel B. Performance entertainer are Actors, Athletes, Dancers, Entertainers Not Elswhere Classified, Musicians. High Wage professions are Medical Professions, Engineering Professions, Managers and Service Professionals. Observations are weighted by local labor market population. Standard errors are reported in brackets, they are clustered at the local labor market level.

B Tables

Table 2: Employment effect of TV

	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)	
Demographics	-	yes	-	-
Local Labor Market Trends	-	-	yes	-

[Note] Each cell is the regression coefficient of a difference in difference regression, with the exception of column 4. "ln(Employment in Entertainment)" is the inverse hyperbolic sine of employment in entertainment. TV signal is a dummy that takes value 1 if signal is avilable in a commuting zone. Subscript "t+1" refers to the lead of the treatment. All specifications control for local labor market fixed effects and time - occupation fixed effects. Panel A covers 1930-1970, panel B and C 1940-1970. Standard errors are reported in brackets, they are clustered at the local labor market level.

Table 3: Effect of TV Studios on Wage Percentile Ratios in Entertainment

	Percentile Ratio (P_x/P_y)				
	(1)	(2)	(3)	(4)	(5)
	P_{99}/P_{95}	P_{99}/P_{75}	P_{99}/P_{50}	P_{95}/P_{50}	P_{75}/P_{50}
Local TV station	0.04	0.131	0.206	0.168	0.0745
	(0.0161)	(0.0377)	(0.0491)	(0.0542)	(0.0154)
Time & Labor Market FE	Yes	Yes	Yes	Yes	Yes

[Notes] Outcome P_x/P_y is the percentile ratio of percentiles x and y of the local entertainer wage distribution. Percentiles are calculated using the provided sample weights. Regressions controll for year and labor market fixed effects and local prodcution cost of filming in years after 1956. Observations are weighted by cell size. Standard error clustered at the local labor market level and reported in brackets.

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

	Share of Income		
	Top 1%	Top 0.1%	Top 10%
Local TV station	3.71	2.37	6.08
	(1.69)	(1.27)	(2.12)
Time & Labor Market FE	Yes	Yes	Yes
% growth	96%	239%	33%
% growth = top 1% growth (p-value)		0.0043	0.0000

[Notes] Top p% is the share of income going to the top p percent of entertainers in a given local labor market-year, in percentage points. The shares are based on Pareto interpolation where necessary. A test of whether the growth rate of the top p% income share ($g^{S_{p\%}}$) is the same as the growth rate of the top 1% income share is reported in the final row. The test is implemented in a regression with the ratio of top income shares as dependent variable. All regressions control for local labor market and year fixed effects. Standard errors are clustered at the local labor market level.

Table 5: Effect of TV Signal on Local County Fairs

	(1)	(2)	(3)	(4)
	Fair visits	Fair ticket receipts	Show receipts	Carnival receipts
TV signal	-0.085	-0.062	-0.083	0.032
	(0.037)	(0.045)	(0.045)	(0.049)
Time & Labor Market FE	Yes	Yes	Yes	Yes

[Notes] Billboard Cavalcade of Fairs 1946-1957 and irregular terrain model for TV signal. Outcome variables use the inverse hyperbolic sine transformation. Monetary variables are in 1945 Dollars. Treatment are the number of TV stations that can be watched in the commuting zone. Data on ride & carnival receipts is unavailable in 1953 and 1955. All regression controll for commuting zone fixed effects and year fixed effects. Standard errors are clustered at the commuting zone level.

Table 6: Spillover Between Labor Markets

	(1)	(2)	(3)
		A. Migration of entertainer	
local TV station	-0.00306	-0.00257	0.00476
	(0.00400)	(0.00436)	(0.00623)
Demographics	–	Yes	-
Local labor market trends	–	–	Yes
	B. Entertainer in top 1% - excluding areas neighboring treatment		
local TV station	94.01	96.77	133.2
	(27.09)	(26.58)	(49.15)
Demographics	–	Yes	-
Local labor market trends	–	–	Yes

[Source] Data US Census 1940-1970. Dependent variable is migration in panel A and entertainer in top 1% of the US wage distribution in panel B. All regressions control for commuting zone fixed effects, year fixed effects and film amenities after 1956. Demographics are median age, income and share female, minority and population density and different trends for urban areas. Standard errors are clustered at the commuting zone level. Neighbors are commuting zones with a land boundary.

Table 7: Effect of TV on Log Entertainment Audience

	(1)	(2)	(3)
	Ln(audience size)		
local TV station	1.432	1.459	1.504
	(0.213)	(0.195)	(0.234)
growth rate in market size	319%	330%	350%
Demographics	–	Yes	-
Local labor market trends	–	–	Yes

[Source] Data US Census 1940-1970. Dependent variable is the capacity of the largest venue in the commuting zone. The number of observations are 2656. All regressions control for commuting zone fixed effects, year fixed effects and film amenities after 1956. Standard errors are clustered at the commuting zone level. Demographics are median age, income and share female, minority and population density and different trends for urban areas.

Table 8: Effect Heterogeneity by Market Structure

	(1)	(2)	(3)	(4)	(5)
	Entertainer in top 1%				
local TV station	90.92	92.62	92.66		
	(36.67)	(0.213)	(0.195)		
local TV station * high tax state	0.842				
	(34.22)				
local TV station * population density		-0.025			
		(0.371)			
local TV station * theatre count			-0.035		
			(0.401)		
local TV station dummy				127.9	10.64
				(63.84)	(27.81)
multiple local TV station					206.8
					(101.1)

[Source] Sources and specification as in baseline. High tax states are defined as states where the marginal tax rates of the top income bracket exceeds the median. Theatre count are the number of employers listed in the Cahn-Gus Hills theatrical guide.