THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Studies of Labor Market Data

FELIX NIKOLAUS KOENIG

A THESIS SUBMITTED TO THE DEPARTMENT OF ECONOMICS OF THE LONDON SCHOOL OF ECONOMICS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY. LONDON, MARCH 2019.

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that this thesis consists of approximately 38,000 words.

Statement of inclusion of previous work

I confirm that work on Chapter 3 began prior to my enrollment in the PhD. The research paper has not been submitted for any prior degree.

Statement of conjoint work

I confirm that Chapter 2 was jointly co-authored with George Fenton and I contributed 50% of this work and Chapter 3 was jointly co-authored with Professor Alan Manning and Professor Barbara Petrongolo and I contributed 33%.

Statement of use of third party for editorial help

I confirm that parts of Chapter 1 of this thesis was copy edited for conventions of language, spelling and grammar by Kaley Joyce and Michael Beaney and Sarah Taylor of the LSE language centre.

Abstract

The thesis uses micro data and quasi experimental research designs to test three theories about labor markets.

The first chapter tests a leading explanation for top income growth, the superstar effect. The superstar effect attributes rising top incomes to expanding market reach of workers. I identify a case of exogenous market reach expansion in the entertainment sector and study the labor market effects. Incomes become markedly more concentrated on the top when entertainers can reach a bigger audience. Wages of stars grow 17% in response to a fourfold increase in market reach. A distinctive pattern of wage changes distinguishes the superstar model from alternative explanations. Growth of top pay occurs simultaneously with widening income differences at the top, a decline in middle-income jobs, an increase in low-paid jobs and a fall in total entertainer employment.

The second chapter tests how labor supply responds to improving entertainment technology. To identify the effect the chapter tracks the roll-out of TV signal. Social security records show that labor supply drops significantly with the introduction of TV. The effects are most pronounced for older workers, in line with descriptive evidence on changing retirement habits. The chapter shows that monetary spending substantially understates the value attached to TV.

The third chapter studies the canonical search and matching model and shows that accounting for realistic job search helps the model to account for labor market fluctuations and addresses the "Shimer puzzle." The chapter provides evidence that reservation wages significantly respond to backward-looking reference points. Introducing such reference-dependent job search to the model reconciles predictions on the cyclicality of both wages and reservation wages with the data. Other proposed solutions to the unemployment volatility and wage flexibility puzzle that hinge on alterations to the wage setting mechanism only work for parameter values outside the range typically estimated.

Acknowledgments

This thesis relied on the support of a great number of people.

Foremost, I thank my supervisor Steve Pischke, who has been a fantastic role model and supported and challenged me in countless ways throughout my PhD. I am particularly appreciative for the countless hours we spent with the all the projects that got discarded on the way.

Another key ingredient to this thesis was the fantastic research environment and outstanding academics at LSE and MIT. I am particularly grateful to David Autor and Alan Manning. David has been an amazing source of inspiration for the last two years. Thanks for your enthusiasm, attention to detail and your round the clock availability. And to Alan for the light hearted approach to research and the fun anecdotes that spiced up many research projects.

My co-authors and peers have been a crucial source of encouragement. Seeing projects through has been a breeze together with George Fenton, Barbara Petrongolo and John van Reenen. Just as valuable was the research community at CEP. Particular thanks goes to Jan Baker, Catherine Thomas, Vincenzo Scrutinio and Anna Valero who were incredibly helpful in drafting this research.

Without doubt, my PhD peers had the biggest impact on this thesis. The foundation for this thesis was laid during many problem set sessions and philosophical coffee breaks with Thomas Drechsel, Friedrich Geiecke, Niklas Moneke and Claudio Schilter. The joint supervisory meetings with Andres Barrios, Giulia Bovini, Giulia Giupponi and Stephan Maurer were invaluable and thanks for those shared memories of break-throughs, set-backs and frustrations. At least as important were the wind-down moments in the gym, pool and in the saddle with Wolfgang Ridinger, Thomas Störk, Arthur Seibold and Torsten Figueiredo Walter.

I am also grateful to the people who welcomed me during my research visits at Goethe University and MIT. To Horst Entorf for offering unconditional support during a difficult period, to Magda Davoli and Sunny Dutt for fantastic company at Campus West End. Thanks to Josh Angrist for hosting me in Cambridge and for being Master Joshway on the mountain bike trails. Thanks to my MIT office mates and the labor group, among others Jacob Moscona, Marco Tabellini, Camille Terrier, Joonas Thukuri, Matthias Wilhelm and Samy Young, the MIT cycling club and my flat mates for introducing me to the BSO, facial conducting and life in Cambridge.

Most of all I am indebted to my family. Foremost to Victoria who has kept me sane, cheered me up and has been the best company I could have wished for. To Anne, Lothar and Franca without whom I would not have made it to and through university. To the Russos and Richardsons, the past years would not have been the same without our amazing travels, bike rides and messages from every corner of the world.

Finally, I am grateful for financial support from the Economic and Social Research Council.

Contents

List of Tables

List of Figures

Introduction

This thesis consists of three chapters that study the functioning of labor markets. Over the past 60 years labor markets have changed dramatically. Two of the most striking trends are the decline in labor market participation and the sharp increase in income inequality. Economists have used models to shed light on the mechanisms that may drive these trends. In this thesis I test three prominent models empirically.

The three chapters use the same methodological approach and apply micro data and modern empirical methods to test hypothesis derived from economic models. The empirical tests use quasi-experimental tools to establish causal effects. The work thus combines empirical work with economic theory, which allows me to highlight the strength as well as areas of misspecification in prominent models of the labor market. The first chapter focus on the wage distribution and tests a leading theory of top income growth - "the superstar theory"; while the second and third chapter focus on explanation for changes to employment. The latter two chapters respectively study the search and matching model and the labor-leisure trade-off.

The first chapter studies top income growth and focuses on a leading explanation for such growth, the so called superstar effect. I present a tractable version of the model to illustrate the implication of the superstar effect. I then show how periods of expanding market reach of workers can be used to distinguish superstar effects from conventional models of labor markets. I use a historic period of location specific expansions in market reach in the entertainment sector to test the key predictions of the superstar model. Newly collected data on the licensing process of TV filming allows me to identify locations where the launch of television filming is delayed for exogenous reasons. The locally staggered variation in market reach gives rise to a differences-in-differences setting which allows me to test the superstar model. My results show that expanding market reach moves local wage distributions closer to winner takes all markets. Wages at the 99th percentile grow 17% while mid-paid jobs disappear and overall employment falls. Specific predictions of the superstar model distinguish the superstar model from alternative channels and confirm that superstar effects are driving the results.

The second chapter tests how labor supply responds to improving entertainment technology. Entertainment has improved rapidly over the past decades. This paper shows that better home entertainment options have led to a substantial decline in labor supply, particularly among the elderly. To identify the effect, we track TV signal during the introduction in the US and exploit variation from a regulated roll-out and terrain interference. Social security records allow us to measure how individual level labor supply responds. Our results confirms descriptive evidence that better leisure activities contributed to changes in retirement habits over the twentieth century. Finally, we use our estimates to quantify the forgone income from watching TV. Our results show that monetary expenditure on TV represents only a small fraction of total expenditure on this technology. Spending based measures like GDP therefore underestimate the value created by free-to-use technologies like TV.

The third chapter studies the currently dominant model of wage cyclicality, the search and matching model. The quantitative predictions of the canonical search model are at odds with the observed fluctuations in wages and employment in the labor market. We emphasize the role of reservation wages in wage cyclicality and argue that reference-dependence in reservation wages can reconcile model predictions and empirical evidence on the cyclicality of both wages and reservation wages. We provide evidence that reservation wages significantly respond to backward-looking reference points, as proxied by rents earned in previous jobs. We also argue that other proposed solutions to the unemployment volatility and wage flexibility puzzle that hinge on alterations to the wage setting mechanism only work for parameter values outside the range typically estimated.

Part I

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

1 Introduction

Rapid top income growth has been a striking feature of many labor markets in recent decades.[1] One of the leading economic explanations for this type of change in the wage distribution is the superstar effect.[2] According to this theory top income growth arises when workers can apply their talent on a bigger scale. As it gets easier to reach many consumers simultaneously, a greater share of consumers will flock to the most talented workers in the profession – the "superstars". Such a shift in demand creates rising incomes at the top and simultaneously reduces incomes for less talented workers. The superstar effect can therefore explain why top incomes are growing much faster than average incomes and rationalize rapid top income concentration. This theory has a long tradition in economics and has been used widely to explain labor market trends, it has however rarely been tested.[3]

This paper uses a historic natural experiment to directly test the predictions of the superstar model. Workers are increasingly able to reach larger markets with the help of modern technologies. I use a historic setting to identify the effect of such changes on labor market returns and use it to test the superstar model. The entertainment sector provides a unique setting for such a test. The launch of TV in the mid 20th century had vastly increased the audience available to entertainers. Before the introduction of TV, a live performance could be watched by a few hundred individuals, while after the introduction of TV, the same performance could be watched by millions. Technological constraints limited TV filming to locations near broadcast antennas, as a result, TV was characterized by multiple local TV stations that independently broadcast content to the local population.[4] For a local entertainer, the construction of a TV station was therefore

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

[2] Applications of the superstar model include Gabaix et al. (2016); Terviö (2008); Gabaix and Landier (2008); Garicano and Hubbard (2007); Cook and Frank (1995).

[3] Classic articles on the superstar model include Tinbergen (1956); Sattinger (1975); Rosen (1981). For a recent applications of this theory to the digital economy see Bas et al. (2018); Guellec and Paunov (2017); OECD (2016); Acemoglu et al. (2014).

[4] TV shows were effectively a non-tradable service. Recording was, in principle, possible in the form of "kinescopes." However, the image quality of this technology was poor and such TV displays were unpopular. Shows produced elsewhere were a poor substitute for local productions. TV networks, which later harmonized programming across the US, initially had a limited influence over local programming.

a substantial shock to market reach, similar to the construction of a hypothetical giant theater that would hold an entire local population.

A growing number of local labor markets got access to TV filming during the staggered local deployment of TV stations. I study the effects of this roll-out in a difference in difference analysis across local entertainer labor markets and find that the launch of a TV station leads to sharp growth of top incomes, while simultaneously eroding demand for mediocre workers. The effects align closely with the superstar model but are at odds with conventional alternative models.

The roll-out of TV has exogenous elements that allow me to address three empirical challenges that made it difficult to test for superstar effects. A first challenge is that changes to market reach need to occur exogenously to local labor market shocks, which is rarely the case with ordinary endogenous technology adoption.[5] In the case of TV on the other hand, technical change is introduced through a government licensing scheme. I exploit regulatory constraints to generate local variation in access to TV that is exogenous to local labor market conditions. One such feature is the sudden interruption of licensing in 1948 that became necessary due to signal interference between stations. I identify stations that were about to launch but narrowly missed out due to the license freeze. Such places that narrowly miss out on TV launches allow me to probe the identification assumptions and test for spurious effects in the government led roll-out process.

A second challenges is to isolate the effect of expanding market reach from other drivers of top income growth. The modern boom in market expanding technologies for example coincided with other trends that affect top incomes, such as deregulation and pay setting norms. The recent correlation of expanding market reach and top income growth may therefore reflect spurious effects. In the entertainment setting I can hold aggregate changes constant and exploit the fact that different parts of the US experience the effect at different times.

A third challenge for a test of superstar effects is that most innovations simultaneously affect many aspects of the economy. Digital technologies, for instance, enable workers to serve bigger markets but also affect up and down stream markets, which makes it difficult to isolate the effect of worker market reach. TV, to the contrary, was used to broadcast entertainment shows and had no use in production of the rest of the economy. This allows me to isolate the effect of changing market reach from effects that occur in other industries.

A further advantage of the entertainment setting is that the entertainers' audience size and it's change through the TV roll-out can be quantified, overcoming one of the key measurement issues. I built a novel dataset from archival records that makes changes in production and consumption of entertainment visible. On the production

[5] Evidence for such endogenous technical change is presented in Blundell et al. (1999), the theory in Acemoglu (1998).

side, the data show where, when and why TV filming became feasible. Specifically, the data include information on the universe of broadcasting licenses of TV stations, their locations and audience sizes, as well as the historical capacity of over 3,000 performance venues. I combine this information with administrative records on the TV station licensing process, including information on how locations were prioritized. On the demand side, the data quantify the shift in labor demand. I digitize archival sources that report spending at roughly 4,000 local entertainment venues and contain information on prices and show revenues. With this data, I can trace demand shifts and associated changes in entertainers' marginal revenue product. These records are linked to US Census micro-data that capture labor market outcomes in entertainment and beyond.

My findings confirm the headline prediction of the superstar model that growing market reach causes top income growth. A local TV station boosts pay at the 99th percentile by about 17% and expands the audience by roughly 300%. To put this wage growth into context, I look at the position of entertainers in the US wage distribution. Local star entertainers rise markedly in the US wage distribution when a TV station is launched in the labor market. The share of local entertainers in the top 1% of the US wage distribution almost doubles. Locations that narrowly miss out on the launch of a TV station see no growth in top entertainer pay. Similarly, I find no effect on other professions. The superstar effect is specific to the time periods, places and professions involved in local TV filming, reinforcing confidence that the effect is caused by TV.

Next, I show that the superstar effect differs from canonical models of technical change. To distinguish superstar effects, I derive additional predictions that are specific to the superstar model. In cross-sectional data, superstar effects and the effect of canonical labor demand shifts are indistinguishable. However, wage changes over time differentiates the superstar model from a wide range of alternative models. Specifically, in a superstar model labor markets move closer to a winner-takes-all market when it becomes easier for workers to reach a bigger market. These effects are captured by four testable predictions: (i) disproportionate wage growth at the top, (ii) decreasing wages for mediocre workers, (iii) falling employment and (iv) growing wage dispersion at the top. The empirical results confirm these patterns. Expanding market reach has a striking U-shaped effect across the wage distribution, characteristic of the superstar effect.[6] The gains at the top occur together with a decline in mid-income jobs and a growing low-pay sector. Moreover, I confirm that pay differences among top earners increase and document substantial employment losses in entertainment. When TV signal becomes available in a local area, entertainment employment declines around 13%. These results show that demand is becoming concentrated on star workers, at the expense of mediocre

[6]The middle of the income distribution also hollows out in models of routinization where technology replaces mid-skilled workers (e.g. Goos et al., 2010). This differs from superstar models, where mid-skilled workers are replaced by star workers and technology acts as a vehicle for stars to project their talent.

workers.

Moreover, I can measure the shift in labor demand directly by studying spending data on different types of entertainment. This allows me to go beyond analyzing labor market outcomes and test the supposed underlying demand shift directly. The results show that local TV filming increases the audience and revenues for the biggest local shows while drastically reducing attendance at ordinary local live entertainment .

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

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

Spillovers between local labor markets could bias estimates based on local labor markets relative to the effect of an aggregate shock. I assess such differences by studying spillovers between markets. A major spillover channel is shut-down in this setting since live entertainment shows are by nature consumed locally and there is no cross-labor market trade in output. The main potential link between local labor markets is entertainer mobility. I quantify the mobility effects and find that they only play a minor role for the findings.

In a second extension, I explore how superstar effects interact with imperfect competition.[7] The predictions of the superstar model change substantially in an imperfectly competitive labor market. Monopsony employers no longer pass on gains from technical progress to workers. The regulated entry of TV stations allows me to

[7] Imperfect competition features prominently in the market access literature. Integration of markets could give rise to entry effects that intensify competitive (Melitz and Ottaviano, 2008). Monopsony power and rent-sharing have also been linked to pay inequality (e.g. Manning, 2003; Benabou and Tirole, 2016).

test this empirically and analyze how competition affects the magnitude of superstar effects. In line with the superstar model but contrary to popular belief, it is not the lack of competition that raises top incomes, but rather more intense competition for talent.[8]

The superstar model is a classic model in economics that was first presented six decades ago and became popular through a series of articles in the 70s and 80s that emphasized that the model could explain dramatic concentration of labor market returns at the top of the distribution (Tinbergen, 1956; Sattinger, 1975; Rosen, 1981). Despite this long tradition, there is no common modeling framework to study superstar effects. In an attempt to structure the literature, I develop a unifying framework that nests many of the existing superstar models (including Costinot and Vogel, 2010; Terviö, 2008; Gabaix and Landier, 2008; Teulings, 1995; Rosen, 1981; Sattinger, 1979, 1975). I use a benchmark version of the model to show how improving production scalability affects the demand for talent and, ultimately, wages.[9] Previous empirical applications use the superstar model to explain the distribution of CEO pay (Edmans and Gabaix, 2016; Gabaix and Landier, 2008; Terviö, 2008). Such studies calibrate key model parameters to the correlation of pay at the top and market size. My study instead focuses on distinguishing the superstar model from leading alternative models. In a link to the previous literature, I additionally provide a comparison of OLS and IV estimates for the key elasticities of the model.

A number of studies have shown that technical change has profound effects on labor markets. Canonical models of technical change include models of efficiency units (Stigler, 1961), skill biased technical change (Acemoglu and Autor, 2011; Katz and Murphy, 1992) and routine bias technical change (Acemoglu and Restrepo, 2018; Autor and Dorn, 2013; Goos et al., 2010). I show how the superstar model differs from such models and derive testable predictions that allow me to distinguish the models in the data. Empirical evidence for the canonical models use variation in technology to test the predictions of those models (e.g. Acemoglu and Restrepo, 2017; Michaels and Graetz, 2018; Akerman et al., 2013). In line with this work, I exploit exogenous technical change, but in contrast to those studies, I analyze a technical change that expands market reach in a single industry and test for superstar effects.

Recent work has applied the superstar model beyond the labor market and showed that superstar effects can account for growing market concentration in product markets. When applied to firms, the superstar model rationalizes increasing dispersion in firm size and changing factor shares (Eeckhout and Kircher, 2018; Autor et al., 2017). There

[8]Rents are emphasized as an explanation for top income growth in Baker (2016); Benabou and Tirole (2016); Piketty et al. (2014); Murphy et al. (1993); Bok (1993). Evidence for rent-sharing has been documented in Kline et al. (2017); Bertrand and Mullainathan (2001), while De Loecker and Eeckhout (2017) find that rents have risen over past decades.

[9]The standard approach uses one-to-one matching. A related literature models superstar effects in terms of span of control, where one worker is matched to multiple units (Geerolf, 2014; Garicano, 2000; Rosen, 1981).

is growing concern that internet-based technologies lead to sharp increases in market concentration, which some observers link to rising mark-ups and rents (for evidence on rising mark-ups see De Loecker and Eeckhout, 2017). I show that market concentration and mark-ups need not go hand in hand. In the superstar model integrated markets reallocate resources to more talented workers and market concentration can arise in a fully competitive setting.

There is a sizable literature that studies the social consequences of television watching. Such studies find that watching TV affects political attitudes, consumer behavior and educational outcomes (among others Cantoni and Bursztyn, 2016; DellaVigna and Kaplan, 2007; Durante et al., 2015; Chong and La Ferrara, 2009; Fenton and Koenig, 2018; Gentzkow, 2006; Gentzkow and Shapiro, 2008; Olken, 2009; Putnam, 1995). Different from the literature on consumption of TV shows, this paper focuses on the production of TV shows. I use novel data on TV filming to test for superstar effects in the labor market for entertainers.

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

2 Model

This section develops a tractable model of the superstar effect that illustrates the key predictions of the model and distinguish it from conventional models of labor demand. The term "superstar effect" has been used to describe different concepts, the aim of this section is to clarify the meaning and derive a definition from the superstar model. I will show that many of the superstar models' predictions can be replicated by conventional models of labor demand. Finally, I illustrate how technical progress generates predictions that differentiate the superstar model from a wide class of alternative models. A more general version of the superstar model is presented in Appendix 8.C. This model provides a unifying framework that nests the various existing versions of the superstar model, shows their connection and is used to illustrate the key model properties.

2.1 A Benchmark Superstar Model

A superstar model is an assignment model where heterogenous workers are matched with heterogenous tasks. In the context of entertainment we can think of workers as actors and of tasks as shows. Actors have different and unique talent (t) and the talent types can be ranked, actors are thus vertically differentiated. Shows differ in their innate productivity characteristics, think of these characteristics as the performance venue's audience capacity, or size denoted by s.[10]

A general superstar model that nests the standard versions of superstar models in a unifying framework is presented in Appendix 8.C. Here I will illustrate the key mechanics of the model by developing a benchmark version, building on Sattinger (1979), that allows for closed form solutions.

Labor Supply and Demand

Since workers are differentiated, we need to characterize the labor supply of each worker type. Assume that each worker supplies one unit of labor inelastically, the labor supply then is the same as the distribution of worker types. In the same way labor demand is characterized by the distribution of venue sizes. The benchmark model makes the simplifying assumption that both actor abilities and show sizes follow a Pareto distribution (More general results are illustrated in Appendix 8.C). Denote the probability that an actors' talent is above a threshold t by p^t and the equivalent for shows by p^s. I denote by x_p the value of a variable x at percentile p . The inverse CDFs of the two Pareto distributions are then given by:

$$p^t = t_p^{-\frac{1}{\alpha}} \tag{1}$$

$$p^s = s_p^{-\frac{1}{\beta}} \tag{2}$$

The distribution of talent and venue size is characterized by the shape parameter of the respective Pareto distribution. Here the shape parameter is the inverse of the exponents, respectively α and β and a bigger value implies greater dispersion. Next, assume that workers and shows are matched one-to-one, each show hires exactly one actor and an actor performs in one show.[11] One-to-one matching has been widely adopted in the superstar literature to keep the model simple (e.g. Gabaix and Landier (2008); Terviö (2008)), but extensions to one-to-many matching lead to similar results (as in Garicano

[10]In the literature, differences in job characteristics are often referred to as "market size" or "firm value". Important to the model is that the characteristics are innate and cannot be changed at the time of hiring. Further, note that these characteristics are not the same as the employer's market value, which depends on both innate characteristics and endogenous factors such as talent employed and the price of talent. In the empirical section I will address how to distinguish a change in S_i from the endogenous firm value Y.

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

(2000); Rosen (1981); Sattinger (1975)).[12]

Production

A matched actor-show pair produces revenue $F(s,t)$. The key assumption of a superstar model is that more talented workers have a comparative advantage in larger markets, which in the entertainment context implies that adding an extra seat to a theater affects revenues more when a better actor is performing. In other words, the superstar model assumes that $F(s,t)$ is super-modular.[13] A Cobb-Douglas production function guarantees this and allows for a simple closed form solution. I therefore assume that production revenues are given by:

$$F(s,t) = \pi s^{\gamma} t^{\delta}$$

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

Equilibrium

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

$$p^{\hat{t}} = p^{s}(\sigma(\hat{t})) \iff \sigma(\hat{t}) = \hat{t}^{\frac{\beta}{\alpha}} \tag{3}$$

$$w'(\hat{t}) = F_t(\sigma(\hat{t}), \hat{t}) = \delta \pi \hat{t}^{(\frac{1}{\xi}-1)} \tag{4}$$

Equation 3 is a formal expression of PAM, it states that percentiles in the size and talent distributions are the same. We can use this equilibrium condition together with the inverse CDF functions 1 and 2 to solve for the matching function $\sigma(t)$. The second

[12]The one-to-many matching features equilibrium cut-offs that determine which share of jobs is performed by which type of workers. Highly talented actors' comparative advantage in juggling many shows implies that they serve a greater share of the shows.

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

equilibrium condition states that the wage increase for a marginally more talented worker equals the marginal product of the worker in the equilibrium assignment. The second equality uses the equilibrium assignment from equation 3 to eliminate s and write wages in as a function of equilibrium talent $\hat{t}$. The exponent is defined as $\xi \equiv \frac{\alpha}{\delta\alpha+\gamma\beta}$.

The resulting equilibrium is perfectly competitive in the sense that there are no match specific rents. Despite the fact that both workers and venues are monopolists over their types no worker earns rents over their next best employment option. This is an artifact of the continuity assumption of types. The outside options for both actors and show producers are infinitesimally worse and thus ensure competitive renumeration of marginal talent units. If we relax the continuity assumption match specific rents can arise. Take the alternative case, where the distribution of show types has jumps; some theater venues are thus discretely bigger than their competition. Here the show producer does not have a direct competitor that would bid up wages and thus he will keep all the productivity gains. A lack of competition among employers therefore dampens wages. While there are no match specific rents, notice that workers earn rents over the outside option which we normalized to zero. Participating in the labor market is therefore beneficial for all inframarginal workers.

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

$$w(t) = \xi\delta\pi t^{1/\xi} \tag{5}$$

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

$$p^w = \lambda {w_p}^{-\frac{\xi}{\alpha}} \tag{6}$$

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

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

$$ln(w_p) = ln(\xi\delta\pi) + \frac{\alpha}{\beta\xi} ln(s_p) \tag{7}$$

Wages are a function of market size and wage growth is thus proportional to changes in market size. I call this relation the "superstar effect." The related literature has used this result to generate two insights. First, dispersion in firm size does not grow quickly enough in a random growth model to generate transition dynamics that account for the sharp rise in income concentration in the US (Gabaix et al. (2016)). However, with a more nuanced growth process, the superstar model matches the data. Second, CEO pay can be explained by this relation when firm values are used as an empirical analogue to the size distribution (Gabaix and Landier (2008); Terviö (2008)). Although these results illustrate the model's potential power, they do not preclude the possibility that other factors cause the relationship between wages and market size. A similar relation arises from alternative models; most notably, models of endogenous technical change link labor productivity and firm productivity (see Blundell et al. (1999) for an empirical illustration).

2.2 The Effect of Technical Change

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

[14]This additional restriction is not required in other versions of the superstar model. For example, Sattinger (1975) assumes log super-modularity in production and does not require additional assumptions on the spacing of the distributions.

distinguish different models of technical change, I derive patterns of wage changes from the superstar model that are distinct from conventional models of technical change.

2.2.1 Superstar Effects and Technical Change

The superstar effect is the result of expanding markets reach. A tractable way of modeling such a change is allowing production to become more scalable and thus reducing the diseconomies to scale in the production function. Assume that this change takes the form of $\delta' = s \cdot \delta$ and $\gamma' = s \cdot \gamma$ with $s > 1$.[15] The new wage distribution (call the new period $t+1$) is therefore found by substitute the new values ξ' and λ' into equation 6:

$$p^{w}_{t+1} = \lambda' w_p^{\frac{-\xi'}{\alpha}} \tag{8}$$

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

$$g_{w_p} = \frac{w_p^{t+1}}{w_p^t} = \psi p^{\frac{-\alpha}{\xi}(s-1)} \tag{9}$$

Where $\psi = (\frac{\lambda'}{\lambda})^{\frac{\alpha(s-1)}{\xi}}$. These equations reveal, that the reduction of diseconomies to scale has differential effects at different parts of the distribution. The effect are summarized in Figure 1, the wage distribution shifts inward and pivots out. The intuition is that more productive workers are matched with bigger shows and therefore operate on a bigger scale, diseconomies to scale are more binding for this group. Such top workers therefore benefit most from better scalability of production. The effect can be seen in two changes in equation 8, the shape and scale parameter of the Pareto wage distribution change. Compared to equation 6 $\xi' = \frac{\xi}{s} < \xi$, which implies that wage differences between workers grow, the wage distribution pivots out and the distribution becomes more right skewed. Besides this top income growth, there is an additional level effect on the wage distribution operating through λ'. This is a level effect that reduces wages at all levels. The level effect is a consequence of expansion in the availability of entertainment, since more entertainment is being produced, the entertainment market clears at a lower price for talent π.[16] As a result the Pareto scale parameter falls

[15] An alternative but ultimately equivalent way of modeling this change is to allow the size distribution to change, for example by increasing the shape parameter α (see Appendix 8.C).

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

($\lambda' < \lambda$) and the wage distribution shifts inward (see equation 8).[17] This case illustrates one of the key features of a superstar model, the potential for cannibalization effects. The greater availability of stars, reduces demand for the rest of the profession and in the limit, a single superstar serves the entire market. In summary, the bottom of the distribution benefits little from better scalability but suffer from the fall in the price for talent units, while at the top of the distribution the bigger scalability over-compensates for the fall in π. Previously unattained income levels are reached at the top and bottom ends of the distribution, while mid-paid jobs simultaneously disappear.

For empirical tests, it will be useful to derive separate predictions for different parts of the distribution. I will illustrate the effect of technical change by deriving which types of jobs are created and which ones are being destroyed. Consider the number of jobs that pay wage w, given by the density of the wage distribution $f(w)$. To derive the density take the derivative of 5 with respect to w and multiply by minus one:

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

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

$$g_e(w) = \frac{f_{t+1}(w)}{f_t(w)} = \frac{\lambda'\xi'}{\lambda\xi} w^{\frac{(s-1)}{s}\frac{\xi}{\alpha}} - 1 \tag{10}$$

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

[17] Notice that if π is unchanged (ie if demand for entertainment is perfectly elastic), λ would rise. I assume that demand is sufficiently inelastic to rule this case out.

that has support in both distributions. By using wage bins I can compute growth rates for wage bins that span the full wage distribution.

2.2.2 Alternative Models of Technical Change

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

To examine the differences between the heterogenous workers who are perfect versus imperfect substitutes, I focus on a single group. Specifically, I will abstract away from the low-skill group and focus on income dispersion among the high-skilled.[19] First consider the baseline case, where labor supply is perfectly inelastic and all workers with skill above $\bar{p}$ are participating in the market.[20] A skill-biased demand shift (SBD) increases the demand for talent $D(\pi)$ to $D'(\pi) > D(\pi)$.[21] Market clearing implies that increase in demand for talent increases the price of talent π to π'. A unit of talent

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

[19]The results for the fully fledged model are equivalent and presented in Appendix 8.C.

[20]This assumption is immaterial here but becomes relevant if one introduces matching (as in Eeckhout and Kircher (2018)).

[21]In the conventional model, the demand shift is a result of a skill augmenting change in productivity (see Appendix 8.C). Here we take the reduced form approach of modeling the skill-biased demand shift as a change in the demand for talent.

becomes more valuable, and the more talent a worker has, the more she benefits from the growth in π. After a SBD shock in period $t+1$ wages at percentile p are given by:

$$w_p^{t+1} = \pi' \cdot q_p = w_p^t \frac{\pi'}{\pi}$$

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

$$g_p^w = \frac{\pi' \cdot q_p}{\pi \cdot q_p} = g^w \tag{11}$$

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

$$p_t = (w/\pi)^{-\frac{1}{\alpha}}$$

The growth in the skill premium to π' leads to an outward shift in the wage distribution that is illustrated in logs in Figure 2. First, notice that the original wage distribution is identical to the result of the superstar model. With the right assumption on its parameters, the SBTC and superstar models yield the same result and makes the two models indistinguishable in cross-sectional data.

2.2.3 Testable Differences

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

[22] To cut through the debate on assumptions related to the talent distribution, I show which talent distribution is required for this model to match the 1939 wage distribution and what shift in the skill premium is needed to account for the growth in top earners between 1939 and 1969. The predicted wage change pattern for the rest of the distribution is shown in the Appendix Figure 12.

and bottom parts of the wage distribution, where mid-income jobs disappear and low-pay jobs emerge. These effects are summarized by four testable propositions:

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

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

Proposition 2.2. *Mediocre worker pay: In a superstar model $w^{t+1}_p < w^t_p$ is feasible, while in a SBD model $w^{t+1}_p > w^t_p$ at all percentiles.*

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

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

With entry and exit, the participation threshold $\bar{p}$ determines which worker are active in the market. Employment declines in a superstar model as stars' growing reach pushes other workers out of the market. With positive demand shocks, by contrast, quantity and wages move in the same direction. The price for talent π rises in response to SBD shocks but falls with superstar effects. The proposition then follows from the market clearing condition in equation 25, meaning that a higher price for talent leads to more participation.

Next I focus on dispersion in top incomes and focus on the predicted change in top income shares.

[23] With additional skill groups this holds approximately for the highly talented individuals within a skill group.

Proposition 2.4. *Dispersion at the top: Income differences within the top tail increase with superstar effects but not with SBD shocks. This implies for top income shares (s_p) at two percentiles p: $s^{t+1}_{1\%}/s^{t+1}_{10\%} > s^{t}_{1\%}/s^{t}_{10\%}$ in a superstar model and $s^{t+1}_{1\%}/s^{t+1}_{10\%} = s^{t}_{1\%}/s^{t}_{10\%}$ in a SBD model .*

This proposition highlights the income dispersion *within* the top tail. A superstar model exhibits a fractal inequality, so that moving up a rank in the talent distribution becomes more valuable. As a result, a growing proportion of the the income share of the top 10% is earned by the top 1% and consequently the ratio of the two $(s_{1\%}/s_{10\%} \uparrow)$ increases. The same increase in fractal inequality does not hold in the SBTC model, where the relative pay differences remain stable. The proposition is derived in Appendix 8.C.

A natural question is whether extensions to the SBTC model allow it to replicate these results. The key distinction highlighted so far is that a SBTC model features groups of perfectly substitutable workers, whereas workers are imperfect substitutes in the superstar model. However, there are additional differences between superstar and SBTC models. To see this, consider the case where there is a continuum of skill groups in the SBTC model. Workers in two different skill groups are imperfect substitutes, with a continuum of skill groups each worker is his own skill group and hence all workers are imperfect substitutes. This extended SBTC model can generate fractal wage inequality, it requires technical change that increases productivity in an escalating fashion towards the top. However, the model will not feature cannibalization effects. A positive demand shock translates into gains across the range of the distribution, which is proved in Appendix 8.C. Even a SBTC model where all workers are imperfectly substitutable will therefore not feature cannibalization effects and will not replicate propositions 2.2 and 2.3.

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

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

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

3 Data

I collect novel data on the production and consumption of entertainment in the middle of the 20th century from archival sources. Consumption data includes local-level consumer spending and attendance at entertainment venues, while the production data includes information on local inputs and production technology. These data are linked to entertainers' labor market records.

3.1 Production Technology

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

Television Filming Data on television facilities come from the "Annual Television Factbooks" which records the address, technical equipment, launch date, assigned channel and call letter for each TV station.[24] I geocode the location of TV studios and match them to the local labor market to track the roll-out. The launch of TV filming provides one of the main sources of variation in the analysis, Figure 3 shows where broadcasting took place in the year 1949, a year with Census wage data. For each year I compute the exposure to local TV filming by summing the number of active stations in the local labor market and therefore assume that all stations were filming locally at that time. There are a handful of exceptions, as a few stations operated a local network. These interconnected stations could relay local shows to nearby stations through upgraded phone lines (run by AT&T) or microwave relay technology (run by Bell). Interconnection was rarely feasible because the technical infrastructure was still in its infancy. In my main specifications I code all members of such networks as treated. This approach avoids potential endogenous selection of filming locations within the network.[25]

TV Licensing Detailed information on the licensing process allows me to identify places that narrowly miss out on TV launches during the license freeze. The freeze in licenses began in 1948 and continued until 1952. The data are based on the weekly bulletins from the Federal Communication Commission (FCC), which are summarized annually in the "Television Factbook." From the same records, I collect information on the rule used to prioritize locations. This rule was published in a few years and reveals that the priority ranking of the TV roll-out was based on fixed location characteristics. This lends credibility to the assumption of the difference in differences regression, that the TV timing did not respond to local demand shocks.

[24] This data source has previously been used by Gentzkow (2006) to build a dataset of TV signal coverage throughout the US.

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

Videotape Filming Local filming was ultimately superseded by centralized productions. With the invention of the videotape in 1956 television production shifted away from local stations, towards places where conditions for filming are most favorable. Centrally produced shows could then be broadcast across the country at low cost and at the time appropriate for the local time zone. The rise of centralized filming leads to a decline of local TV filming which allows me to test whether local superstar effects disappear.[26] To do so, I will control for places where filming centralizes, which raises the potential challenge of an endogenous control variable. To avoid this endogeneity issue, I proxy locations of centralized filming with locations where movie filming took place in the 1920s. This measure is pre-determined and picks up location incentives that come from permanent regional characteristics. The data on the location of film shoots come from the "Internet and Movie Database" (ImDB). ImDB is a widely used platform (self-proclaimed number one worldwide) for information on movies and holds metadata on over 4 million movies. In 1920, around 200 movies were produced in the US. For each labor market I compute the share of movies produced in this market in 1920.

3.2 Demand Data

Television Broadcasting Data on TV signal allows me to establish when local entertainer start facing competition from TV entertainment. Places that are exposed to TV signal are not necessarily the same as labor markets that produce TV shows, as signal airwaves travels beyond the local labor market of the TV station. Information on TV stations' signal reach comes from Fenton and Koenig (2018), who re-construct historic catchment areas.[27] Figure 4 shows the variation in TV signal in 1950 and illustrates areas that narrowly miss out on TV signal due to the freeze in licensing. I combine the information on each TV station's catchment area with Census data on household location and TV ownership and compute the audience of TV shows. The median TV station could reach approximately 75,000 households. Even the smallest TV audiences substantially exceeded the show audiences of local venues.[28] Additionally, I compute

[26]The invention of the videotape is another example of a technology that expanded market reach. However, this variation is less suitable for a test of the superstar model, the location of filming for videotape is not exogenously imposed. Instead markets with the lowest production cost were selected for filming. Production cost are determined by endogenous factors such as local wages and tax rates as well as fixed location characteristics (e.g. sunshine hours, availability of equipment and expertise, local scenery).

[27]For details on the data construction see Fenton and Koenig (2018). They use an irregular terrain model (ITM) to calculate signal propagation. In this model, signal reach depends on the technical properties of an antenna (channel, frequency, height, etc.) and on terrain that blocks airwave travel (e.g. mountains). The ITM has also been used in a number of other studies (Olken (2009); Enikolopov et al. (2011); Durante et al. (2015))

[28]For the US, alternative signal data has been collected in Gentzkow (2006). He uses modern media markets as proxy for historic TV signal. Modern media markets are linked to the first TV station that falls

the same change in market reach in Dollar terms. I collect station specific price data from "rate cards" and use them to compute hourly revenue figures for each market.

Theatre Data To measure audience size in the pre-TV period I collect archival records on the seating capacity of live performance venues across the US. This information comes from a historic companion book for the entertainment profession, the 1921 "Julius Cahn-Gus Hill theatrical guide," which aims to provide "complete coverage of performance venues in US cities, towns and villages."[29] The data covers seating capacity and ticket prices of over 3,000 performance venues that, taken together, cover more than 80% of US local labor markets.[30] On average, a performance venue has 872 seats, but capacity varies between a few hundred seats to several thousand. The most iconic performance venue at the time was the New York Hippodrome, which was hailed as the "world's largest theater" at a capacity of over 5,000 seats. The largest venue in a labor market had, on average, 1,165 seats. I use these data to quantify the shock in market reach from the launch of television. The measure of audience size combines the live audience data with the audience of TV shows, while the Dollar value of a show is based on ticket prices times local audience for live shows and advertisement rates for TV shows.[31]

County Fairs Additionally, I collect information on expenditure at local entertainment outlets. My data spans ticket sales and revenues for over 4,000 fairs spanning 11 years (1946-1957) and the majority of US labor markets. The data come from the "Cavalcade of Fairs," which contains detailed records on county fairs and is published annually as a supplement to Billboard magazine. Fairs provide a range of amusement activities, usually including a carnival with rides, food stalls, activities and a grandstand show with performances by local dance squads and music groups, sport competitions and similar highlights. The records report spending in three categories: fair ticket receipts, show entrance receipts (e.g. grandstand) and carnival receipts[32] (e.g. fair rides, merchandise and food). I aggregate the spending categories at two levels: the county-year level and at the more aggregated local labor market-year level, which allows me to analyze

within that radius and the launch date of that station is used as the date where TV became available in the area of todays media market.

[29] According to the author "Information has been sought from every source obtainable - even from the Mayors of each of the cities." Undoubtedly the coverage will be better for larger venues and small or pop-up venues will be missed. Since we focus on star venues this omission may be a lesser concern.

[30] Spot checking confirms that the data accurately cover physical performance establishments. The data does not cover mobile performance venues, such as circus tents. Since the analysis is concerned with the largest local performance venues, this omission is likely not a major problem.

[31] Details on Revenue data are in Appendix 8.D.2. For TV shows, prices are imputed based on an demand elasticity estimated in a subset of 451 markets where data is available.

[32] Carnival receipts are unavailable in 1953 and 1955.

demand across these regions separately for leisure activities that are differentially close substitutes for television.

3.3 Labor Market Data

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

The core of the analysis focuses on occupations that appear on television. Three-digit occupation information identifies five relevant entertainment occupations: actors, athletes, dancers, musicians and entertainers not elsewhere classified. The last group is relevant because it includes most circus and vaudeville actors, one of the most important forms of entertainment at the time.[33]

In many settings, the reclassification of occupations over time poses a problem. Entertainment occupations, however, are well established and there is little change to their definitions throughout the sample period. There are a few exceptions; most relevant for the above groups is that the athlete category is discontinued in 1970. To account for such time shifts in the occupation definition, the regressions will control for occupation-specific year effects. In defining labor markets, I follow Autor and Dorn (2013) and define local labor markets as urban centers together with their respective commuter belts, so called "commuting zones". I extend the data produced by Autor and Dorn (2013) backwards and produce consistent labor markets for the Census data going back to 1930.[34] The final data covers 722 consistent local labor markets spanning the mainland US over time. On average, a commuting zone has about 400,000 inhabitants and approximately 500 workers in entertainment occupations.

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

The analysis studies labor demand shocks at the local labor market level and I compute outcomes at this level for each occupation and year. A first set of outcome computes entertainers' position in the US wage distribution. This follows Chetty et al.

[33]The original string occupation title is available in the 1940 Census and confirms that the category includes acrobats, clowns, animal trainers, etc.

[34]See Appendix 8.D.2 for details on the variable construction

(2014) and measures wage inequality by ranking entertainers relative to a benchmark group. A key advantage of the rank position metric is that it is scale independent and thus makes it easier to compare changes in pay inequality over time. A further advantage of this measure is that it allows to side-steps top coding issues in the wage variable. The share of workers with a wage above a threshold, say the 99th percentile, can be computed even with a top code in place as long as the top code bites above the 99th percentile. In my data this is the case and I can thus compute the measure throughout the sample period. To compute a set of variables that cover the entire wage distribution, I divide the number of entertainers in a given range of the US distribution by the total number of entertainers in the market. I prevent that fluctuations in the denominator bias my results by fixing the denominator above the treatment level. Take for example the share of entertainers whose wage falls in the top 1% of the US wage distribution ($D^{US1\%} = 1$):

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

The share of top paid entertainers is thus computed by dividing the number of top-earning entertainers in market m at time t by the number of entertainers in a standard labor market. The denominator does not vary at the level of my treatment and amounts to a normalization.[35] This guarantees that $p^{99}_{m,t}$ captures effects on the top of the distribution rather than fluctuations in the number of entertainers. Results without the normalization, as presented in the Appendix 10, are in line with the baseline effects. I also compute additional outcome measures: top-paid entertainers as a share of the local population and top income shares of entertainers.[36]

Finally, I build a short panel of the career of TV stars. This uses the de-anonymized records of the 1940 Census and matches local TV stars of the 1950s to their pre-TV careers. Information on the stars of the 1950s come from the "Radio Annual, Television Yearbook " which publishes the "Who is Who of TV". For 60 out of 89 cases a unique Census record can be identified; clearly, this is only a subset of all entertainers.[37] The advantage of these linked records is that we obtain information about entertainers' pre-TV careers.

[35]To simplify interpretation of the treatment effects, I normalize by the average number of entertainers in treated labor markets (instead of averages across all markets). The regression coefficient therefore has a natural interpretation as a percentage point change in the treated market.

[36]If the top tail of the distribution is not observed, I use Pareto interpolation to estimate top income shares. The procedure follows a large literature that uses Pareto interpolation to estimate top income shares (Kuznets and Jenks (1953); Feenberg and Poterba (1993); Piketty and Saez (2003); Blanchet et al. (2017)) and is described in Appendix 8.D.2.

[37]Manually searching vitas generates information on place of birth, birth date and parents. Combined with the information on names and places of residence, I identify entertainers' 1940 Census records.

4 Empirical Results

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

An ideal test of superstar effects would randomly assign production technologies across labor markets that allow for varying degrees of market reach of workers. To get close to this ideal, I exploit the staggered introduction of television that varies entertainer market reach across local labor markets and test the effects in a difference in differences regression. This regression compares local labor markets $m = 1, ...M$ over time t:

$$Y_{mot} = \alpha_m + \delta_{ot} + \gamma X_{mt} + \beta TV_{mt} + \epsilon_{mot} \tag{12}$$

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

Early TV stations mainly filmed locally and hired local entertainers and thus affected demand for entertainer in local labor markets. Non-local shows were a poor substitute for local productions for two reasons. First, the infrastructure to air shows simultaneously across stations was lacking. Sterne (1999) gives a detailed account of pioneering efforts to build a national TV network and the major technical obstacles. While in principle storing and transporting shows was feasible, the technology was costly and turned out to be unpopular because it led to poor image quality.[38] Second, regulation restricted studio locations by specifying that "the main studio be located in

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

the principal community served" (FCC annual report 195). The effect of the local launch of a TV station is captured by the coefficient β.

The main source of variation in TV_{mt} comes from the staggered deployment of TV stations. The first commercial television stations were launched on July 1st, 1941, but many regions did not get stations until years later.[39] The roll-out was based on an FCC license process that specified where TV was to be launched and issued licenses accordingly. The initial roll-out was hampered by production restrictions on TV-related equipment during World War II. From 1945 onward, television spread rapidly and by 1949, 124 stations were active. Figure 3 illustrates which local labor markets had been treated. Over subsequent years launch dates across areas differed by as many as 15 years because multiple delays interrupted the roll-out.

A second source of variation in TV_{mt} comes from the eventual decline of local TV filming, impelled by the invention of the videotape. The invention of the Ampex videotape made it possible to store and transport TV productions cheaply, which transformed the TV production industry.[40] Shows from outside the local labor market became a close substitute for local live shows. This led to the demise of local TV production and the concentration of TV production in two hubs, Los Angeles and New York.[41] The videotape proved an instant hit with TV stations. For example, when the product was presented at the National Convention of Broadcasters in 1956, over 70 videotape recorders were ordered immediately by TV stations across the country. The same year, CBS started to use the technology, and the other networks followed suit the next year. Local TV stations' effect on entertainers would subsequently fade, while production hubs, by contrast, started to serve vastly bigger audiences. To capture the effects on the hubs I interact a dummy for the time period of national production with a time-invariant and pre-determined proxy for local production cost.[42] This variable will pick up incentives to move filming to a given place, while the measure avoids the potential endogenous control problem. At the same time the period of national production allows me to test if the effects of local antennas disappear when their importance for TV filming declines.

For identification we can thus exploit both the introduction and removal of the treatment. The key identification assumption is that TV launch dates are unrelated to local trends. By observing the treatment removal, I have access to a powerful parallel trend check. Treatment effects ought to disappear after the removal of the treatment.

[39] Experimental broadcasting existed since the 1920s and had familiarized the population with the new technology. Prior to the launch of commercial television the private ownership of TV sets was however minimal. In four cities experimental broadcasters where later turned into commercial television channels.

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

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

[42] See Section 3 for the construction of the proxy.

Moreover, I can sharpen the identification by exploiting an exogenous interruption of the planned roll-out process.

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

4.1 Effect on Top Earners

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

The magnitude of these effects is easier to interpret if compared to wage changes in the aggregate US distribution. The rest of the paper therefore focuses on the position of entertainers in the US wage distribution (see Section 3 for the variable definitions). A first test looks at $p^{99}(\xi_t)$, the share of entertainers among the top 1% of US wage earners. In labor markets that subsequently received TV stations about 4% of entertainers were paid at this top level in 1939. With TV the share of top earners increases by roughly 4 percentage points and hence the share of local top earning entertainers roughly doubles (see panel B in Table 1). Recall that the denominator of the outcome variable does not vary at the treatment level and the effects therefore capture changes at the top of the distribution, rather than changes in the denominator. This effect on the top of the entertainer pay distribution can also be seen by focusing on the per capita share of top paid entertainers. This measure is again not affected by occupational choice in the local labor market. Results for the per capita measure are reported in panel C of Table 1 and are comparable to the previous results.

A related question is whether TV broadened the availability of high quality entertainment. The quality measure for entertainment is difficult to come by as quality is subjective. I avoid this debate and look at willingness to pay for different entertainers and thus use prices as a reasonable proxy for the value consumers attach to different entertainers. I build a short panel for a subset of local TV entertainers that allows me to test whether the most valued entertainers benefit from TV.[43] The panel reveals that TV did not lead to substantial leapfrogging in the wage distribution, the vast majority of TV stars were in the top tail of the wage distribution even before TV (Figure 7). Television therefore predominately promoted the market reach of the most popular entertainers.[44]

[43] See Section 3 for the details on the data

[44] An alternative interpretation is that TV promotes the same kind of talent that was required to be successful in the pre-TV era.

4.1.1 Probing the Identification Assumption – TV filming

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

An initial check is to control directly for time-varying changes in local labor markets in the regression. I run two specifications, one controlling for time varying local characteristics and one that allows for local labor market specific trends (column 2 and 3 in Table 1). The second approach, is a very demanding specification as it adds more than 700 additional location specific trends and standard errors increase accordingly. Both specifications find effects very similar to the baseline, indicating that differential local trends are not driving the findings.

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

The principal reason for the sudden shut-down was an error in the assignment model. The FCC used a signal propagation model to calculate which broadcast frequencies and catchment areas were safe to use, avoiding interference between stations. An error in the model resulted in interference occurring among licensed stations. The FCC ordered a review of the model to avoid such interference problems from becoming worse and put all ongoing license procedures on hold. This interruption has previously been noted by scholars that study the social consequences of TV watching. Besides varying TV filming locations, this interruption also left some regions without access to TV signal, which studies have used to analyze the social consequences of TV watching (Gentzkow (2006); Gentzkow and Shapiro (2008)). In this study I focus on different variation, the roll-out of TV studies, and use newly collected data for a novel identification strategy.

I digitize data that allow me to observe where licenses were imminent but did not proceed as planned because of the FCC review. A simple identification approach would compare TV launches before and after the freeze. A threat to this type of identification strategy arises from selection into treatment if post-freeze stations are not the same as the ones exogenously held up by the freeze. This concern is particularly relevant here, since revising the roll-out priority was the dedicated goal of the freeze. The newly collected administrative records allow me to avoid such selection bias and reveal directly which locations were held up by the freeze. Figure 3 shows the affected local labor markets. During the freeze period, the FCC undertook extensive field studies and expert hearings to improve the scientific standard of their signal model.[45] As a result, licensing did not resume until 1952 and the onset of TV was delayed by nearly four years in many markets.[46]

We can use such "stations that did not happen" to test whether the introduction of television coincided with spurious location specific shocks to top pay. I compare local labor markets that narrowly miss out to untreated locations and test for spurious trends by comparing the two types of untreated locations in a dynamic difference in differences regression:

$$Y_{mot} = \alpha_m + \delta_{ot} + \gamma X_{mt} + \sum_t \beta_t TV_{mt}^{blocked} + \epsilon_{mot} \tag{13}$$

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

Figure 9a illustrates the time path of the coefficients on "stations that did not happen". The point estimate is a precise zero. Local labor markets that narrowly missed out on a TV station experienced no top income growth for entertainers. Areas affected by the freeze and untreated labor markets follow the same time path. This test is arguably more convincing than a pre-trend test or a test based on placebo occupations, as we can test for local shocks in the same occupations and year. For peace of mind, such related checks of pre-trends and placebo occupations are reported in Appendices 8.D.1 and 8.D.1.

[45]The revision of the model was further delayed by the onset of color transmission standards which required additional field tests.

[46]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.

Common Trend Test Common trend tests add further credibility to the identification assumption. In my setting I observe treatment and control groups in an untreated state both before and after the removal of local TV show filming. I use both pre- and post-treatment periods to test the common trends assumption. If common-trends hold, the treatment effect arises when local TV productions are introduced and disappears when they are removed.[47] I test this in a dynamic difference in differences regression, similar to equation 13 but using local TV stations ($TV_{m,t}$) as regressor instead of blocked stations ($TV_{m,t}^{blocked}$). The time-path of β_t reveals how differences between treatment and control group change over time and are plotted in Figure 9b. A difference in treatment and control areas appears during local TV filming and disappears after the end of local TV. By 1969 the differences between treatment and control group returned to the pre-treatment level, which suggests that the common trend assumption holds.[48]

4.2 Concentration of Consumer Demand

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

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

First, I estimate how TV affected market reach of star entertainers. I run the baseline specification, regressing log of audience size of the biggest local shows on the introduction of local television production. The results reveal that the most widely watched entertainers experienced a dramatic growth in audience size (see panel A. of Table 2). The launch of a television station increased the audience of the largest shows by about 150 log points, which converted to a growth rate implies a growth of over

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

[48] A conventional pre-trend check is reported in Appendix 8.D.1.

300%, or a fourfold increase in market size. Additionally, I quantify the change in market reach in dollar terms and thus am able to estimate the change in marginal revenue product that went hand in hand with the growth in audience size. In dollar terms market reach of stars roughly tripled (see panel B. of Table 2). The launch of a TV station thus dramatically increased the market value of top talent.[49]

Second, I estimate how TV affected demand for ordinary local live entertainment. To study such effects I want to know if customers of local entertainment were able to watch TV and hence whether local entertainers have to compete with TV entertainment. Notice that the variation in TV signal differs from the previous variation in TV filming because signal often travels beyond the local labor market where filming takes place. I study how demand for local entertainment changes as TV channels become available locally and find that demand for ordinary local entertainment drops substantially when customers can watch TV. When a station is launched local county fair ticket revenues drop by about 5% (column 1 of Table 3). The decline in revenue reflects a similar decline in fair attendance (column 2). People staying away from county fairs can therefore account for fairs' declining ticket revenues. These results are, however, noisy as they hide substantial heterogeneity.

Demand effects differ markedly for different types of entertainment. More specifically, demand for entertainment that is similar to TV falls significantly, while demand for entertainment that is very different from TV holds up. I collect data on receipts for two extremes of substitutability: grandstand shows and traditional carnivals. Grandstand shows included vaudeville acts, thrill shows, dance groups and beauty pageants and were similar to many TV shows at the time. Traditional carnival activity on the other hand, including candy sales and amusement rides, engages more than the visual senses and is inherently less substitutable by television. As columns 3 and 4 of Table 3 report, grandstand show receipts see a large decline, while carnival receipts are unaffected. This shows that even within the entertainment spending category, close substitutes to TV are most affected by the availability of TV signal, providing confidence that we are in fact picking up the effect of TV.

Next, I analyze the same data at a finer local level, which arguable allows me to measure TV exposure with greater accuracy. So far the unit of observation was the local labor market, in line with the study of labor market effects. A finer regional analysis of spending is feasible because the precise location of fairs is available in my data. Running the regressions at the county level increases the power of the set-up. However, this faces the drawback that we must assume that fairgoers come from the county where the fair is held. Results at the county level are in line with the previous findings (panel B of Table 3). TV again decreases demand for fair entertainment. The point estimates are smaller than before, which plausibly reflects the fact that we are mismeasuring TV

[49]Note that audience grows faster than revenues, reflecting a drop in the unit price for entertainment. Star entertainment became more widely available and simultaneously more affordable.

signal exposure for a sizable fraction of fairgoers who live outside the county where the fair is located. Overall, these result confirm that TV signal reduces demand for local live entertainment and increased demand for star entertainers. The marginal revenue productivity of entertainers shifts substantially in favor of stars.[50]

4.3 Effect on Non-Stars

4.3.1 Cannibalization of Demand for Non-Stars

Hollowing out A prediction of the superstar model is that technology causes demand to shift from a profession's mediocre workers towards its stars. We already saw this effect in the demand data where the revenue product of county fair entertainers declines because of TV. Now I focus on the labor market effects for a broader group of non-star entertainers. We should observe declining pay for non-stars (see Proposition 2.2). Consider, entertainers who are below the top 90th percentile of the US wage distribution but still in the upper quartile. These are entertainers who receive above-average pay but are far from the top of the entertainer distribution. The launch of TV production has a significantly negative effect on this group, as the number of jobs that pay in this range declines by around 50%. The results look similar between the median and the 75th percentile (results are reported in Figure 10). Television therefore leads to a substantial decline in well-paid jobs below the star level. This suggests that it is substantially worse to be a mediocre entertainer during the TV era.

The corollary to disappearing mid-paid jobs is the growing low-pay sector. Analyzing the share of entertainers paid below the median, we observe a modest rise in the share of entertainers with wages at the very bottom of the distribution, with little change in the second quartile. Television thus reduces the payoff of non-star talent and creates a growing low-pay sector.

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

To test for employment effects among entertainers, I compare local labor markets with differential access to TV signal. The Census records information on employment information for longer than wages, I therefore expand the sample period backward by a decade to 1930. The full sample therefore covers 1930-1970 and results are reported alongside results for the baseline period. For this extended period data on TV signal is not available at a channel level, instead I use a dummy for access to TV

[50]Notice that the productivity results rule out that the observed wage effects are purely driven by a shift in bargaining power.

signal as regressor, moreover median income is unavailable, hence specifications with demographic controls include only the remaining five demographic controlls.[51]

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

4.3.2 Probing the Identification Assumption – TV signal

Since the previous results use variation in TV signal rather than variation in TV filming, it is salient to probe the identifying assumption again. As before, I first control for proxies of local demand shifts with time varying local characteristics and local trends. Columns 2 and 3 of Table 4, respectively, show that the coefficient of interest remains unchanged, implying that we are identifying fairly sharp changes around the time of treatment. Next I test for common trends in treatment and control group directly.[52] I implement a pre-trend test by introducing a lead to the treatment variable in the difference in differences regression. This captures differential changes in treatment and control areas, right before the onset of the actual treatment. If we have parallel trends the effect of the lead should be zero. Indeed the coefficient on the lead is insignificant and qualitatively small and indicates that the treatment and control groups are on similar trends in the run up to the treatment (column 4).

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

4.3.3 Fractal Inequality

A final implication of a superstar model is that technical change widens the wage difference between stars and their slightly less talented peers, while skill-biased demand shocks move wages proportionally (see Proposition 2.1). A non-parametric test of this prediction repeats the baseline difference in differences regression, focusing on lower percentiles. Take, for example, entertainers who are below the top 1% but still among the top 5% of the US wage distribution. I find that television has a more modest 12%

[51] Median income is missing in 1930 and TV channel data in 1970.

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

effect on this group. This effect just below the top of the wage distribution is only one tenth the size of the effect at the very top. Television therefore disproportionally benefits the superstars. To confirm this pattern we can look at the next lower wage bin. The effect of television on entertainers in the top 10% but below the top 95th percentile is insignificant with a negative point estimate, again confirming that television's effect fades quickly as we move away from the top stars in the market.

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

Next I focus on pay dispersion within entertainment. This is closely related to the previous results but focuses on an inequality measures widely used in the literature, top income shares. Proposition 2.4 suggests that the growth in these shares should escalate toward the top of the distribution. Top income shares have been used extensively to document growing inequality at the very top (Piketty and Saez (2003); Piketty (2014)). To compute top income shares for different parts of the top tail we need rich data about the distribution. I therefore focus on the larger 350 markets with sufficient observations and follow the literature in computing top income shares using Pareto interpolation as described in Appendix 8.D.2.

I implement a test for growing pay dispersion in the top tail by studying the effect of TV on top income shares. Prior to TV, the fraction of income going to the 1% highest earners in a local labor market was, on average, 3.8%.[53] TV filming increased the top income share by 3.7 percentage points, which means the top 1% income share almost doubled. Most of the gains accrued to the very highest earners in the top 1%. The top 0.1% of entertainers saw their income share rise by 2.4%. This group is only one tenth of the top 1% but accounts for over half of the rise for the top 1% income share. The share of income going to the top 1% grows by 100% and the equivalent share for the top 0.1% grows by 300%, but the top 10% share grows by only 30%. This shows that the wage growth is most pronounced at the very top end of the distribution. A formal test of equal growth rates of wages in the top tail is strongly rejected. Such increasing effects in the top tail are at odds with models of skill-biased demand where wages grow proportionally across all percentiles within a skill group. The findings however aline with the superstar model where wage growth is skewed towards the top of the distribution.

In the appendix, I confirm these results with a set of quantile regressions (see Appendix 8.D.1). These estimates look at changes in the wage quantiles within entertainment and show the same pattern. The effect of technology declines remarkably quickly in the top tail. TV appearances help a small group of top stars, has moderate

[53]The equivalent number for the US economy as a whole is about 10%. It is however unsurprising that within a given region and industry income is less dispersed.

effects on backup stars and has no discernible benefit for other top earners.

4.4 Links Between Markets

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

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

A second spillover comes from mobility of workers. First, note that the biggest labor markets tend to receive TV first and, as a result, the ranking of places by audience reach remains largely unchanged. The TV roll-out thus generates limited new incentive to relocate. To back this claim up empirically, I use data on mobility reported in the US Census.[55] I find that TV has small effects on mobility rates. The coefficient on television in a regression on the probability that entertainers move are insignificant. In fact, the point estimates are negative and the results rule out that mobility increased by more than 2% (columns 1 - 3 of Table 6). Migration therefore cannot explain a major share of the baseline findings.[56]

An alternative approach is to focus on labor markets that are further apart. Moving is arguably easiest between neighboring markets, so we should see most of the migration take place across that margin. By excluding control areas that neighbor treated labor markets, we will clarify the importance of such spillover effects.[57] Results that exclude neighboring areas are close to the baseline (see panel B of Table 6), suggesting again that relocation between neighboring markets does not explain the findings.

[54]TV shows, by contrast, can be watched beyond the local labor market. This occurs if TV signal travels beyond the local labor market. I have data on signal reach and can thus account for links between markets.

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

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

[57]This excludes 2,990 labor markets, or 1/4 of the sample.

5 Magnitude of Superstar Effects

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

$$ln(w^{99}_{m,t}) = \alpha_0 + \alpha_1 ln(s^{99}_{m,t}) + \epsilon^{99}_{m,t} \tag{14}$$

where $w^{99}_{m,t}$ is the 99th percentile of the entertainer wage distribution in market m and year t and $s^{99}_{m,t}$ the size of the market that such entertainers can reach. In a first step, I estimate this relation using a naive OLS estimator on a single cross-section. This uses the variation in the size of the biggest local theatre size in 1939 as regressor. A large literature has estimated similar regressions by correlating firm size and pay. In line with those results my cross-sectional OLS estimate of α_1 is highly significant with a point estimate of 0.23 (see panel A. of Table 7) Moving from a local labor market with a small theatre to a market twice the size, increases pay for a top earner by 23%. The effect may of course reflect differences in local labor markets, rather than the effect of market reach. Indeed, after controlling for local characteristics, the effect disappears almost entirely (column 2 of the same Table).[58]

To estimate the causal effect of market reach, I turn to the exogenous shock in market reach from TV. I use an IV approach that instruments market size with the roll-out of TV. The first stage regression, the effect of TV on audience size, is large and highly significant (recall Table 2) and the associated first stage F-statistic is well above conventional cutoffs with a value of 20. The IV estimator of the elasticity α_1 is also highly significant with a point estimate of 0.17. This implies that wages at the 99th percentile grow 17% when market size doubles. While this wage effect is sizable, the effect is 30% lower than the cross-sectional OLS estimate above. This suggests that the causal effect of market reach is smaller than the correlation of market size and top pay suggests.

The literature on product market concentration measures changes in market size in terms of revenues, rather than measuring concentration of customers. To link my results to this literature, I compute show revenues. The dollar value of an entertainment

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

show is calculated as the product of the audience measure and ticket prices for theaters or advertisement prices for TV shows. Notice that aggregate chances in prices will not affect the regressor, since the log specification absorbs such price changes in the time fixed effects. Results are affected by changing price dispersion between shows.[59] Using revenue as regressor makes it easier to compare results across settings but has the drawback that it uses an outcome, prices for talent, as part of the regressor of interest. Non-withstanding this limitation, I estimate the elasticity of top pay to market value and find again a highly significant effect with a point estimate of 0.22. One dollar greater concentration in the product market therefore leads to 22 cents higher pay for star workers.[60]

6 Imperfect Competition and Superstar Effects

For policy makers it is key to understand how superstar effects interact with imperfect competition. The benchmark superstar model is perfectly competitive and growing top incomes are the result of changing demand for talent. The models' predictions change sharply with imperfect competition, as employers with market power will not pass-on the surplus from greater scalability of production. Monopsony power will thus reduce the predicted top income growth. To test this prediction I use variation from the licensing process that limits employer entry into labor markets and thus exogenously generates monopsony power. To implement such a test empirically I allow for differential effects of TV in markets with a single TV station and markets with multiple TV stations. Since this dummifies the previously continuous treatment variable, I first report the baseline regression with a dummy treatment variable (column 1 of Table 8). I then introduce the additional dummy for labor markets with multiple TV stations (column 2). The effect on labor markets with multiple TV stations is therefore the sum of the two dummies. The differences between monopsonistic and competitive labor are striking. Markets with a monopsony employer see almost no top income growth, while gains are large when there is more than one employer. Using information on the freeze, I identify labor markets that would have had competition from a second station if the freeze had not blocked the competitor. The results confirm the baseline findings; locations that would have experienced employer competition but miss out see next to no top-wage growth (column 3). Only when employers face competition, does greater market scale translate into rising wages.

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

[60] Note that this estimate is bigger than the elasticity with respect to audience size. A fact that arises because the launch of TV reduced the per-head cost of top entertainment. The reduced form of both elasticity IV estimators is the same, a smaller first-stage therefore increases the IV estimate.

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

We can also study the reverse, the possibility that a TV station's exogenous entry breaks up previously non-competitive structures. To investigate this possibility, I allow the effect of TV to differ across labor markets with different numbers of pre-TV employers. The result, as reported in column 4 of Table 8, is a fairly precise zero. There is no differential effect across this dimension. This suggests that the pre-TV labor market of entertainers was reasonably competitive or that the differences that arose from imperfect competition are negligible relative to the effect of greater scalability. Another possibility is that the number of employers in the pre-period is a poor measure for competition. To address this, I use an alternative proxy of labor market competitiveness, population density. Here again, I find no effect (column 5), which suggests that pre-TV labor market competitiveness does not greatly influence superstar effects. Given the magnitude of the top income growth, this is unsurprising. Appendix 8.D.1 further explores how policy induced variation in the structure of local labor market affects superstar effects and specifically tests the impact of tax wedges and education levels.

7 Conclusion

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

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

The increase in production scalability has profound effects on inequality at both the top and bottom of the distribution. These results are at odds with conventional models of technical change but align with the predictions of a superstar model. To assess the magnitude of superstar effects, this paper provides top income elasticities with respect to market size. The estimates imply that one extra dollar in product market concentration leads to 22 cent higher pay at the 99th percentile. Similarly, the share of income going to the top 1% increases by 30% if market size doubles.

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

8 Appendix I: Superstar Earners and Market Size

8.A Tables

Table 1: Effect of TV on Entertainer Top Earners

	(1)	(2)	(3)
	Panel A: Ln(99th Percentile of Entertainer Wages)		
Local TV station	0.171	0.157	0.149
	(0.029)	(0.066)	(0.093)
Effect/Baseline	18.6%	16.9%	16.1%
Cluster	702	702	702
	Panel B: Entertainer among Top 1% of US Earners (% of Entertainers)		
Local TV Station	4.14	4.31	5.93
	(1.26)	(1.27)	(2.21)
Effect/Baseline	92%	96%	132%
Cluster	722	722	722
	Panel C: Entertainer among Top 1% of US Earners (Per Capita)		
Local TV Station	0.40	0.40	0.31
	(0.10)	(0.10)	(0.10)
Effect/Baseline	133%	133%	103%
Cluster	722	722	722
Time & Labor Market FE	Yes	Yes	Yes
Demographics	–	Yes	–
Local labor market trends	–	–	Yes

[Note] Dependent variables are, Panel A: The entertainer wage at the 99th percentile, Panel B: share top-paid entertainers as described in the text, Panel C: top-paid entertainer divided by the local population in 10,000s. Specifications: Each cell reports the regression coefficient of a separate difference in differences regression. The treatment is the number of TV stations in the local area. All regressions control for commuting zone fixed effects, time fixed effects and local production cost of filming in years after the invention of the videotape. Demographic controls are median age, median income, % female, % black, population density and trends for urban areas; specifications with local labor market trends include a separate linear trend for all local labor markets. Entertainers are Actors, Athletes, Dancers, Entertainers Not Elsewhere Classified, Musicians. Panel A implements the quantile DiD using the estimator developed by Chetverikov et al. (2016); cells where the 99th percentile cannot be computed are dropped. The unit of observation in this approach is a commuting zone - year cell, while Panel B and C are run at the more disaggregated commuting zone - occupation - year level to additionally control for year-occupation fixed effects. Panel A uses 2,264 CZ-year observations and Panel B and C 13,718 occupation-CZ-year observations, demographic data is missing for one CZ in 1940. The "outcome growth" reports treatment effects as percentage growth over the baseline value of the outcome variable. Observations are weighted by local labor market population. Standard errors are reported in brackets and are clustered at the local labor market level. Sources: US Census 1940-1970.

Table 2: Effect of TV on Market Reach of Local Stars

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

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

Table 3: Effect of TV on Log Spending at Local County Fairs

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

[Note] Dependent variables are summed across county fairs in location m in year t at annual frequency from 1946 to 1957. All variables use the the inverse hyperbolic sine transformation to approximate the log function, while preserving 0s and monetary variables are in 1945 US Dollars. In Panel A the unit of observation m is a local labor market and in Panel B a county. Treatment is the number of TV stations that can be watched in the commuting zone. Data on carnival receipts (col 4) are unavailable for 1953 and 1955. Panel A uses 8,664 local labor market observations (7,220 in column 4), while Panel B uses 37,332 county observations (in col 4 31,110). Standard errors, reported in brackets, are clustered at the local labor market level in Panel A and at the county level in Panel B. Source: Billboard Cavalcade of Fairs 1946-1957 and Fenton and Koenig (2018).

Table 4: Effect of TV on Entertainer Employment

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

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

Table 5: Effect of TV on Top Income Shares in Entertainment

	(1)	(2)	(3)
		Share of Income	
	Top 0.1%	Top 1%	Top 10%
Local TV station	2.37	3.71	6.08
	(1.27)	(1.69)	(2.12)
Time & Labor Market FE	Yes	Yes	Yes
Growth in outcome	239%	96%	33%
P-value: same growth as top 1% share	0.0043	—	0.0000

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

Table 6: Effect of TV on Mobility Between Labor Markets

	(1)	(2)	(3)
	Panel A: Share Entertainers who Migrated		
Local TV station	-0.014	-0.017	-0.010
	(0.015)	(0.015)	(0.020)
	Panel B: Share Entertainer among Top 1% of US Earners (excl. neighbor)		
Local TV station	4.30	4.46	6.16
	(1.31)	(1.30)	(2.27)
Time-Occupation & Labor Market FE	Yes	Yes	Yes
Demographics	–	Yes	–
Local labor market trends	–	–	Yes

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

Table 7: Elasticity of Entertainer Top Pay to Market Reach

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

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

Table 8: Effect Heterogeneity by Market Structure

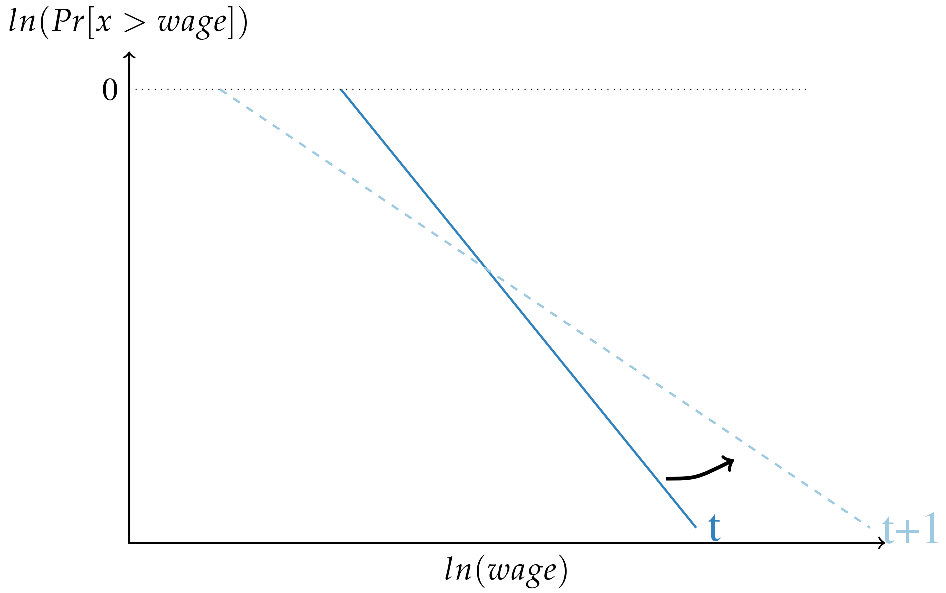
	(1)	(2)	(3)	(4)	(5)
	Share Entertainer in US Top 1%				
Local TV station (dummy)	5.90	0.753	-0.57		
	(3.06)	(1.91)	(0.36)		
Multiple local TV station (dummy)		9.07	10.37		
		(4.99)	(4.70)		
Frozen competitor			1.43		
			(2.10)		
Local TV station				4.25	4.25
				(2.25)	(2.25)
Local TV station × theatre count				-0.00	
				(0.02)	
Local TV station × population density					-0.00
					(0.02)

[Note] Sources and specification as in baseline. Theatre count are the number of employers listed in the Cahn-Gus Hills theatrical guide.

8.B Figures

Figure 1: Effect of Technical Change on Wage Distribution – Superstar Model

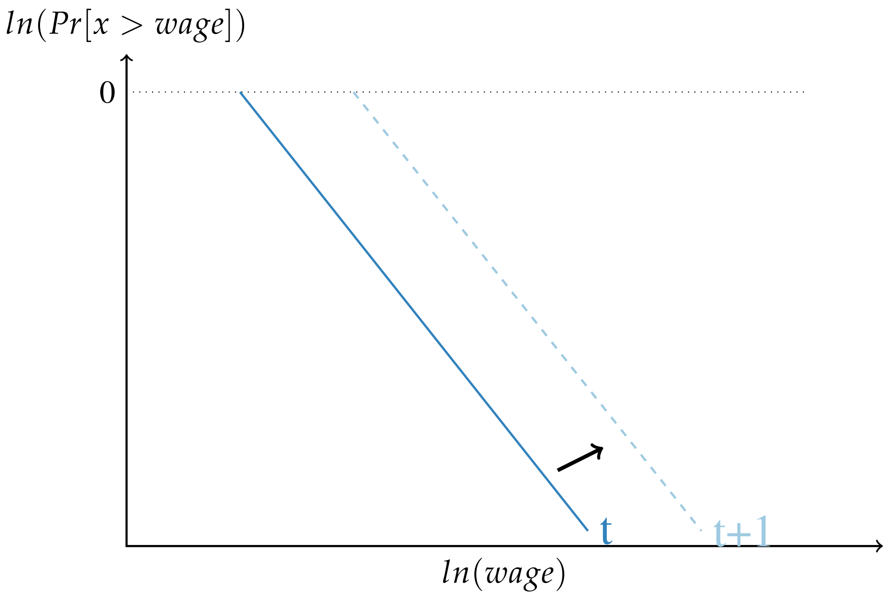
(a) *Wage Distribution*



(b) *Employment Growth at Different Wage Levels*

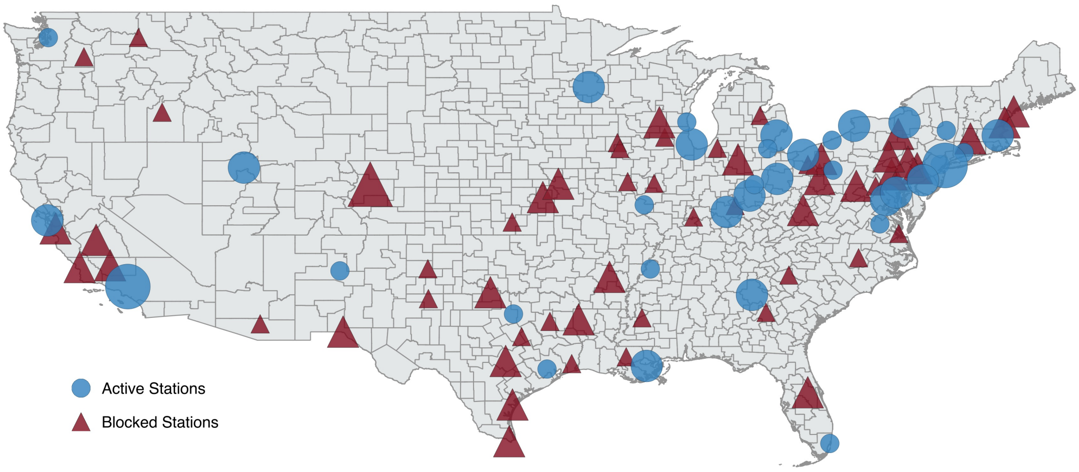
[Notes] Panel A shows the wage distribution given by equation 8 and illustrates wage changes when production becomes scalable by $s > 1$ between t and $t + 1$. Panel B illustrates the same change in terms of employment growth across wage bins. The figure shows equation 10 for parameterization $g_w = 0.2x^{(1.3)} - 1$.

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



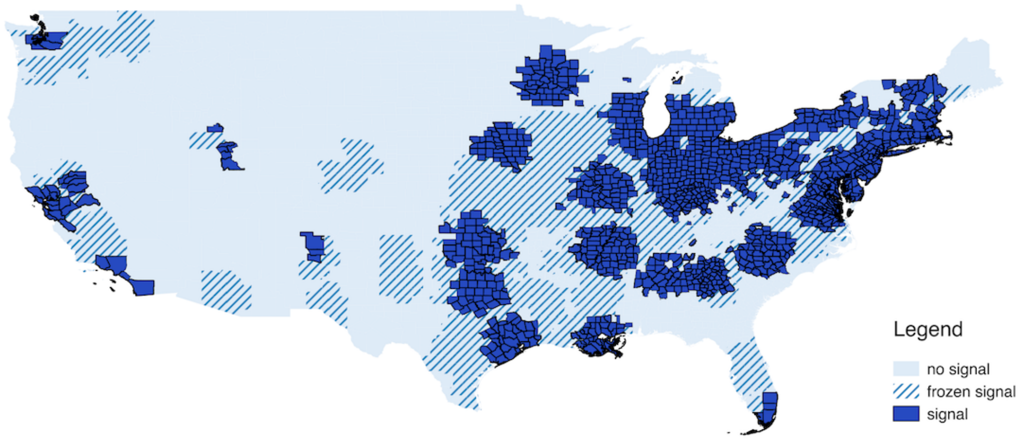
[Note] The figure shows an increase in the skill premium for a group of workers with heterogenous but perfectly substitutable talent.

Figure 3: Intensity of TV Filming in 1949



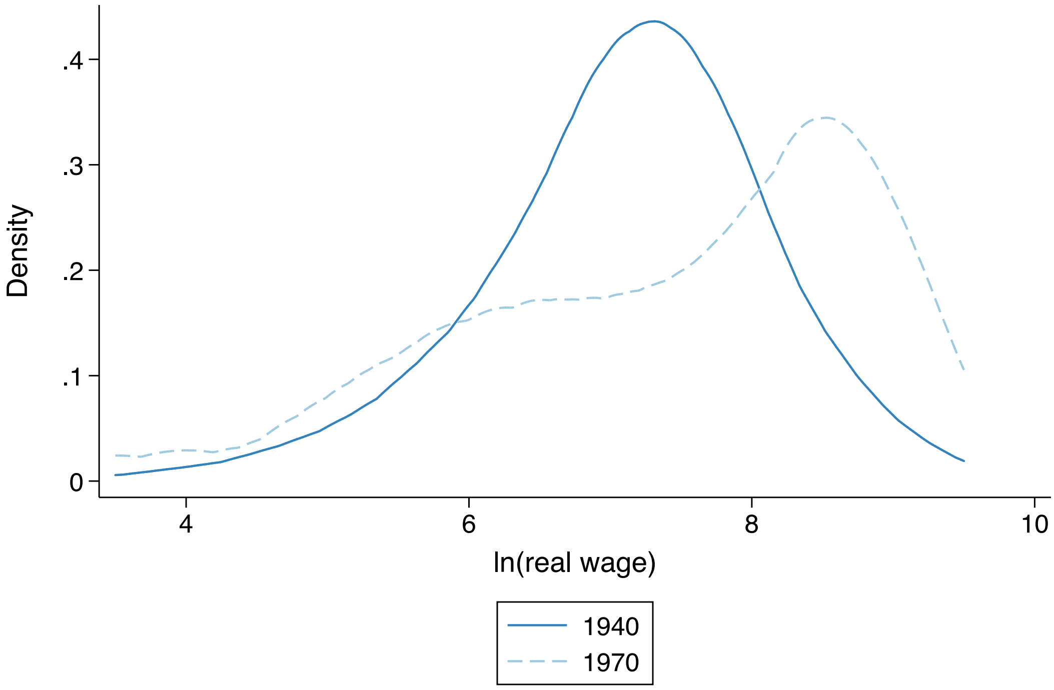
[Notes] Symbols show the location of TV filming and the size of a symbol indicates the number of TV stations per local labor market. Active stations are blue circles, frozen stations red triangles. Source: FCC reports.

Figure 4: TV Signal of Licensed and Frozen Stations in 1949



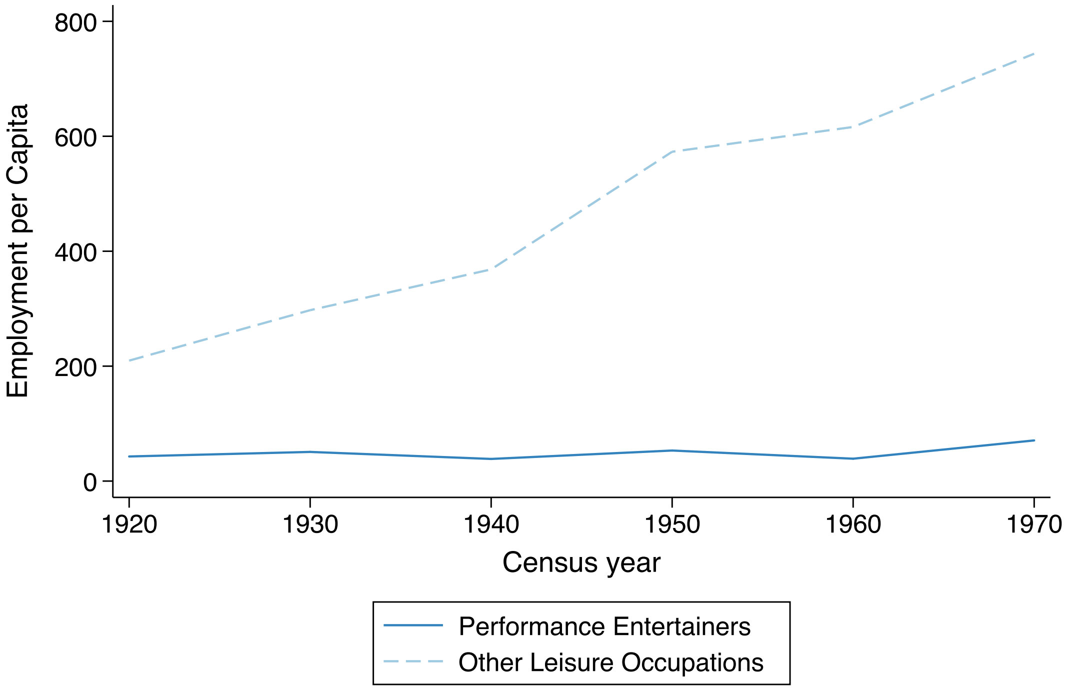
[Note] Areas in dark blue can watch TV, while shaded areas would have had TV signal from blocked TV stations. Signal coverage is calculated using an Irregular Terrain Model (ITM). Technical station data from FCC files, as reported in TV Digest and Television yearbooks, are fed into the model . Signal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

Figure 5: Entertainer Wage Distribution 1940 and 1970



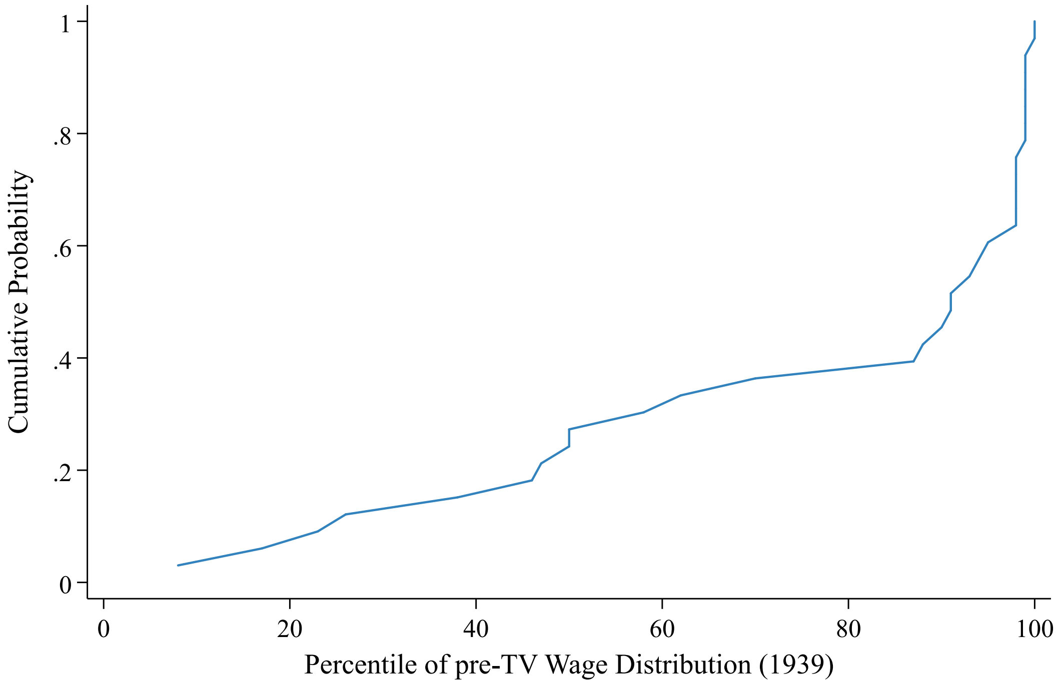
[Note] Log real wage distribution for performance entertainers from the lower 48 states of the US from the 1940 and 1970 Censuses. Real wages in 1950 USD using Census sample weights. Density is estimated using the Epanechnikov smoothing kernel with a bandwidth of 0.4. Common top code applied at $85,000.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

Figure 6: Entertainment Employment per Capita



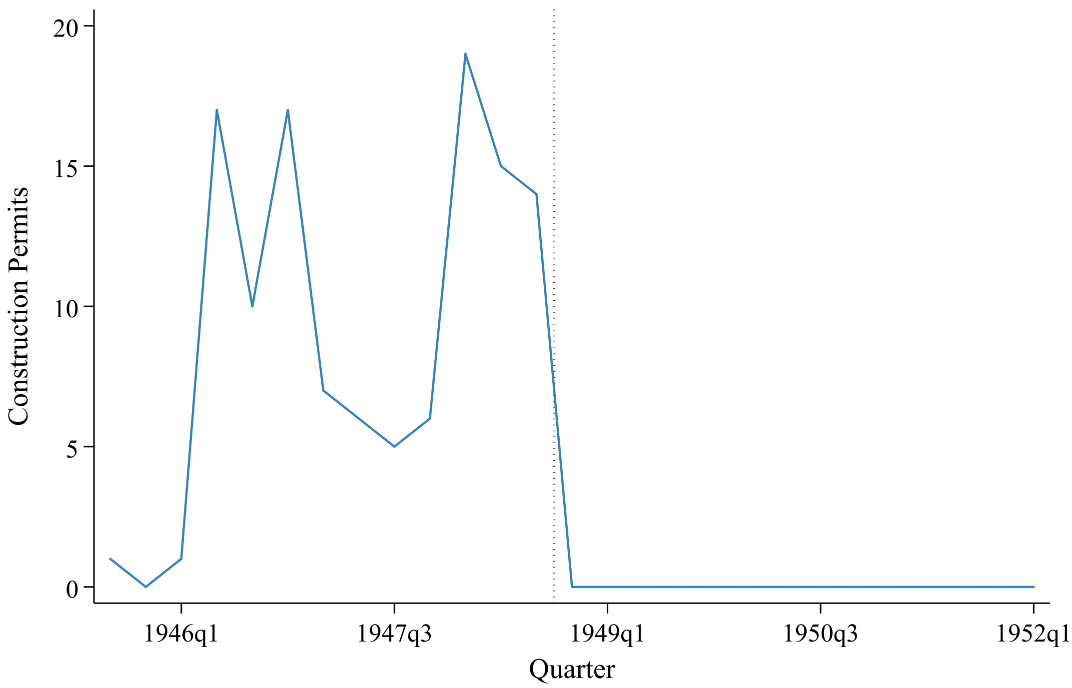
[Note] Employment is measured per 100,000 inhabitants. The mean for performance entertainers is 49 and for other leisure occupations 468. For consistency with early Census vintages, the employment measure includes the unemployed when they report an occupation. Other Entertainment includes "drink and dine" occupations and "interactive leisure" professionals. Definitions along with the definition of performance entertainers are given in the text. Sources: US Census 1940, 1970.

Figure 7: Position of Future TV Stars in the 1939 US Wage Distribution



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

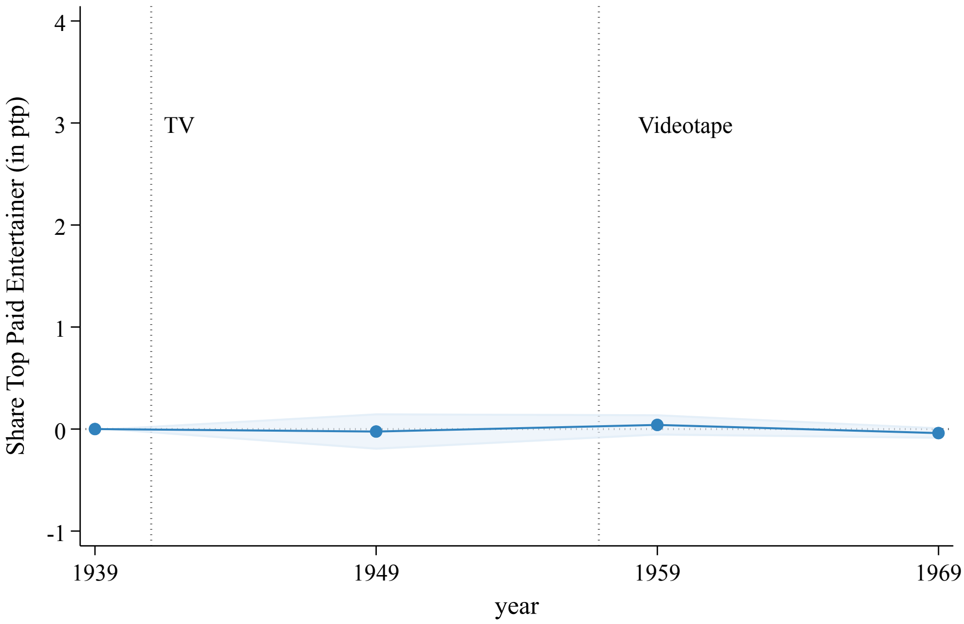
Figure 8: Number of TV Licenses Granted



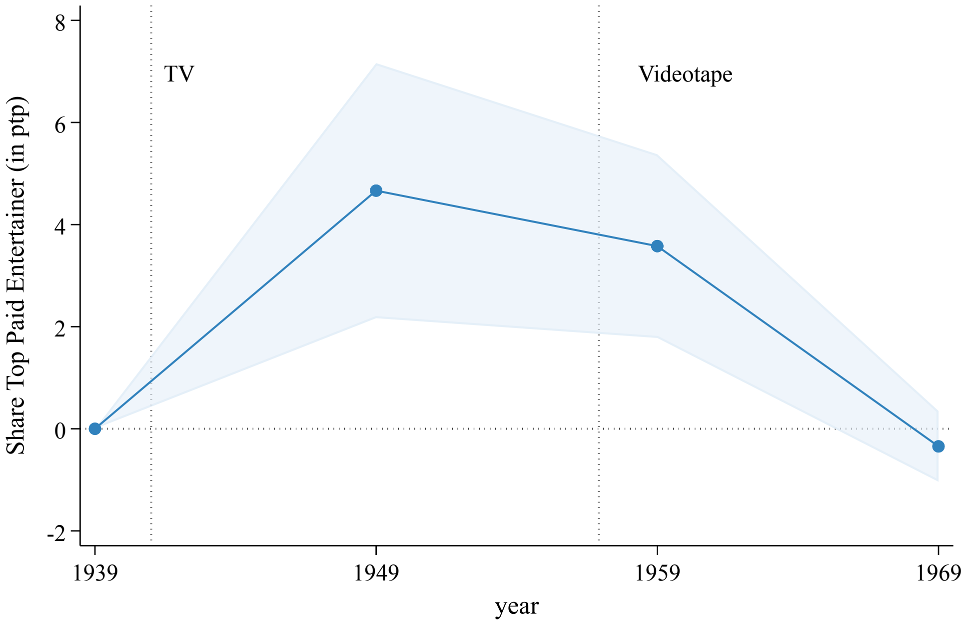
[Note] Missing issue dates of construction permits are inferred from start of operation dates. Source: TV Digest reports.

Figure 9: Dynamic Treatment Effect of TV on

(a) *Blocked TV Stations*

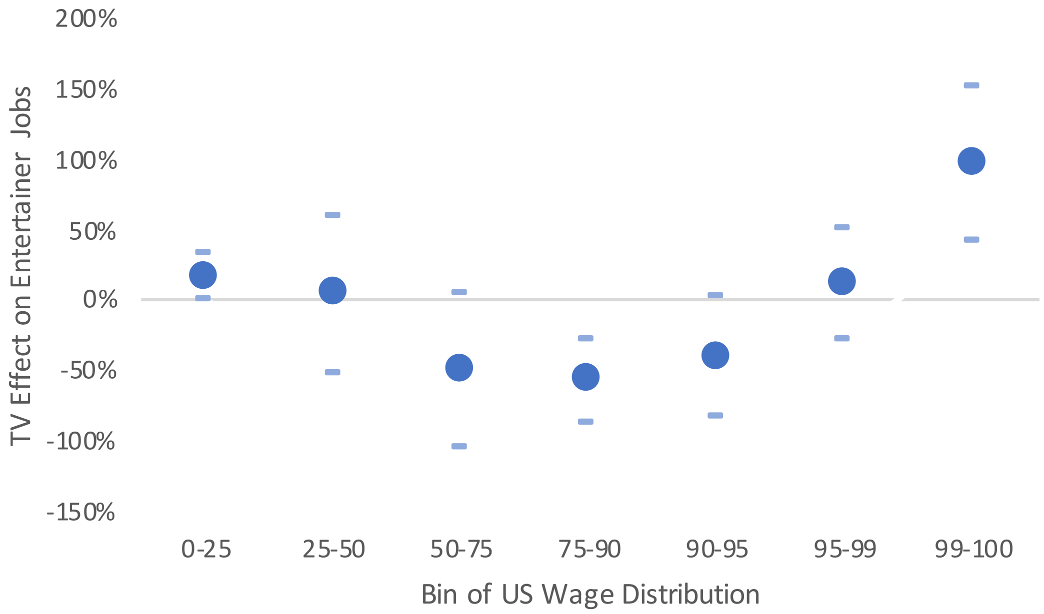


(b) *Active TV stations*



[Note] Figure plots treatment coefficients from two dynamic difference in differences regressions. Panel a) shows the coefficient on $FrozenTV_{m,t}$ (comparison groups are untreated areas) and panel b) shows the coefficient on $TV_{m,t}$. Top-paid entertainers are in the top 1% of the US income distribution. Vertical lines mark the beginning of local TV ("TV") and the end of local TV ("Videotape"). The area shaded in light blue marks the 95% confidence interval. Standard errors are clustered at the local labor market level.

Figure 10: Effect of TV on Entertainer Employment Growth at Different Wage Levels



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

8.C Theory Extensions

A General Superstar Model

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

Labor Supply

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

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

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

Production

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

Assumption P1: absolute advantage $C_a < 0$

Assumption P2: comparative advantage $C_{ta} < 0$

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

Assumption P2.a: comparative advantage II: $\frac{\partial lnC_a(a,t)}{\partial t} < \frac{\partial lnC(a,t)}{\partial t}$

Labor Demand

Demand for the different types of labor is determined by the distribution of tasks. To derive demand for tasks, we need to model the use of tasks. Assume tasks contribute to an aggregate production function $Y(f_t(i), f_t(j), ...)$.[61] Moreover we assume that there is a continuum of tasks $i\epsilon[0,1]$ and Y has well defined derivatives for all i.

Assumption D1: firms maximize profits by choosing tasks:

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

and the market for output clears

$$\int_0^1 P(t) f_t(t) dt = K$$

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

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

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

Equilibrium in the Labor Market

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

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

1. *the assignment meets positive assortative matching (PAM):*

$$t'(a) > 0 \tag{15}$$

2. *wages guarantee IC by:*

$$\frac{\partial ln(w(a,t))}{\partial a} = -\frac{\partial ln(C(a,t))}{\partial a} \tag{16}$$

[61] An equivalent approach thinks of tasks as contributing to a utility function.

The first equilibrium result is that more able workers perform harder tasks.[62] For the proof *of PAM:* see Sattinger notes (uses L1, P2). The intuition is that because markets are perfectly competitive and output of workers is perfectly substitutable within each task, the allocation of talent is completely determined by comparative advantage. Since better workers have a growing edge the harder a task gets, equilibrium assignment perfectly sorts worker ability and task difficulty. The second equilibrium condition ensures that the assignment is incentive compatible. The equality states that wages grow in line with productivity, which is a classic result of assignment models. Workers are paid their marginal product because there is perfect competition in the production of tasks.

Sattinger (1979)

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

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

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

Sattinger (1975)

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

Terviö (2008)

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

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

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

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

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

[62] Costinot & Vogel (2010) derive further structure of the assignment function from the market clearing conditions. They require 1-to-1 matching, see proofs.

Gabaix & Landier (2008)

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

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

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

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

Teulings (1995)

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

$$P(t) = Y^{\eta} f_t(t)^{-\eta} \tag{17}$$

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

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

Costinot & Vogel (2010)

D1: Aggregate production is similar to Teulings and follows a CES: $Y = \left[\int_{\gamma}^1 f_t(t)^{-(\eta-1)} dt\right]^{\frac{-1}{\eta-1}}$ and $K = Y$. Different from Teulings, they additionally allow that tasks are excluded from production. They prevent that the marginal product of a task goes to infinity if demand goes to zero by introducing a participation threshold γ. Only tasks above γ are used.

Additionally they impose 1-to-1 matching. This allows them to arrive at a closed form solution for the slope of the matching function (see proof below)

$$t'(a) = \frac{f(a)}{C(a,t)} \frac{P(t)^{\frac{1}{\eta}}}{Y} > 0$$

Rosen (1982)

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

D1: Rosen assumes all tasks as perfect substitute in aggregate production. Hence, each quality unite $(q(t))$ produces the same amount of output, independent of the task. Tasks only differ in how often they are used in production. As before demand for a task is denoted $f_t(t)$. Total output of quality unity is therefore $Y = \int q(t)f_t(t)dt$. The aggregate producer problem becomes:[63]

$$max_{m_t} \int [q(t) - p(t)]\, f_t(t)dt$$

$$\int f_t(t)p(t)dt = K$$

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

Equilibrium

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

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

$$w_t(a,t) = 1/C(a,t)f_d'(t) - C_t(a,t)/C(a,t)^2 f_d(t) = 0$$

this implied that $f' > 0$. Since the cardinal value of tasks has no meaning, we can use

[63]Notice that we do not impose $K = \int f_d(t)w(t)$, spending in the sector does not necessarily equal income. As a consequence productivity growth means that fewer workers are needed to produce the demand K. The intuition for this assumption is that there is an outside good and spending on can therefore differ from income in the sector. Formally, K then depends on relative prices of the outside good and the elasticity of substitution between the superstar good and the outside good. For simplicity I suppress that complication.

any monotone transformation of the task label. Since tasks here denote market size it is convenient to chose $t = f_d(t)$, this implies $f'_d = 1$. Taking the derivative of the IC condition with respect to a we get:

$$\frac{\partial f_d(t)}{\partial a} = t'(a) = -\frac{\partial^2 lnC(a,t)}{\partial t \partial a} > 0 \tag{18}$$

This result pins down the matching of workers to tasks. From comparative advantage ($C_{at} > 0$) it follows that t is increasing in. As in the previous models this model features PAM. Additionally, the model delivers a functional form for the matching function. Using Rosen's terminology, where $t(a)$ is market size, greater talents serve a bigger market.

Equipped with the IC and PAM condition we can study the wage distribution. Wages at the optimal assignment are $w(a,t) = t(a)/C(a,t(a))$. How much do incomes differ in this economy? The slope of the wage distribution is given by:

$$w_a(a,t) = -\frac{\partial lnC(a,t(a))}{\partial a} t(a) > 0$$

Which uses the envelope theorem. To assess how the income distribution compares to the ability distribution, take the derivative again: . After rewriting the second derivative is:

$$w_{aa}(a,t) = -\frac{\partial^2 lnC(a,t(a))}{\partial a \partial a} t(a) - t'(a)\left[\frac{\partial lnC(a,t(a))}{\partial a} + \frac{\partial^2 lnC(a,t(a))}{\partial a \partial t} t(a)\right]$$

$$= -t'(a)(\tilde{C}_a + t(a)\frac{d}{dt}\tilde{C}_a)$$

with $\tilde{x}_y = \frac{\partial ln(x)}{\partial y}$.[64] The second derivative is positive as long as $-\tilde{C}_a > t(a)\frac{d}{dt}\tilde{C}_a$. This implies that the marginal product of a worker exceeds the marginal output gain from increasing market size t for that worker. Another way of thinking about the same restriction is that as long as diseconomies of scale are not too extreme, wages are more dispersed than talent. This result is the insight of Sattinger (1975) that wages are more dispersed than talent and skewed to the right. In the Rosen model this result is extended to a model with endogenous task demand.

Rosen illustrates two interesting comparative statics of this model. First, consider the case where producing at a large scale becomes easier, falling diseconomies of scale imply $\frac{\partial^2 lnC(a,t)}{\partial a \partial a} \downarrow$. This change makes wages more convex in talent ($w_{aa}(a,t) \uparrow$), top

[64] To arrive at the final equality use $\frac{d}{dt}\tilde{C}_a = \frac{\partial^2 lnC(a,t)}{\partial a \partial a}\frac{1}{t'(a)} + \frac{\partial^2 lnC(a,t)}{\partial a \partial t}$, which assumes t(a) is invertible, aka one-to-one matching.

income inequality therefore increases. The intuition for this result is that workers who operate in the largest market benefit most when the cost of large markets is relaxed. Since we have PAM, it is the high ability workers who benefit most from falling diseconomies of scale.

A second comparative static is the case of growing comparative advantage ($\frac{\partial^2 lnC(a,t(a))}{\partial a \partial t} \uparrow$). From the PAM condition 18 it follows that $t'(a) \uparrow$for all levels of a. Hence all workers work in bigger markets, moreover markets for the best worker grow the most. The expansion of market size for all types of workers implies wages for some workers will fall. Call the new market size distribution $\hat{t}(a) > t(a)$, the budget constraint becomes $\int \hat{t}(a)p(t)dt = K$. Since $\hat{t}(a) > t(a)$ for all a, the budget constraint requires that $p(t)$ declines. Recall that wages are $w(a,t) = p(t)t$, a drop in $p(t)$ may therefore offset the rise in t and thus lead to falling wages, even though all workers became more productive. The model can deliver falling wages without technical regress.

Contradiction with Canonical Skill Biased Technical Change Models

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

Assume both CRS and comparative advantage holds. Recall that CRS implies that if we increase all inputs by κ, total output growths by κ:

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

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

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

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

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

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

Equilibrium of the Superstar Model

Worker talent and firm follow a continuous and differentiable distribution. This will guarantee competitive like behavior of the market, despite the fact that each worker and firm type is unique. Denote talent at percentile p by $T_p = W^{-1}(p)$ and firm size at percentile i by $S_i = F^{-1}(i)$. For simplicity I assume that only one worker can appear on each stage, each firm therefore hires at most one worker.[65] This assumption implies one of the three key assumptions of superstar effects, the imperfect substitutability of workers with team of workers. The production function is given by $\widetilde{Y(S,T)}$. We can use the CDF above to equivalently express output in terms of stage size (S) and the talent rank of the worker (p). $Y(S,p)$. has the following properties: $Y_S > 0$, $Y_p > 0$, $Y_{pp} \leq 0$ and $Y_{Sp} > 0$ where subscripts denote partial derivatives. The final assumption guarantees that talented workers have a comparative advantage in bigger markets. This assumption is another key assumption of the model, also known as single crossing assumption. It will generate positive assortative matching in equilibrium.

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

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

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

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

i) assortative matching: $i = p$

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

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

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

$$Y_i(p) - w(p) \geq Y_i(p') - w(p') \quad \forall \;\; p' \epsilon [0,1] \tag{19}$$

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

[65] The model extends to a setting were many workers are matched to a given market (see Sattinger (1993); Eeckhout and Kircher (2018))

$w^{res}$and the reservation profits π^{res}. We assume they are the same for all workers and firms. The PC condition is thus given by:

$$Y_i(p) - w(p) \geq \pi^{res} \quad \forall \;\; p\epsilon[0,1] \tag{20}$$

$$w(p) \geq w^{res} \quad \forall \;\; p\epsilon[\bar{p},1] \tag{21}$$

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

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

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

$$w'(p) = Y_p(S_i, p) \tag{22}$$

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

Condition *ii)* pins down the wage distribution up to a constant. Wages are increasing for more talented workers and under mild assumptions returns are convex, that is small differences in talent translate into growing wage differences at the top.[66] Consider worker at percentile p, a firm will pay this worker $w'(p)$ more than a slightly less talented worker. The wage increase is exactly equal to the additional output that the worker produces over the next best worker. Note that a firm pays what the worker is worth to it, while the same worker would be worth less to a smaller firm (since $Y_{pS} > 0$). Despite the worse outside option of the worker, the firm passes all the output gains to the worker. This may seem surprising but holds because we assume firm types are distributed continuously. The outside option of the worker is thus only infinitesimally worse and the worker receives the full productivity gains of moving to a bigger firm. Take the alternative case where the distribution of firm types has jumps, some theatre venues are thus discretely bigger than their competition. Here the firm would keep all of the productivity gains. The lack of competition among employers therefore dampens wages. Condition *iii)* pins down employment levels in the market. All workers with

[66]The wage distribution is convex : $w''(p) = \frac{Y_{pS}}{f(S_p)} + Y_{pp} > 0$ as long as the return to talent is not diminishing sharply and talent at any percentile is sufficiently scarce $-Y_{pp} < \frac{Y_{pS}}{f(S_p)}$

talent above $\overline{p}$ are active. In the extreme case of a winner takes all market $\overline{p} \rightarrow 1$ and only the most talented worker prevails.

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

$$w'(p) = \pi \cdot T'(p) \cdot S_p^{\gamma} \tag{23}$$

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

$$w(p) = \int_{\bar{p}}^{p} \pi \cdot T'(j) \cdot S_j^{\gamma} dj \tag{24}$$

In equilibrium the supply and demand for talent $D(\pi)$ clear at the market clearing price for talent (π):[67]

$$D(\pi) = \int_{\bar{p}}^{1} T(j) \cdot S_j^{\gamma} dj \tag{25}$$

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

To analyze wage changes we need more information on the functional form of the firm size and talent distribution. Assuming a Pareto distribution allows for a tractable solution to the model. This is a strong assumption that is not required but helps the exposition of the model. The conclusions hold more broadly, Appendix 8.C discusses the assumption in more detail and derives a more general solution. The CDF of talent and stage size ($W(p)$ and $F(i)$) is given by $T_p = (1-p)^{-\beta}$and $S_i = (1-i)^{-\alpha}$, where

[67] A downward sloping demand guarantees that an equilibrium exists. The RHS is increasing in π, since more workers enter the market when the returns are high. $\bar{p}$ therefore increases when π falls. The assumption is that the supply response happens along the participation margin. The model has however been extended to include an hours response Scheuer and Werning (2017).

β and α are the shape parameters of the Pareto distributions. A higher value of these parameters means there is greater dispersion. Substituting the distributional equations into equation 24, wages become:

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

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

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

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

Growing Dispersion in Size

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

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

with $\psi \equiv \frac{\pi^{t+1}}{\pi^t} \frac{(\alpha\gamma+\beta)}{(s*\alpha\gamma+\beta)}$, superscripts denote the time period. As benchmark consider the case were labor supply is perfectly inelastic, hence $\bar{p}$ is fixed. There is therefore no entry and exit and percentiles do not carry time superscripts. Allowing for entry has negligible effects on top wages and the results thus would carry through approximately. At the bottom of the distribution entry would matter.[68]

Dividing both sides by w_p^t we get wage growth (g_p^w) from a superstar effect:

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

Notice that the growth rate depends on p. Wages grow more at the top of the distribution. The effect on the wage schedule is illustrated in figure 11. Two features stand out. For one, superstar effects pivot the wage schedule. Second the impact across the wage distribution is U-shaped. Gains for the superstars come at the expense of less talented workers, while wages at the bottom remain pinned down by the outside option. Two factors drive these effects. For one, top entertainers can use their talent

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

more intensely through the new technology and more consumers get to see the star entertainer. This effect is captured by last term in equation 29. The biggest wage gains occur at the top of the distribution and the returns to being a superstar have risen.

There is a second effect that comes from the greater availability of stars. The wider availability of stars has reduced the price for a unit of talent ($\pi^t > \pi^{t+1}$).[69] If the audience of an entertainer is unchanged marginal revenue product therefore decreases. The wage at any given stage declines. This effect is captured by $\psi < 1$ and is a level shift downward in earnings for all entertainers. Taking these effects together the wage schedule has shifted downwards and pivoted upward.

Distributional Assumptions

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

$$w_{pS}(p^*) = Y_{pS}(p^*) + Y_{pp}(p^*)\frac{\partial p^*}{\partial S} = \frac{w''(p^*)}{\theta'(p^*)} > 0 \tag{30}$$

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

The closed form solution derived in the text also holds for a broader set of assumptions. The assumption that talent is Pareto distributed can be relaxed. Extreme value theory has shown that many functions look similar in the upper tail. Gabaix and Landier (2008) show that we can apply this to the assignment problem and solve for top incomes, given mild assumptions on the talent distribution. They noted that for "regular" continuous distributions the tail of the distribution can be approximated by:

[69] The intuition for this result is that the average market size increases with α. Formally, we can look at the RHS of equation 25. 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.

$T'(p) = -Bp^{\beta-1}$. This holds exactly, as for Pareto distributions, or up to a slowly varying function.

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

Skill Biased Technical Change and Pay Dispersion

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

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

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

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

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

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

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

$$F^w(W) = Pr(w_{t^H} > W) = Pr(t^H > \frac{W}{w_H}) = G^H(\frac{W}{w_H})$$

The top tail of the wage distribution follows the same distribution as t^H.[70] The wage at percentile $p(W_p)$ is than given by:

$$W_p = w_H G_H^{-1}(p)$$

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

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

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

Next consider wages. The baseline case assumes that labor supply is inelastic, hence the talent distribution ($G^H(t)$) is unchanged. Allowing for a labor supply response complicates notation and generates little additional insight.[71] The wages at p are given by:

$$\tilde{W_p} = \tilde{w_H} G_H^{-1}(p)$$

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

$$g_p^w = \frac{\tilde{W_p}}{W_p} = \frac{\tilde{w_H} G_H^{-1}(p)}{w_H G_H^{-1}(p)} = \frac{\tilde{w_H}}{w_H} = g^w$$

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

Technical Change and Pay Dispersion in the Income Tail

This section derives proposition 2.4. The top income share is defined as the sum of incomes of individuals above percentile p divided by total income (G):

---

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

[71] The higher wage induces entry of workers where $\tilde{w}_t$ growths above the outside option b. These are workers with low levels of t and as a result the distribution of talent changes at the bottom end. For ordinary talent distributions this has little effect on the top tail of $G_H^{-1}(p)$. The result that follow therefore carry through approximately at the the top of the distribution.

$$s_p = \int_p^{\infty} \int_{\bar{p}'}^{i} w_j dj di / G$$

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

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

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

This contrasts with the impact of superstar effects. Note that the top income share above percentile p for a Pareto distributed variable is given by $s_p = (1-p)^{1-\lambda}$, with λ^{-1} the shape parameter of the distribution.[72] Here wages follow a Pareto distribution with shape parameter $\lambda = \alpha\gamma + \beta$. The growth in the top income share from superstar effects is therefore given by:

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

The second equality uses the property of a Pareto variable, while the final equality cancels terms. Top incomes shares are again growing. But the pattern is different from SBD shocks. Here the growth rate is increasing in p. This implies that the income share of the top 0.1% growths faster than the share that goes to the top 1%, which in turn growths faster than the share of the top 10%. A growing fraction of the top 10% income share is taken home by the top 1%.

Proof: No Cannibalisation in SBTC Models

This section proofs that technical progress rules out falling wages in the SBTC model. I study a flexible SBTC model with arbitrary many skill groups 1 ... n. The production

[72] This result has been used extensively to calculate top income and wealth shares. Even for variables that do not follow a Pareto distribution, there is still a lambda now varying with p. Many income variables are approximately Pareto and lambda is only slowly varying and the result holds approximately.

function is given by:

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

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

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

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

$$\frac{\partial F}{\partial \theta} = \sum L_i \frac{\partial \alpha_i}{\partial \theta} \frac{\partial F}{\partial L_i} > 0 \tag{31}$$

We want to show that this implies that:

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

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

Superstar Effects and the Link of Talent and Pay

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

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

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

$$ln\left(\frac{w(\tilde{n})}{w(n)}\right) = ln\left(\frac{T'(\tilde{n})}{T'(n)}\right) + \gamma * ln\left(\frac{S(\tilde{n})}{S(n)}\right) + ln\left(\frac{\tilde{n}}{n}\right) = ln(\omega) + ln\left(\frac{\tilde{n}}{n}\right) \quad (32)$$

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

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

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

$$ln\left(\frac{w(\tilde{n})}{w(n)}\right) = ln(\omega)(\frac{\epsilon}{\epsilon + 1}) \quad (33)$$

The wage premium thus exceeds the talent premium by a constant factor. In a superstar model wages amplify differences in returns to talent. The model thus breaks the proportionality of talent and pay. The mark up is particularly stark if ϵ is large. That is the case if wages are very dispersed. In terms of model primitives this corresponds to the case where the relative scarcity of talent and market size (α/β) is large and the returns to market size don't diminish quickly (γ).

Superstar Parameters and Employment Elasticities

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

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

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

Where we define G(x) as the share of individuals above x. Consider a small increase in the number of top earners. If the order of individuals in the distribution has remained the same we can re-write the expression.[73] 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 employment to an associated shift in wages. Assuming a Pareto distribution will again proof useful. A convenient property of the Pareto distribution is that the tail of the distribution has a well defined shape with $\frac{f(x)}{G(x)} = \alpha/x$. Using this fact we can re-write the above equation in terms of elasticities $\varepsilon_{i,j}$ with α the Pareto coefficient:

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

[73] 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.

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

This gives us a simple expression for the link between the elasticity of number of top earners and the elasticity of pay. We can apply this expression to the results in this study and get an alternative estimate for the elasticity of top pay to market size. I estimate the α parameter on the pre-TV wage distribution using the full count, non top-coded Census of 1940. I experiment with a number of estimation strategies with similar results.[74] 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.

Using the results above the estimates to compute the elasticity of top pay employment to market size we find $\varepsilon_{TE,m} = 0.45$.[75] Using the relation derived here we can translate this into an elasticity of income. The implied elasticity of top wages to market size is $\varepsilon_{w,m} \approx 0.15$. A doubling in market size will thus raise top wages around 15%. This is remarkably close to the wage elasticity that we estimated directly from the data (recall it was 0.12).

The wage elasticity would be bigger if we relaxed the assumption that the effects are order preserving. Without such homogeneous treatment effects individuals from further down in the wage distribution could become earning superstars. This would require a large wage rise for these people and thus potentially increase the estimated elasticity. To assess how much this matters in practice, we need to know where in the income distribution local TV stars came from. Figure 7 plots this. The figure matches local TV stars to their pre-TV earnings in the 1939 Census.[76] 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.

---

[74] 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.

[75] This is based on panel B of table 1 and table 2.

[76] 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.

8.D Data and Robustness Checks

8.D.1 Robustness checks

Top Income Metrics

The baseline outcome variable normalizes the number of top earners by aggregate employment in entertainment. This has the convenient effect that the result is a percentage change. The numerator doesn't vary at the local labor market level, changes in this variable should therefore be captured by the year fixed effect. We may however worry that since the variable enters multiplicatively, the additive year fixed effect doesn't completely control for changes in the denominator. In column 2 Table 11 I therefore rerun the baseline regression using the count of top earners as outcome. In an average labor market 18 individuals are in the top percentile. TV more than doubles the number of top earners. Column 1 repeats the baseline regression. The normalization changes the units of the results, but the basic conclusion remains unchanged. This confirms that the normalization has no substantive effect on the result.

Figure 16 illustrated the evolution of various alternative top income measures. The figure shows the the 99th percentile of the Census wage distribution over time. This is the threshold that defines top earners in the baseline estimates. The figure contrasts this threshold with alternative top income thresholds. These include the thresholds calculated by (Piketty and Saez, 2003) and the 95th percentile of the wage distribution and the 95th percentile of the entertainer wage distribution. All of these are below the wage top-code applied in the data. The series move similarly. In practice it will therefore matter little how a top earner is defined. Table 11 confirms this formally. It repeats the previous analysis using other top income measures. Column 1 repeats the baseline estimate. Column 3 uses the top income percentile as defined by (Piketty and Saez, 2003). With this definition of top earners slightly more entertainers are top earners. The effect of TV remains however unchanged. The the number of people in the top percentile about doubles.

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

Pre-Trend

A challenge for estimating pre-trends with this sample is that wage data in the Census is first collected in 1939. Since the Census is decennial this only allows for a single pre-treatment period. To estimate pre-trends I therefore combine the Census data with data from Internal Revenue Services (IRS) tax return data. In 1916 the IRS published aggregate information on top earners by occupation-state bins. Data for actors and athletes are reported. I link the Census data with the tax data and run the regressions at the state level. Table 13 reports the results. Column 1 repeats the baseline estimate with data aggregated at the state level. Despite the aggregation at the state level the effect remains highly significant. Column 2 adds the additional 1916 data from the IRS. The results stay unchanged. Column 3 shows the differences in top earners in treatment and control group for the various years. It shows a marked jump up in top earners in the treated group in the year of local TV production. The coefficient on the pre-trend is not significant because the standard errors are large. If anything the pre-period saw to a decrease relative decrease in top earners in the treatment areas. Even if taken at face value the pre-trends thus can't explain the identified positive effect of TV.

Placebo Occupations

Television only changed the production function of a handful of occupations, we can therefore use alternative occupations as placebo group. The ideal placebo group will pick up changes in top income in the local economy. The main high pay occupations are therefore used as placebo group, these professions are medics, engineers, managers and service professionals. If TV assignment is indeed orthogonal to local labor market conditions, we would expect that such placebo occupations are unaffected. Results for the placebo group are reported in 14. TV does not show up in top pay of the placebo occupations. The only occupation group with a significant positive effect are performance entertainers. Column 1 shows that the placebo group doesn't experience any growth in top incomes. Moreover, the estimated effect on performance entertainers remains similar to the baseline in Table 14. Column 2 allows for separate impact of television across the different placebo occupations. Only performance entertainers experience the significant and large top earner rise.

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

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

Quantile Regressions

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

A shortcoming of the quantile regression is that the estimates are sensitive to entry and exit. The magnitude of the quantile effect is therefore hard to interpret. However, the relative magnitude across percentiles is still informative and the test relies exclusively on such relative patterns. Recall that SBD predicts a homogeneous growth rate, while the superstar model predicts larger wage growth rates at the top. To test whether either model matches the data, I run quantile regressions at various percentiles. I restrict myself to quantiles for the median and above since the results were derived by using an approximation for the top of the distribution. I follow the procedure in Chetverikov et al. (2016) to implement the difference in difference for quantile regressions. The estimated coefficients are plotted in figure 18, alongside the prediction of the SBD model. The effect is biggest at the top of the distribution and effects are notably smaller at the lower percentiles. This result is in line with the superstar model but contradicts a model of SBD. Table 15 reports the panel quantile estimates using the Powell (2016) approach.

Policy Effects in a Superstar Setting

A leading policy to battle inequality is investment in education. Arguably, modern production technologies require greater skill and are therefore driving up demand for skilled workers. In line with this argument, the wage premium for skilled workers has been rising (Acemoglu and Autor (2011); Katz and Murphy (1992)). Investment in education would increasing the relative supply of skilled labor and thereby reduce inequality. In superstar models by contrast, the level of education does not affect inequality. In such models, the rank position in the ability distribution determines pay differences. Changes in the skill level of the workforce have no material effect on inequality in this model. I can test this prediction empirically by interacting the treatment with the local high school graduation share, which is admittedly a rough proxy for education levels but has been widely used in the literature on inequality. The interaction is insignificant, suggesting that the superstar effect is independent of the skill level in the local labor market (column 1 of Table 17). Since the standard errors are large, these results can however, not be interpreted as conclusive evidence against skill investments.

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

I test whether superstar effects differ under different tax regimes. This test exploits variation in top income tax rates across US states. Data on states' historical tax rates are not centrally collected. I compiled such data from the study of historical state taxation in Penniman and Heller (1959), who collect detailed information on income tax legislation across US states during the sample period. Using this information I construct a dummy variable that is equal to one for high-tax states, aka states with tax rates above the median.[77] I test how higher tax rates affect the rise in top incomes in a superstar setting. This estimate combines the effect of out-migration and reduced labor supply by stayers. Column 2 of Table 17 shows there is no significant difference between high- and low-tax states. While the standard errors are large, the point estimate on the interaction term is quantitatively close to zero. There is thus no evidence that high taxes lead to substantial distortions in superstar markets, nor that taxes are able to substantially slow the rise of superstar earning.

8.D.2 Data construction

Local labor markets

- The analysis defines local labor markets as commuting zones (CZ). A labor market is an urban center and the surrounding commuters belt. The CZs fully cover the mainland US. The regions are delineated by minimizing flows across boundaries and maximizing flows within labor markets.
- David Dorn provides crosswalks of Census geographic identifiers to commuting zones (Autor and Dorn, 2013). I use these crosswalks for the 1950 and 1970 data
- No crosswalk is available for the 1960 geographic Census identifier in the 5% sample and the 1940 Census data. Recent data restoration allows for more detailed location identification than was previously possible (mini-PUMAs).
- to crosswalk the 1940 data, I use maps that define boundaries of the identified areas. In GIS software I compute the overlap of 1940 counties and 1990 CZ. In most cases counties fall into a single CZ. A handful of counties are split between CZ. For cases where more than 3 percent of the area falls into another CZ, I construct a weight that assigns an observation to both commuting zones. The two observations are given weights so that they together count as a single

[77] I use a binary variable because marginal tax rates are difficult to interpret in this context. Deductibility rules generate a wedge between MTR and headline rates. This is less of a problem for comparing high- and low-tax states to the extent that deductibility rules don't change whether a state is a low- or high-tax state.

observation. The weight is the share of the county's area falling into the CZ. The same procedure is followed for 1960 mini PUMAs

- Carson city county (ICSPR 650510) poses a problem. This county only emerges as a merger of Ormsby and Carson City in 1969, but observations in IPUMS are already assigned to this county in 1940. I assign them to Ormsby county (650250)
- CZ 28602 has no employed individual in the complete count data in 1940.

Worker data

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

- there are 722 commuting zones (CZ) covering the mainland USA. These regions are consistently defined over time.
- there are 28 relevant occupations. 1950 occupation codes are
  - Treatment group: 1, 5, 31, 51, 57
  - Placebo group: 0, 32, 41, 42, 43, 44, 45, 46, 47, 48, 49, 55, 73, 75, 82, 200, 201, 204, 205, 230, 280, 290, 480
- controls are population aggregates in the area: share high skilled (high school and above for people over 25), share non white, median age, sample size per CZ, median wage and age
- Aggregates are calculated using the provided sample weights
- variables used incwage, occ1950 (in combination with empstat), wkswork2, hrswork2

Employment

- Occupation based on the 1950 classification of IPUMS (Occ1950). This data is available for years 1940-1970. For previous years the data is constructed using IPUMS methodology from the original occupation classification.
- Occupational definitions change over time. IPUMS provides a detailed methodology to achieve close matches across various vintages of the US census. Luckily the occupations used in this analysis are little affected by changes over time. More

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

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

- The definition of employment changes after the 1930 Census. Before the change, the data doesn't distinguish between employment and unemployment. In the baseline analysis I therefore focus on the period from 1940 onwards. For this period the change doesn't pose a problem. An alternative approach is to build a harmonized variable for a longer period, this includes the unemployed in the employment count for all years. I build this alternative variable and perform robustness checks with it. The results remain similar. For two reasons the impact of this change on the results is smaller than one might first think. First, most unemployed don't report an occupation and thus don't fall into the sample of interest.[78] Second, the rate of unemployed is modest compared to employment

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

and thus including them doesn't dramatically change the numbers.

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

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

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

Wage data

- Census data on wages refer to the previous calendar year

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

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

- Wage data is in real 1950 terms

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

previously worked. I construct an employment series that includes such workers for the entire sample period. This measure is a noisy version of employment as some job losers continue to count as employed. Since the share of these workers is small, the correction has only small effects on the results.

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

- As a robustness check I use earners above the 99th percentile within their occupation.

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

Pareto Interpolation

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

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

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

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

- It turns out that OLS is a poor approach here. The Gauß Markov assumptions are violated making OLS inefficient and bias. The outcome variables are order statistics, resulting in heteroskedacticity and correlation of errors across observations. Moreover, the log transformation implies that $E(\epsilon_i) = E(log\varepsilon_i) \neq 0$, making OLS biased. The latter problem can be addressed by replacing the regressor with the Harmonic index (Blanchet (2016)). And efficiency can be achieved with MLE.[80] Polivka (2000) and Armour et al. (2015) give an overview how MLE can be applied to this problem. A further challenge is misspecification. The Pareto distribution is used as an approximation and may not fit the data perfectly. In particular the distribution may fit better at the top than the bottom of the distribution. Even at the top of the distribution changing Pareto coefficients may be required to fit the data (Blanchet et al. (2017)). Misspecification is particularly problematic for the more efficient estimators (Finkelstein et al. (2006)). I will test the performance of three estimators using real-world data by drawing samples from the full-count Census. This allows us to assess how estimators cope in data with i) small samples, ii) top coding and iii) bunching at tax thresholds and round numbers. I test the following estimators:

- Estimator with n total observations, T top coded observations, $rank_j$ the rank in the wage distribution (1 being the top), w_j wage at rank j and ω the smallest wage in the sample:
- MLE: $\hat{\beta}^{MLE} = \frac{1}{n}\sum_{j=1}^{n} log(w_j/\omega)$
- MLE (top code adjusted): $\hat{\beta}^{MLE_{TC}} = \frac{T}{n}\sum_{j=T}^{n} log(w_j/\omega) + T * log(w^{TC}/\omega)$
- OLS: $log(w_j) = \delta - \beta^{OLS} * ln(\frac{rank_j}{n+1}) + \epsilon_j$
- Close to cut-off: $\hat{\beta^A} = (\sum_{j=1}^{3} \frac{ln(w_j/w_{j-1})}{ln(rank_j/rank_{j-1})})^{-1}$
- Extrapolation: The standard method of calculating top income shares fits a Pareto curve through the observed data and computes income shares as area under the curve. For the Pareto distribution the fraction that falls in the tail is captured by a single Parameter. We can thus compute any top income share once we know the tail index of the Pareto distribution. For other distributions the tail index varies for different percentiles, in that case we have one shape parameter that allows to compute the top 1% income share and a different one to compute the top 0.1%

considered

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

share. A well known feature of extreme value theory is that in the the tail many regular distribution only differ by a slow moving function from the Pareto. Using the Pareto parameter estimate just below the cut-off may thus yield a reasonable approximation even if the data generating process is not Pareto.

- Table 16 shows the results. They suggest that OLS and MLE perform relatively poorly in small samples of the data of interest. I find that the best performing estimator is the average of the alpha values just below the top code. The difference to OLS and MLE estimates is the weight attached to values far from the top-code. OLS and MLE give a non zero weight to observations further away from the top-code. This approach will yield greater bias if the Pareto distribution is not a perfect fit and observations far from the top-code are poor proxies for the distribution beyond the top-code. Consistent with this, I find that the OLS and MLE perform worse in smaller samples. For the application here I therefore focus on Pareto interpolation based on observations closest to the top-code. It should be stressed that this result is specific to the data in this context. More general results for Pareto inference with real-world data should be conducted to establish the wider relevance.
- For each local labor market and year I derive the Pareto coefficient. At the bottom of the income distribution the Pareto distribution has been found be a poor fit, I therefore discard Pareto parameters based on observations at the bottom quarter of the distribution. The results are however robust to including those observations. Next, I use the local labor market- year specific Pareto coefficient to estimate top income shares. Here I make use of the fact that for a Pareto distribution top income shares are given by: $S_{p\%} = (1-p)^{\frac{\alpha-1}{\alpha}}$.

Controls

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

IRS Taxable Income Tables

Data from the Internal Revenue Service (IRS) allows me to extend income data backward beyond what is feasible with the Census.[81] To obtain records for entertainers,

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

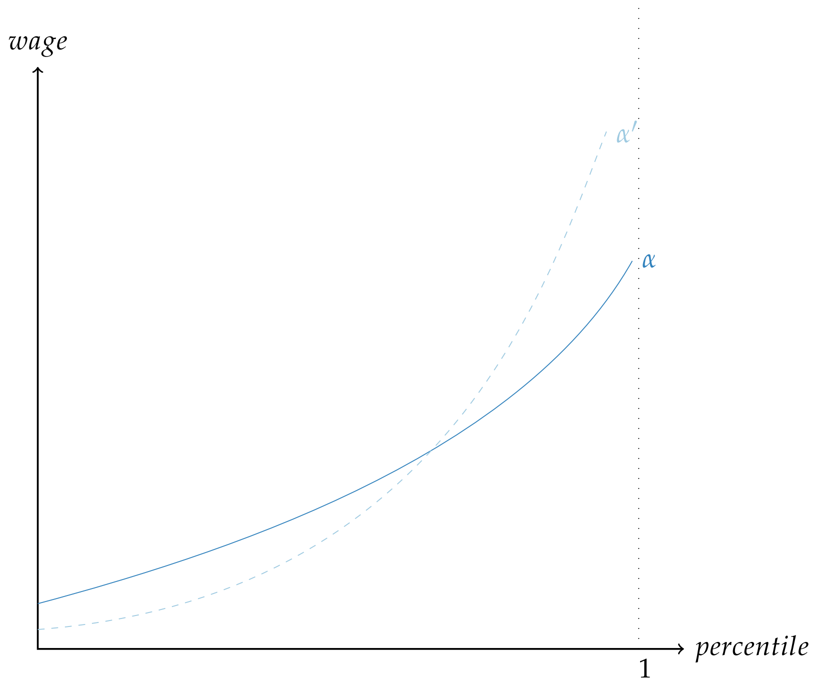
I digitize a set of taxable income tables that lists income brackets by state and occupation. The breakdown of the data by occupation and state is only available for the year 1916.

Marginal Tax Rates

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

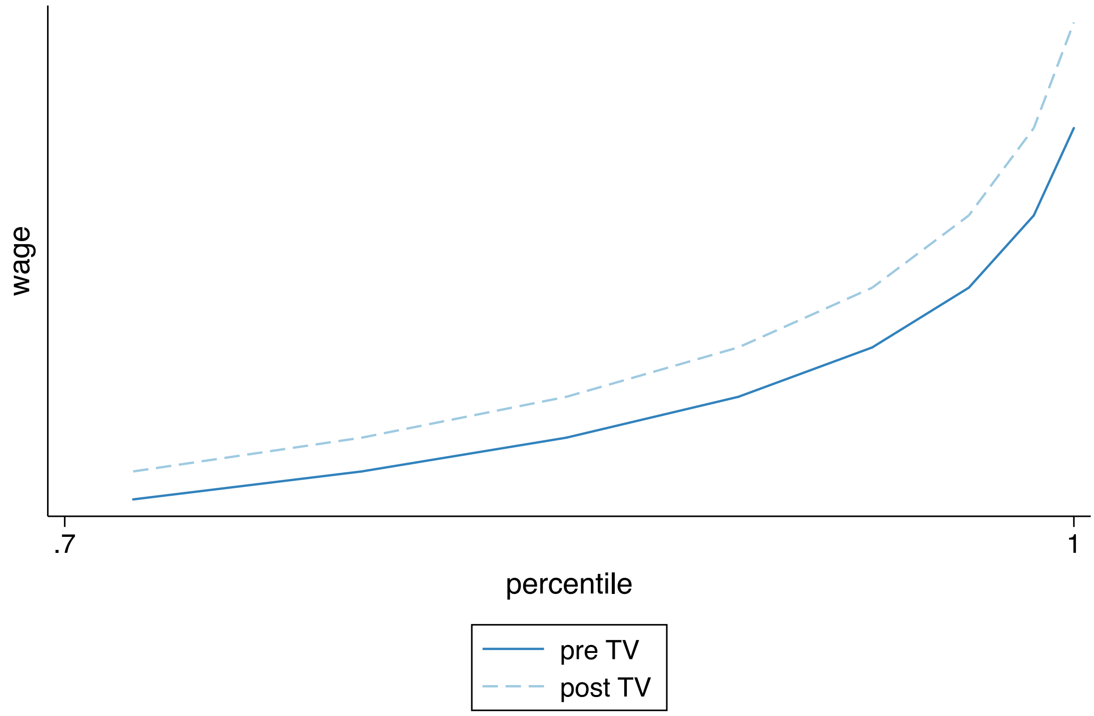
8.D.3 Tables & Figures used in Appendix

Figure 11: Superstar Wage Distribution



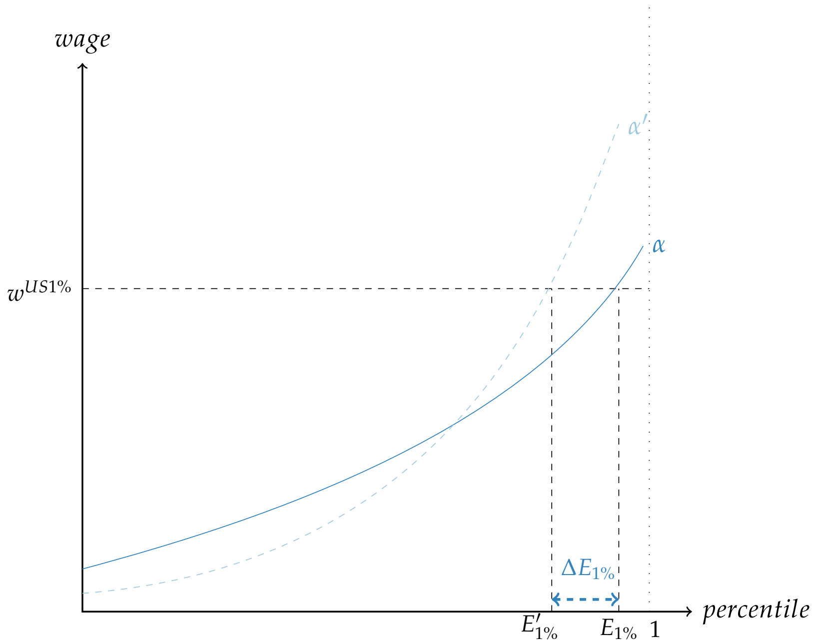
Notes: Wages based on a superstar model ($w_p = \pi \cdot \kappa \cdot (1-p)^{-(\alpha\gamma-\beta)}$). α is the shape parameter of the market size distribution ($\alpha' > \alpha$). The percentiles shown are the upper tail of the wage distribution. With exit they correspond to the percentiles in the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

Figure 12: Effect of Technical Change on Wage Distribution - Skill Biased Demand Model



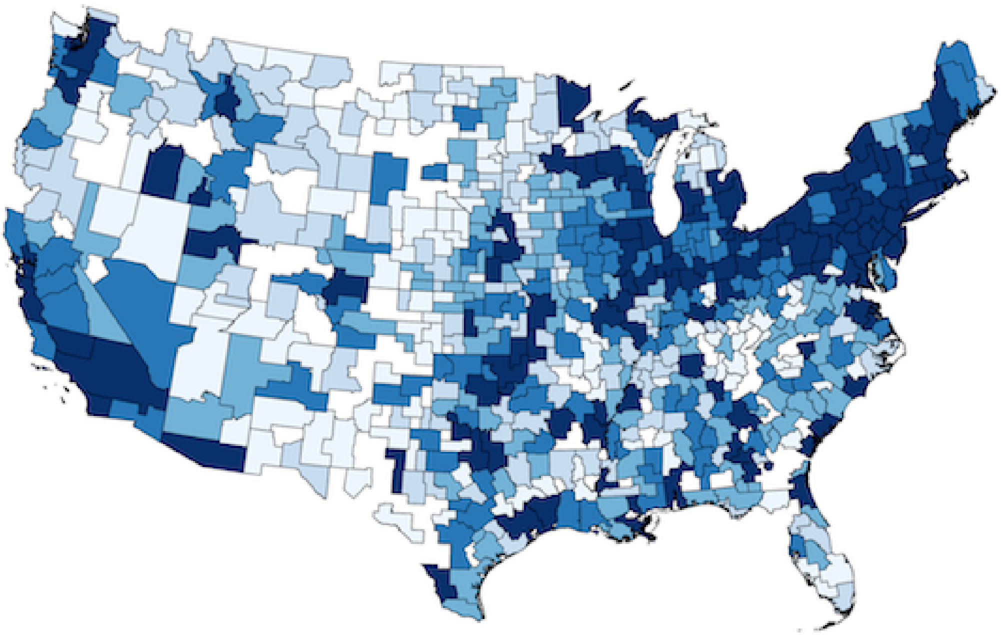
[Note] The figure shows the wage distribution above the 70th percentile. The talent distribution has been chosen to match the 1940 wage distribution. The change in the skill premium matches the growth in the share of top earners.

Figure 13: Superstar Effect on Top Earner



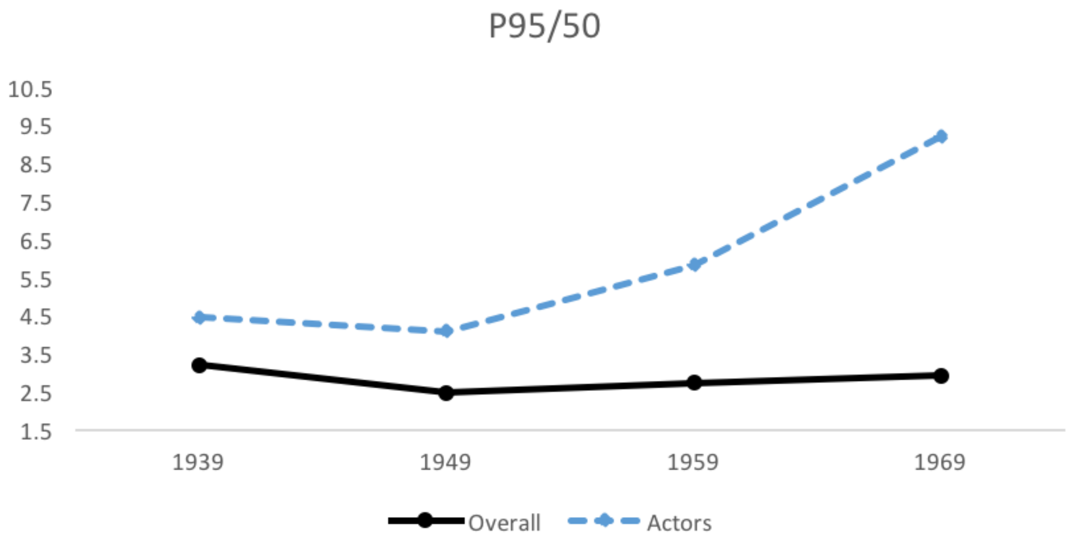
Notes: Details as in figure 11. $w^{US1\%}$ is a wage threshold that defines a top earner, e.g. the national top percentile. $E_{1\%}$ and $E'_{1\%}$ are the share of entertainers above the threshold. $\Delta E_{1\%}$ is the change in top earners when market size becomes more dispersed (move from α to α').

Figure 14: Theatre Seating Capacity



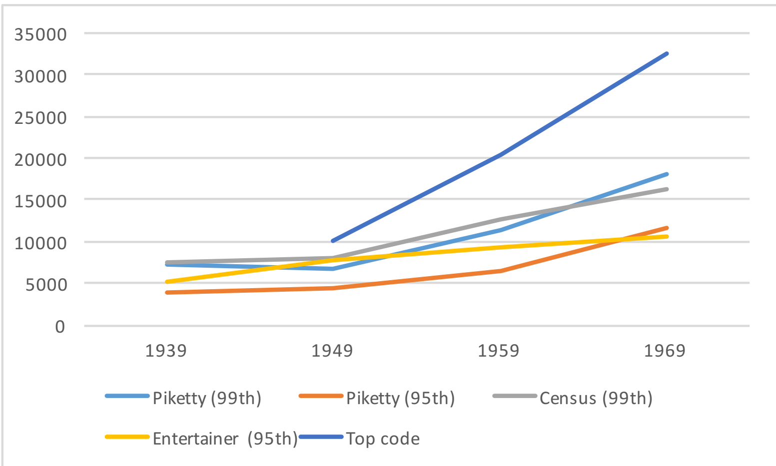
Notes: Performance venues are the venues listed in Julius Cahn-Gus Hill's 1921 theatrical guide. Size refers to the average seating capacity of the largest two venues in the commuting zone.

Figure 15: P95-P50 Gap



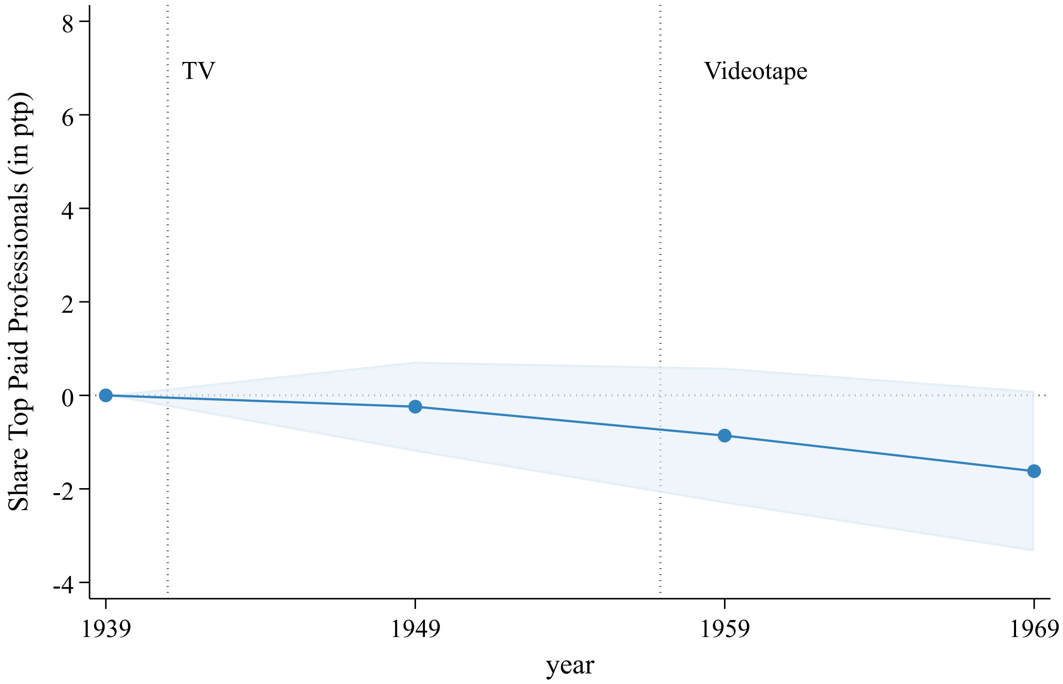
Notes: Figure reports the ratio of wages at the 95th and median. Percentiles are from the wage distribution reported in the US decennial Census for the lower 48 states.

Figure 16: Top Income Percentile Values



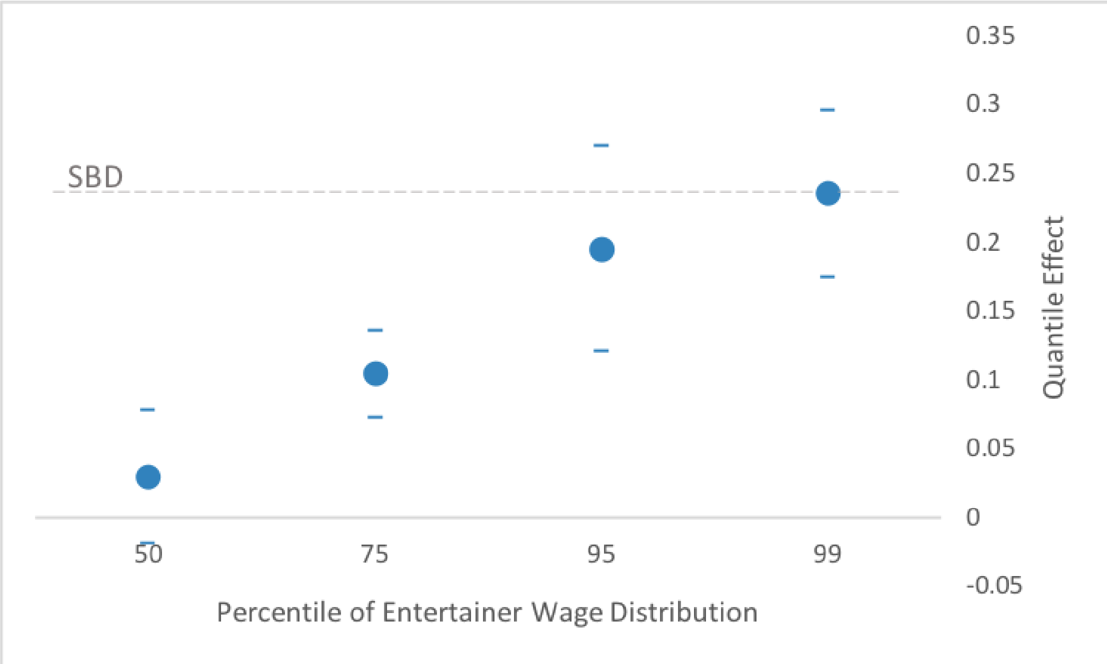
*Notes:*The Figure shows the top code cut-off in the US Census data and top percentiles of the wage distribution in the Census years. The name in the legend refers to the sourcee of the wage distribution: Census refers to percentiles in the Census data wage distribution, Entertainer to percentile in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

Figure 17: Dynamic Treatment Effect of TV stations - Placebo Occupations



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

Figure 18: Quantile Effects of Television



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

Table 9: Effect of TV on Top Earner - Placebo Occupations

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

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

Table 10: Effect of TV on Top Earner - Alternative Top Income Measures

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

Notes: See table 1 Panel B denominator is the average number of entertainers per labor market in occupation o at time t. Denominator in Panel C is the total number of entertainers in local labor market c at time t.

Table 11: Alternative Top Income Measures

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

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

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

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

Notes: The outcome is a dummy that takes the value 100 if an individual is in the top 1% in the US distribution. Columns 1-3 are based on 83,748 individuals and column 4 on 3,438,002 individuals. Placebo occupations are non affected free time professions: drink & dining and active leisure and typical high pay professions: management, medicine, engineering, professional services (finance, accounting, law). The number of observations are 100308. Regressions use provided Census weights and cluster by local labor market.

Table 13: Effect of TV on Top Earner - State Level

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

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

Table 14: Earning Effect - triple diff

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

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

Table 15: Quantile Effect of TV

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

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

Table 16: Small Sample Performance of Pareto Shape Parameter Estimators

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

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

Table 17: Policy Effects in a Superstar Setting

	(1)	(2)
Local TV station	4.83	4.59
	(4.56)	(1.77)
Local TV station × % with high-school degree	-1.42	
	(12.23)	
Local TV station × high tax state		0.10
		(1.65)

Notes: Sources and specification as in baseline. High-tax states are defined as states where the marginal tax rates of the top income bracket exceed the median; data availability restricts observations to 12,977 in this column.

Part II
The Labor Supply Response to Entertainment Technology

9 Introduction

Entertainment technologies have undergone a massive expansion in variety, quality, and availability, from the early advent of radio and TV to more recent innovations like YouTube and Netflix. While these technologies are widely studied and ubiquitous in daily life, little is known about their implications for individuals' leisure choices. Basic economic theory predicts that all else equal, a positive shock to leisure utility will reduce labor supply. The central contribution of this paper is to test that prediction using the natural experiment afforded by the 1950s roll-out of TV in the US.

Next to sleep and work, nothing occupies more of Americans' time than TV. By the end of the 1950s, most households owned a TV and watched for several hours a day. Viewing hours have grown steadily since and are at record levels today (Figure 19). In recent years, online streaming has made TV access even easier.[82] The rapid growth of such services suggests that TV shows may become an even more important part of our daily live. The large historic change in time use illustrates how drastically TV appears to have improved leisure utility. We study if TV only crowds out other leisure activities or if it also affected the labor-leisure trade-off.

A first challenge is to find exogenous variation in access to entertainment technology. The growth in TV time does coincide with an aggregate increase in leisure hours, but correlated trends are not informative about cause and effect. We use the US television roll-out as testing ground. At the time of it's introduction TV signal reached about 100 miles, so locations just inside and just outside of broadcast rings comprise plausible treatment and control groups. Changes in technical license specifications expand and restrict signal in an undirected manner. We further exploit regulatory restrictions and terrain blockages that interfered with TV signals. The regulator, FCC, froze licensing of new stations from 1948 to 1952, so similar cities saw broadcasts start years apart for reasons unrelated to economic conditions. Pre-trends checks and placebo tests show that the interference was orthogonal to labor market conditions and therefore provides plausibly exogenous variation.

A second challenge is to distinguish labor supply responses from changes in labor demand. Many technologies affect both sides of the labor market. Computers and the Internet, for instance, provided new forms of entertainment but also resulted in

[82] The Internet has accelerated video consumption. Streaming services make up over 70 percent of bandwidth use. Gaming, by contrast, accounts for about 4 percent (see Sandvine, 2015).

a skill-biased labor demand shift (Akerman et al. (2013)). Equilibrium employment effects reflect both types of shifts. TV offers a cleaner testing ground for the effect of entertainment technology on labor supply, since it sharply improved entertainment but had little effect on production. The obvious exception is the entertainment sector, where TV led to superstar demand effects for entertainers (Chapter I). Demand shocks in entertainment however do not pose a major threat to identification because the entertainment sector is small relative to the overall economy. An individual employment response to TV access therefore likely reflects a labor supply shift. We use social security records to study such shifts at the individual level.

Related Literature - Key related work falls into four categories. The first studies the determinants of labor force participation. Labor demand factors like trade and technology are widely discussed. Autor et al. (2015), for example, show that trade with China depressed US labor demand, and Acemoglu and Restrepo (2017) attribute reduced employment in part to industrial robots. Much of the literature on the supply side focuses on the effect of taxes and benefits. Abraham and Kearney (2018), in a review of the causes of the decline in the employment-to-population ratio, highlight several possible explanations related to transfer programs, including Social Security Disability Insurance, the Supplemental Nutritional Assistance Program, and the Affordable Care Act. Other supply side issues highlighted in the literature are social norms and the opioid epidemic (see respectively Goldin and Olivetti (2013) andKrueger (2017)). These various forces are most likely not independent. We can test how the effect of social insurance changes with the availability of TV entertainment.

The most closely related paper to ours is Aguiar et al. (2017), who document dramatic and concurrent declines in work hours and increases in video gaming among young men. They then develop a model of labor supply and "innovations in leisure luxuries" and find that improvements in gaming predict a 1.5 to 3.0 percent decline in work hours. In the absence of exogenous variation in gaming quality, these figures rely strongly on modeling assumptions. For that reason, Abraham and Kearney (2018) note that "We do not attempt to assign a magnitude to the possible contribution of improved leisure technology, in particular gaming technology, but call attention to the provocative hypothesis that has been advanced about its possible effects on young men's participation. This is an issue deserving additional attention." One interpretation of our contribution is as a quasi-experimental test of that same mechanism, of TV as a leisure technology capable of shifting labor supply.

Secondly, our findings relate to research on trends in retirement and the determinants of those trends. Today's conception of retirement as an opportunity for "golden years" of leisure is a relatively recent phenomenon. Through the nineteenth century, most men worked as long as they were physically capable. Retirement happened, rather than being chosen. This changed in the twentieth century (for a historic overview see Goldin (1998)). In the US, almost half the men over the age of 64 worked in 1940. That share had halved by 1970. Costa (1998) gives several possible reasons for this shift, including more generous social insurance and greater availability of compelling, low-cost leisure activities like TV. We find that the effect of TV is largest among older people. As theory would predict, workers respond more if they are marginally attached. Moreover, social security improves outside options drastically at age 65. TV is more likely to nudge someone already around this age out of the labor market than someone mid-career.

Third, the paper speaks to the literature on the value of free-to-use services. While TV broadcasters clearly engage in market transactions, much of the utility they provide is not captured by monetary transactions. This problem has been widely associated with freely available internet services, such as Spottify, Youtube or Facebook. A growing literature discusses the implications of such goods for GDP (see for example Syverson (2017); Byrne et al. (2016)). Nordhaus (2006) discusses methods to account for the value of non-marketable goods. A standard approach is to put a monetary value on the time invested in the good. Applications to the value of the Internet have found modest willingness to pay in monetary terms but large expenditure of time (Goolsbee and Klenow (2018)). Accounting for the value of this time vastly increases the valuation of the Internet (Brynjolfsson and Oh (2012)). Critics however point out that only a fraction of the time spend on the Internet would otherwise be spend productively. To value the time expenditure one should distinguish between hours that replaced leisure and hours that replaced labor (Syverson (2017)). We can use our estimates to do exactly that. The labor supply response to the TV will reveal what share of TV hours crowded out labor and how much came from leisure. We can then proceed to compute what share of the value of TV is missing from GDP figures.

The final related literature aims to estimate the causal effects of TV in the US. Notably, Gentzkow (2006) finds that TV caused a decline in voter turnout, and Gentzkow and Shapiro (2008) conclude that contrary to conventional wisdom, TV exposure during childhood led to modest increases in test scores during adolescence. These studies use the 1950s TV roll-out We use similar variation in TV access during that period. The variation was non-marginal, in the sense that in the 1950s many households were watching for several hours a day while others were still years from access.

The rest of this paper is organized as follows. Section 2 introduces our TV access data. We combine newly digitized data on the locations and technical features of TV towers with signal propagation formulas to compute the decibel signal strength and broadcast range of all commercial stations in operation from 1948 to 1960. Section 3 describes our two sources of labor supply data, highlighting their respective strengths and weaknesses. The results follow in section 4. Our baseline regressions, as well as a set of placebo tests, show that TV reduced labor force participation, especially among older workers. Section 5 introduces the welfare analysis and uses our estimates to compute the monetary value of non-monetary transactions. Section 6 concludes with a summary and discussion.

10 Measuring TV Access

10.1 The Irregular Terrain Model

We estimate the signal strength of each station at the geographic center of each US county from 1948 to 1960 using the Irregular Terrain Model (ITM) of Olken (2009).[83] Several recent economics papers have used this approach in different settings. Olken (2009) first used a version of the ITM created by Hufford (2002) in a study of how radio and TV impacted social capital in Indonesia. Enikolopov et al. (2011) used the ITM in a paper on Russian media and voting, DellaVigna et al. (2014) on the role of radio in the Serbo-Croatian conflict, Yanagizawa-Drott (2014) on propaganda and violence in Rwanda, and Durante et al. (2015) on how Silvio Berlusconi's TV network affected Italian politics.

The ITM is a model of signal propagation that computes signal intensity at a receiving location as a function of topography, distance from the tower, and tower specifications. As its name suggests, the ITM accounts for terrain changes between a tower and a viewer. Durante et al. (2015) explain that a "television signal is transmitted over the air according to the laws of physics for electromagnetic propagation. In the free space, signal strength would decrease with the square of the distance from the transmitter, however in reality patterns of decay are much more complex due to difraction caused by mountains and other obstacles." They note also that the "ITM was originally developed by the US government for frequency-planning purposes and allows one to accurately predict signal strength across narrow geographical cells."

10.2 *Television Factbook* Data

We collect three sets of data on broadcasting from early editions of the *Television Factbook*, a trade publication for advertisers and other industry players. First, beginning in 1948, the *Factbook* published the technical characteristics of all commercial stations in operation. We use these as inputs for the ITM. Specifically, for each station in each year from 1948 to 1960, our digitized *Factbook* data include latitude and longitude, height above ground, channel number and frequency, visual and aural power, and other details like call letters and start date. There were 41 stations on air in 1948. Already by 1960, there were 570.[84]

[83] Previous research approximates historical signal reach with contemporary media markets. In Appendix A we present evidence of measurement error in that approximation.

[84] Latitude and longitude are first published in the 1952 *Factbook*. Earlier years give station addresses, which we geocode. The *Factbook* was published four times per year in 1948 and 1949 and twice per year

The second and third groups of data involve secondary extensions of original broadcasts. A town across a mountain range from a nearby city would be cut off from that city's TV signals, and the ITM would correctly measure that town as having no TV access through the air. However, some towns constructed antennas on top of the mountains to capture signals and then wire the broadcasts into the otherwise obstructed homes. This was the birth of cable TV and was known at the time as Community Antenna Television (CATV).[85] We have digitized the *Factbook* directories of CATV locations, start dates, and estimated number of subscribers. Finally, an alternative to piping a signal through a CATV system was to rebroadcast it through the air with small antennas called translators. The *Factbooks* record the locations of licensed translators beginning in 1957, and we have digitized them through 1960.

This allows us to create what is, to our knowledge, the most exhaustive possible measure of TV access during the US roll-out. However, the data do have limitations. For example, while translators served relatively small populations, the *Factbooks* do not list any unlicensed translators that towns or individuals might have erected. We also must always choose a *point* at which to calculate signal strength and then take that point as an estimate of signal strength throughout some cell, like a county. Despite these and other challenges, the *Factbook* data and ITM together allow for precise and flexible measurement of TV access.

10.3 Visualizing the ITM

We map signal strength as a visual check on the ITM's output. We have calculated the strength of each station at each county centroid and, for visualization, mapped only the strongest signal in each county. The units are decibels, which is a relative measure. A strength of zero on this scale is top quality reception, so a strength of -50 means 50 decibels below top quality reception. While there exists no single cutoff below which a station was not available–signals fade, they do not switch on and off–we take -50 as our baseline threshold for the signal quality at which a station was effectively unwatchable.

Figures 20 through 23 show ITM-estimated signal strength in the US for select years of the roll-out, beginning with 1948 in Figure 20. Note that in areas with limited variation in elevation–Dallas and Minneapolis-St. Paul, for example–signals decay slowly, whereas mountains limit the reach of stations based in places like Salt Lake City. Figures 21 and 22 then illustrate the importance of the FCC freeze, which halted licensing of new stations from 1948 to 1952. There was therefore a policy-induced delay in the roll-out and, after the freeze was lifted, a jump in coverage after 1952. Gradual expansion in coverage then proceeded until virtually all of the population had TV access

from 1950 to 1960. We digitize the latest edition available in each year.

[85] In 1966, both the *American Economic Review* and the *Quarterly Journal of Economics* published articles on CATV. See Fisher et al. in the references.

by 1960, shown in Figure 23.

11 Labor Supply Data

We use two sources of data on labor markets. The first is individual-level employment data from Social Security records, the second county-level Census data. Here we briefly discuss each in turn.

11.1 Individual Social Security Records

Our first set of labor supply data come from the 1978 Current Population Survey (CPS)/Social Security Summary Earnings Records (SER) exact match file (ICPSR 9039), which links respondents from the March 1978 CPS to their work histories in Social Security records, including quarters worked for each year from 1937 to 1978. The strengths of this employment data are that it covers the period of the TV roll-out; it is from administrative records, which are more reliable than self-reported data on work; it includes the full set of demographics from the CPS, which we use both as controls and for analysis of heterogenous treatment effects; and it tracks people over time, so we can include individual fixed effects in our regressions to address some potential threats to identification.

The data have three key limitations. First, we do not have intensive margin measures of hours worked. We observe only the number of quarters worked in a year. The second issue concerns place of residence. Ideally, we would know the county in which each respondent lived during each year of the TV roll-out. The match file, however, includes only the Metropolitan Statistical Area (MSA) of the respondent in the 1978 CPS. For analysis, we assume away mobility and assign individuals to their 1978 MSA. We then define the signal strength of a TV channel in an MSA as the population-weighted average of signal strengths across counties in that MSA. Finally, the 1978 CPS sample is not representative of the 1950s population. We aim to address these concerns with a second source of data.

11.2 County-Level Census Data

From the 1950 and 1960 Censuses we obtain county-level labor force participation rates and demographic characteristics. The Census data are representative and allow us to exploit more precise geographic variation in signal strength from the ITM. Obvious drawbacks, relative to the Social Security data, are that the Census data are repeated cross-sections available only every ten years, rather than annual observations of individuals. The definition of work also differs, with the Census asking whether an individual was working or looking for work during a reference week. The complementary differences between the two data sources allow a robustness check on our baseline results, which we present next.

12 Design and Analysis

To estimate the effect of entertainment technology on labor supply we would like to isolate the effect of the technology from confounding factors. The roll-out process of television antennas makes TV signal available to different parts of the US at different times, and we specifically aim to make use of exogenous variation in the availability of television that arises from signal decay and policy restrictions. Our data allow us to compare how work patterns differ between two individuals, one who falls within the signal reach of a local antenna and another who doesn't. We will analyze this variation in a difference-in-difference regression. Areas differ in the time that they first get access to TV and in how many channels are available.

The number of TV channels that can be watched in area a at time t is given by TV_{at} and let labor supply of individual i of gender g in area a at time t be L_{ait}. The difference-in-difference regression is

$$L_{aigt} = \gamma_{tg} + \delta_i + \beta \cdot TV_{at} + \epsilon_{aigt}.$$

The regression controls for individual fixed effects (δ_i) and therefore exploits variation in work patterns within individuals. This fixed effect also absorbs the area fixed effect. The time fixed effects absorb aggregate trends in labor supply. In the post-war period labor force participation did not change much. This however marks substantial heterogeneity by gender. Male labor force participation was declining, while participation among women was growing. To account for these different time trends we allow for separate year fixed effects by gender. Note also that TV stations were not set up randomly across the country. Early stations were mainly set up in urban centers. The fixed effects will control for time invariant differences between the treatment and control regions. These fixed effects will for example absorb that urban residents are more likely to work than rural residents. The key identifying assumption is that variation in TV_{at} is unrelated to simultaneous labor market shocks.

Whether an individual is treated depends on the area of residence a. We treat this as fixed throughout the sample period. Annual data on individual location is not available and would potentially introduce a further problem as the relocation decision might make TV exposure endogenous. Ideally we would like to assign people to a residence that is unaffected by television, for example their place of birth or the residence before the launch of TV. This would allow us to estimate an intent to treat effect. As described above, in our baseline Social Security data we only observe the place of residence in the 1978 CPS, and we therefore assign individuals to their 1978 residence.

While people move regularly, most moves occur within the geographic areas that we analyze. These moves do not affect our results. Among the moves across boundaries, most are unproblematic for this analysis. Our results are only affected if mobility is correlated with changes in both TV and labor supply. In other words, people would have to move towards TV stations at the time when a TV channel begins broadcasts and simultaneously adjust their labor supply. We can also test for the direction and magnitude of the potential bias that mobility could introduce.

12.1 Results: Individual Social Security Records

Our baseline measures of labor supply in the Social Security data are a dummy that indicates whether an individual worked in a year and quarters worked per year. We focus on employment rather than labor force participation because the data does not include periods of unemployment. During the sample period unemployment was low. It averaged below five percent and had no clear trend. In the aggregate labor force participation therefore moved in line with changes in employment, the effect of TV on employment is similar to the effect on labor force participation.

The introduction of a new television channel results in a significant decline in the labor supply. Individuals work about 4 days less, or 0.018 fewer quarters, after the introduction of a new TV station (Table 18, column 1). The average number of quarters worked is 1.8. Labor supply therefore falls by roughly 1 percent. Most of the fall in labor supply comes from the extensive margin. In Panel B we analyze whether a person worked at all in a given year. We regress a dummy that takes the value 1 if a person worked in any quarter of a given year. The introduction of a TV channel reduces employment by 0.5 percentage points. About half of the population aged 16 plus is working in any given year. A TV station therefore reduces employment by roughly 1 percent, similar to the effect on quarters worked. The main effect of television is therefore an exit from employment, consistent with our findings below that older workers are most affected.

Next we address potential shortcomings in the Social Security data and show that our results carry through to alternative specifications. We first check that our results carry through if we exclude the immediate post-war period, which may be an economically unusual period. We find very similar results if we exclude the first four post-war years (column 2). Our data tracks the work history of individuals who were part of the CPS in 1978. The data therefore does not cover people who died or left the US prior to 1978, so we have a selected sample of individuals. By including individual fixed effects we eliminate the potential bias from the selection, but a second form of selection is that the CPS oversamples certain groups of the population. We can address the oversampling by using the weights provided in the 1978 CPS. These weights are designed to yield a representative sample of the US population in 1978. We assign each individual a fixed sample weight throughout time and thus down-weight groups that were oversampled in the original data. The weighted least squares estimates are similar to the baseline results (column 3). This suggests that oversampling in the CPS is not a major concern for our findings. Finally, we run the conventional difference-in-difference regression by replacing the individual fixed effect with a fixed effect for each area a. This is a restricted version of the baseline regression. It is equivalent to an aggregate regression at the time-area level, weighting each observation by the cell size. It therefore does not control for the selection into the sample. The results are very similar to the baseline and suggest that such selection is not a major source of bias (column 4).

The difference-in-difference analysis is credibly causal only if the treatment and control groups have parallel pre-trends. We check pre-trends with a standard specification,

$$L_{ait} = \gamma_{tg} + \delta_i + \sum_{j=-4}^{4} \beta_{t-j} \cdot TV_{at-j} + \epsilon_{ait},$$

which captures the evolution of the treatment effect in the 4 years before and after the launch of a new TV channel. The β coefficients are plotted in Figures 24 and 25. The figures show that treatment and control regions look similar in the years leading up to the launch of a TV channel. The differences are close to zero and show no clear trend. When a TV station begins broadcasts, significant differences between the treatment and control group emerge, with less work in the treatment group in the years that follow. The sharp change around the time of treatment indicates that the difference-in-difference specification is capturing the effects of TV. We can for example rule out that slow moving differences between the treatment and control areas, such as differentially aging populations, is driving our results.

12.2 Results: County-level Census Data

Next we turn to the question whether the results from the Social Security sample is representative for the wider population. We repeat the same regression design with US Census data. To assess the difference between the samples we construct specifications that can be run on both datasets. The smallest unit of observation that is available in both datasets is the MSA average in 1950 and 1960. The comparable data includes 133 MSA regions and 266 observations. We are unable to correct for the difference in sampling and employment definition, but the share of people not working is comparable in the two datasets at about 50 percent. Table 19 compares the effect of TV across the two samples. The results paint a consistent picture. The availability of a TV signal leads to a sizable reduction in employment. The point estimate in the Census data is smaller at 0.15 percent compared to 0.80 percent in the Social Security data, and the precision is greater in the Census because of the larger sample. The difference in the point estimates may be explained by the different sampling. The Census data covers all people, including ones that never work and thus aren't included in the SSA data. The group of never takers is therefore likely larger in the Census, which would explain the smaller treatment effect.

We then make use of the additional geographic variation available in the Census. So far we focused on TV access measured as the average exposure within an MSA. In our next set of results we first keep treatment at the MSA level but control for fixed differences across counties (with about 3,000 fixed effects). This is equivalent to running a first difference regression at the county level. The results are essentially unchanged. With information on the county of residence we can construct a more precise measure of TV exposure. We can look at local variation in TV exposure, even within local labor markets. This will allow us to compare people exposed to the same labor market shocks and same local policies but with differential TV access. The regression at the county level shows that employment declines significantly once a TV channel goes live in a county. The employment rate falls by 0.08 percentage points. By focusing on the local TV variation we gain precision and find a smaller point estimate.

A potential explanation for the smaller estimate is selective mobility. Unlike in the SSA data, the Census does not allow us to link the same individuals over time. We thus do not observe moves across counties. If mobility is systematically related to TV changes such moves could bias the findings. In particular the treatment at the MSA level may be a more precise treatment measure if moves across counties are more responsive to the introduction of TV than moves across MSAs. We can gage the severity of such problems by controlling for changes in the characteristics of the local labor force. If mobility is not selectively related to TV the results should not change much once we control for the composition of the labor force. This test cannot rule out that unobserved characteristics of the local workforce changed. However, to the extent that such unobservable factors correlate with observed characteristics the test will give a sense of the importance of such factors. The results are qualitatively unchanged once we control for changes in the composition of the local labor force (column 5), meaning selective mobility does not pose a serious problem in this setting.

The long-run effect of TV depends on the point at which adding another channel no longer impacts labor supply. The effect may increase with the first few stations as new channels add diversity and improve the quality of entertainment, but the hundredth channel should make less of a difference. Figure 26 shows how the TV treatment effect changes with the number of channels. The effect of the first five channels is nearly linear and the impact diminishes thereafter. From the seventh channel onwards there appears to be no additional effect. A back-of-the envelope calculation then suggests that in the long-run TV reduced the employment share between 1 to 3.5 percentage points. The reduction in working hours in response to computer games from Aguiar et al. (2017) also falls in this range.

12.3 Placebo Tests

A regulatory shock allows us to both sharpen identification and also run placebo tests on our baseline specification. In 1948, with applications pending, the FCC froze licensing of new stations while it revised its spectrum allocation policy. Figure 8 shows the discontinuous drop in FCC permits granted, from about fifteen per quarter to zero. The FCC issued no new licenses from 1948 to 1952. This gives rise to "zombie stations" that had applied to operate and would have gone live but for FCC intervention. The ban affected all pending stations, independent of local characteristics. Figure 4 maps 1950 US TV coverage with three categories: counties that had TV access, counties that did not, and counties that *would have* had access in the absence of the freeze.

We exploit this variation for two additional identification strategies. First, we test if the allocation of TV licenses is correlated with local labor supply trends in ways that our baseline regressions do not control for. This placebo test compares regions without a TV signal to regions that would have had a TV signal without the freeze. Note that none of the counties in this test have TV access. The placebo TV stations should have no effect in our regressions since these stations submitted applications but were not constructed (until after freeze was lifted). Effects of the zombie stations could indicate that our baseline findings are spurious. Panel A in table 20 reports the results of the placebo test. We find no significant changes in labor supply in places where TV was planned but blocked. Column 1 shows results for employment, and column 3 gives results for quarters worked. While the results are noisy because of the smaller sample, they suggest that treatment timing is unrelated to local labor market trends. Relative to the simple pre-trend check, we can now also rule out that trends change sharply around the time of intended treatment. In a second test, we make use of the sharp drop in license approvals and compare places whose station applications fall just to the right of the approval cutoff, and thus got frozen out of broadcasting, to places whose stations were approved. Columns 2 and 4 of Table 20 show that this test–which restricts the control group to stations that applied for licenses but were denied–yields similar results to our baseline estimates. Labor supply declines by about 0.4 percent .[86]

12.4 Terrain Variation

Terrain interference creates further variation in the local exposure to television. Terrain can affect a signal in two ways. Peaks that are in the line of sight between a TV broadcast and a receiving antenna will block the signal, and, conversely, if the broadcast tower is on higher ground the signal will travel further. One advantage of the terrain variation is that terrain is a pre-determined factor, so we can analyze variation in TV access that is independent of human decisions. We leverage the terrain variation by comparing areas

[86] We see similar results with this test in the Census data, albeit the limitation to two years makes such estimates noisy. Results are available upon request.

exposed to TV to places that are blocked because of terrain. Absent elevation variation like mountain ranges, a TV signal reaches all places within a given radius, say 100 miles, of the broadcasting antenna. We call this radius the “line of sight” (LoS) radius. Areas without TV access within the LoS radius are places where terrain blocks signal transmission. If we think of these places as excluded from television for exogenous reasons, they make a good control group. We can analyze the impact of terrain in a difference-in-difference regression. The TV exposure variable is the same as before, but now the control group includes only areas within the LoS radius.

At the MSA level, terrain leads to little variation in TV access. To use more precise local variation we run the regression with the county-level Census data on employment rates. The terrain variation allows for the same two tests that we ran with the freeze variation in Panel A of 20. The first test compares two types of untreated counties– those that were close to a TV tower but had no signal because of terrain, and those that were far away from a TV tower and therefore also had no signal. This test allows us to probe the parallel trends assumption. The results are reported in panel B of table 20. If anything, the point estimates suggest that counties near TV stations trended towards tighter labor markets, which would bias us against detecting a negative effect of TV on labor supply. The second test compares treated counties to counties blocked from TV by terrain. Column 2 shows that people in the county that receives a TV signal increase their leisure consumption relative to neighbors who are blocked from signal. Leisure consumption in treated counties increases 0.18 percentage point relative to counties blocked by terrain. The effect is in line with the baseline county estimates, with the terrain variation yielding a slightly larger treatment effect than the baseline range of 0.08 to 0.16.[87]

12.5 Heterogeneous Effects: The Role of Retirement

The leisure response is largest among workers close to retirment age. We repeat the baseline regression allowing treatment effects to differ for older workers. The regression controls for improving health trends by comparing individuals of the same age with differential access to TV. Access to TV leads to lower participation among all age groups, but the magnitude of the effects differ widely. For workers under the age of 60 the reduction in labor supply is about 0.5 percentage points, while for workers over 60 the effect is almost three times as big at around 1.6 percentage points (see column 1 of table 21). Figure 29 shows the effects across the age distribution. There are modest effects across all age groups, but the impact on those near retirement is much larger. Above the age of 60, the effect increases somewhat and reaches about 2 percentage points for people at retirement age.

[87] Controlling for the composition of the local workforce does not alter the identified effects (not reported). Moves across counties therefore do not appear to drive these effects.

Table 21 reports heterogenous effects for other standard sub-groups. Note that we do not find heterogenous effects among people who are more mobile. This is further evidence that mobility does not seriously bias our results. TV has close to no effect for African Americans, which is consistent with the lower TV ownership rates among blacks during the roll-out. The interaction term for women is marginally significant, and its sign and magnitude suggest that our baseline results are driven by men. Given that much daytime programming in the 1950s was targeted towards women, this is somewhat surprising and deserves further study. On the other hand, if TV primarily affected the retirement decision, and relatively few women were working in the 1950s, then we would expect to see men driving the results.

13 Welfare Estimates

To put our estimates in context we study the monetary value of TV by calculating the willingness to pay. In line with the literature on the value on non-market based goods we use a revealed preferences approach to identify the underlying willingness to pay for TV. We do not aim to conduct a full welfare analyses and do not account for possible social externalities or time-inconsistencies in preferences. A useful benchmark is to think about the change in wages that would lead to an equivalent allocation. To compute this we need an estimate for the response of employment to wages. Estimates of the extensive margin Frisch elasticity suggest that employment between 3 and 7 percent in response to a 10 percent decrease in wages (see Chetty et al. (2013), Gourio and Noual (2009), Mustre-del-Rio (2015), and Park (2017)). The effect of television is therefore equivalent to a 5 percent decline in wages.[88]

A related question is how much of the value of TV is reflected in monetary transactions. The rise of free apps and online services has made this issue salient recently. TV had a similar flavor already 70 years ago. People have two types of resources they can invest, time and money. The expenditure share of TV is small and dwarfed by the share of time expenditure. In 1965, the first year we have access to time-use data, Americans spend about 7.5% of their time watching TV, while the income expenditure share was less than 1% (see panel A in Table 22).[89] Measuring the value of TV by spending on TV is therefore missing part of the value created by TV. Time investment leads to forgone earnings and adds to implicit payment for TV. However, not all time spent watching TV replaces productive activity, in fact we would think that TV mostly crowds out other leisure activities.

We can use the labor supply estimates derived above to compute in what proportion TV displaced work and non-work activities. The Census results estimate the effects for a

[88] This is a conservative estimate since the Frisch elasticity holds wealth constant. A smaller elasticity implies that TV is equivalent to a bigger swing in wages.

[89] TV station revenues are 0.27% and consumer expenditure on TV is 0.51% of GDP.

representative sample of the US population, we therefore focus on those estimates. The estimated employment loss from a TV channel ranges from 0.075 to 0.176 percentage points. As baseline we take the mid range estimate of 0.126. This implies a steady state reduction of employment by 1.26%.[90] The change can explain an expansion in the time share spent on TV of 0.3 percentage points (column 1 of table 22), only a fraction of the total time share of TV. [91] The estimate does not capture an intensive margin response, however employment was mainly a binary choice in the middle of the 20th century, part-time or part-year positions were rare. In line with this our estimates on quarters work and retirement indicate that most response is at the extensive margin. Summing all this up, our results imply that about 4% of TV time crowded out work, over 90% of the time spend with TV replaced other activities.[92]

To assess implications for GDP measurement we would like to quantify the value of time expenditure in monetary terms. This requires additional assumptions. In particular, we need to assume what an hour of time is worth. Previous studies could not distinguish between productive and unproductive hours and attached all hours the value of the market wage. However, the preferred outside option to an hour of TV is only rarely an hour of work. We can distinguish between hours that replaced work and ones that don't. A natural starting point for the value of an hours work is the average market wage. This assumes that compliers are representative for the US working population.[93] Our estimates suggest that this is a conservative assumption, many of the compliers are experienced workers who likely earn more. We can compute how much labor income people are giving up to watch TV. The forgone work hours are worth \$8.5 billion or 1.18% of GDP (see table 22). The non-monetary payment for TV entertainment is thus sizable. It exceeds the monetary spending on TV by nearly 50%. This highlights the importance of accounting for non-monetary transactions. Accounting for both time and money spending implies that the total willingness to pay for TV is 1.9% of GDP.

So far we implicitly valued an hour of non-work activity at 0. This is standard in the literature, yet not all non-work activities are unproductive. TV potentially crowds out home production. We allow for this in two ways. First we compute the value of a non-market hours using the value of home production reported in the Consumer

[90]Our results indicate that steady state is reached once the number of stations reaches 10 (see Figure 26), we assume that by 1965 this level of coverage is achieved everywhere.

[91]We use data from McGrattan, Rogerson (2004) on hours worked to compute the additional time available for TV watching. They use the Census data as us to compute hours. We impute the value for 1965 as a linear approximation between the values for 1960 (40.24) and 1970 (38.83).

[92]A caveat of these results is that we use variation in TV in the 1950s and apply them to the 1960s. We would understate the effect of TV on work if quality improved. To assess the sensitivity of the results, we vary the TV effect estimate. Using the upper end of plausible effects we find that 22% of TV hours replaced work activity. Still over 3/4 of TV time replaced other leisure activities.

[93]Note that a lumpy employment choice implies that not all time won by withdrawing from work is necessarily spent watching TV. For the welfare analysis we want the money value of what a person is giving up and thus account for the full wage.

Expenditure Survey. The survey focuses on a subset of home production areas, value related to home food preparation, and thus provides a conservative estimate for the productivity of an average non-work hour. We complement the approach with a second, more optimistic view on the value of non-work time. He we compute the share of all non-work time spent on home production.[94] Roughly 10% of nonmarket time is spend on productive activities. To construct a scenario that gives an upper bound we make generous assumptions about the value. First, we value home production at the market wage. This approach likely overstates the value of non-market time as many of the tasks performed at home can be bought for less in the market. Second, we assume that TV is just as likely to replace productive activities, such as dish-washing and maintenance work, as it is to replace leisure activities. This scenario will thus give an upper bound estimate for the value of nonmarket hours displaced by TV. The implied range of values goes from \$20 million in the pessimistic case to \$ 4.9 billion in the optimistic case. This increases the estimate for the value of TV by up to 0.78 percentage points of GDP (Panel B of Table 22). Non-market transactions thus capture between 60% - 70% of the value of TV. Decades before the rise of the Internet free goods therefore already play a major role. This has important implications for concerns about GDP accounting in the digital economy. The digital economy only leads to measurement error in GDP figures if free products replace previous market activity. To the extend that the Internet and mobile phone apps are a substitute for TV, the un-measured component of the economy may not change as much as feared.

14 Conclusion

Our findings suggest that entertainment technology led to an economically significant shift in labor supply. Access to an additional TV channel reduces employment rates between 0.1 and 0.5 percentage points. This result is robust across three different sources of variation arising from regulations and terrain that halted signal transmission. The finding also holds across data sets and can be measured both within individual workers over time and also at the local labor market level. Most of the effect comes from workers close to retirement. This confirms one of the key hypotheses of retirement scholars that the wider availability of leisure activities played an important role in the rise of a "leisured pensioner" class. Many older workers withdrew earlier from the labor force when TV brought low-cost, low-impact entertainment to the home.

We also document that most of the value of TV is not captured in GDP figures. Using a revealed preference exercise we document that people are willing to give up sizable amounts of labor income. Accounting for the value of time investment drastically increases the estimated value of TV. Monetary expenditure on TV sets and

[94] We use the data reported in Aguiar & Hurst (2007) and use figure in Table II for core nonmarket work hours.

advertisement capture less than 20% of the value of TV. This finding has important implications for the current debate on missing GDP effects of free to use digital services like Facebook, YouTube or Instagram. To the extend that such new forms of entertainment replace TV, they may only introduce modest additional measurement error in GDP and productivity.

15 Appendix II: Labor Supply and Innovation in Entertainment

15.A Tables

Table 18: Effect of TV on Labor Supply - Individual Level

	Labor supply				
	(1)	(2)	(3)	(4)	(5)
	Panel A: % Not Working				
TV channels	0.634***	0.593***	0.714***	0.445**	0.483***
	(0.133)	(0.129)	(0.167)	(0.164)	(0.128)
	Panel B: Quarters Not Worked				
TV channels	0.022***	0.020***	0.025***	0.017*	0.015**
	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)
Years	1946-1960	1946-1960	1946-1960	1949-1960	1946-1960
Year-sex FE, MSA FE	Yes	Yes	Yes	Yes	Yes
Individual FE	-	Yes	Yes	Yes	Yes
Year-sex-region FE	-	-	Yes	-	-
CPS weights	-	-	-	-	Yes

Notes: Employment data comes from Social Security records and variables are defined in the text. The specifications are based on 565,189 and 566,614 observations respectively. TV channels are the number of channels available on average in the MSA, windzorized at 12 channels. The average percent of people not working is 49 (in column 4 and 5 it is 46 and 48 respectively) and the average number of quarters worked is 2.2 (in column 4 it is 2.1). Standard errors are clustered at the level of a metropolitan statistical area (134 cluster). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 19: Effect of TV on Labor Supply - County Level

	% Not Working (Δ 1960-1950)				
	(1)	(2)	(3)	(4)	(5)
TV channels	0.796*	0.149*	0.163*	0.081*	0.075*
	(0.311)	(0.074)	(0.072)	(0.034)	(0.036)
Data source	SSA	Census	Census	Census	Census
Mean % not working	47	49	48	48	48
Treatment level	MSA	MSA	MSA	County	County
MSA FE	Yes	Yes	-	-	-
County FE			Yes	Yes	Yes
Demographics					Yes

Notes: Data covers two years: 1950 and 1960. Treatment level defines the area for which TV access is measured. At the county level we take signal at the county centroid. At the MSA level we take the population weighted average over county measures. Standard errors are clustered at the level of the treatment. There are 133 MSAs (Census has no record for DC) and 3,078 counties that we can assign to a unique MSA. Demographics are percent with high school degree, percent urban population, median age. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 20: Placebo Test with Blocked TV Channels

	% Not Working		Quarters Not Worked	
	(1)	(2)	(3)	(4)
	Panel A: Freeze Variation (MSA level - SSA)			
Blocked TV channels	0.14		-0.021	
	(1.87)		(0.066)	
TV channels		0.40**		0.014**
		(0.14)		(0.006)
	Panel B: Terrain Variation (county level - Census)			
Blocked TV channels	-0.143		-	
	(0.293)			
TV channels		0.176***		-
		(0.041)		

Notes: Panel A: SSA individual data for freeze years 1946-1953 in column 1 and 3 and full sample in column 2 and 4. Panel B: Census county data 1950, 1960. Observation are weighted by cell-size. Control group are places without TV access in column 1 and 3 and places with blocked TV signal in column 2 and 4 (in Panel A blocked by regulator, in Panel B by terrain). Information on quarters worked is not available at the county level. Standard errors are clustered at the treatment level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 21: Heterogeneous Effects

	% Not Working				
	(1)	(2)	(3)	(4)	(5)
TV channels	0.533***	0.619***	0.558***	0.410***	0.641***
	(0.127)	(0.173)	(0.135)	(0.118)	(0.130)
TV channels	1.126***				
× elderly	(0.140)				
TV channels		0.025			
× mobile individual		(0.119)			
TV channels			0.077		
× high school drop-out			(0.054)		
TV channels				-0.400*	
× women				(0.198)	
TV channels					-0.616***
× black					(0.097)

Notes: Data and specification are as in Table 18 column 2. The elderly are defined as age>59 and mobile individuals are people who report to have moved between 1975 and 1976. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

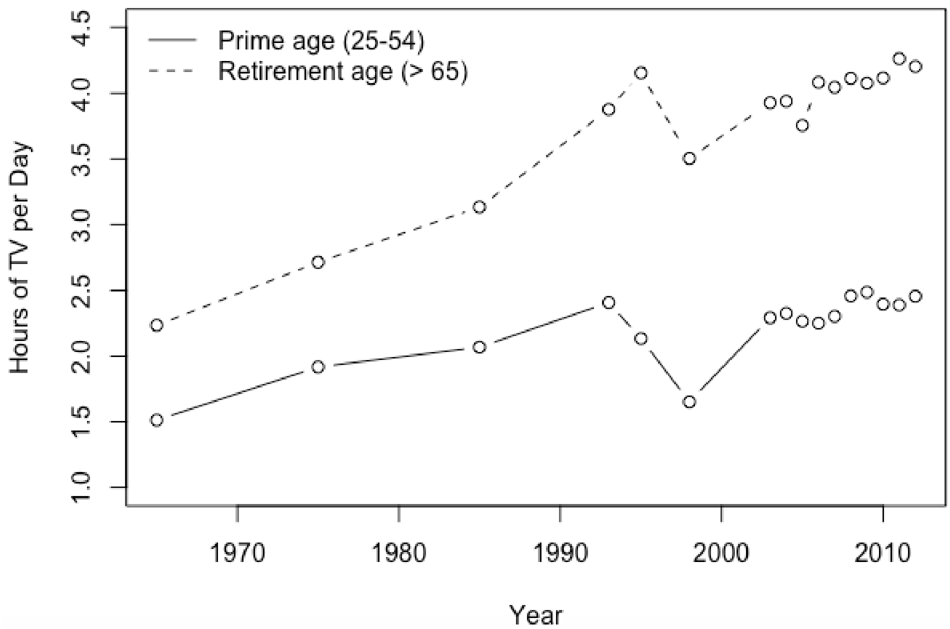
Table 22: Willingness to Pay for TV

	(1) % of time	(2) Value (bio. $)	(3) Value (% of GDP)
		Panel A: Money Spent on TV	
Total	-	5.6	0.79%
- TV Purchase	-	3.7	0.51%
- TV Station Revenue	-	2.0	0.27%
		Panel B: Time Spent with TV	
Total	7.5%	8.50-14.05	1.19% - 1.96%
- From Work	0.3 %	8.48	1.18%
- From Other Activity	7.2%	-	-
of this			
i) productive time	-	0.02	0.00%
ii) productive time (upper bound)	-	5.57	0.78%
Total Value of TV		**14.1 - 19.65**	**1.98% - 2.73%**

Notes: Monetary Expenditure is a share of GDP. Time expenditure is a share of total time. Data on TV set expenditure and hourly value of home production comes from the 1960 CE survey, deflated to 1965 prices. Revenues of TV stations comes from the 1965 “Statistical Abstract of the United States” table 978. Aggregate economic data from the St. Louis Federal Reserve. See the text for details on the computation

15.B Figures

Figure 19: Average Hours of TV per Day in the US



Notes: Data are from the Historic American Time Use Study (H-ATUS). The hours refer to "primary activity." The dip in the 1990s may reflect oversampling of parents. Nielsen reports higher utilization rates, average hours are 5.5 in 1960 and over 4 hours in 1950, but the data do not distinguish primary and secondary activities.

Figure 20: 1948 ITM Signal Strength

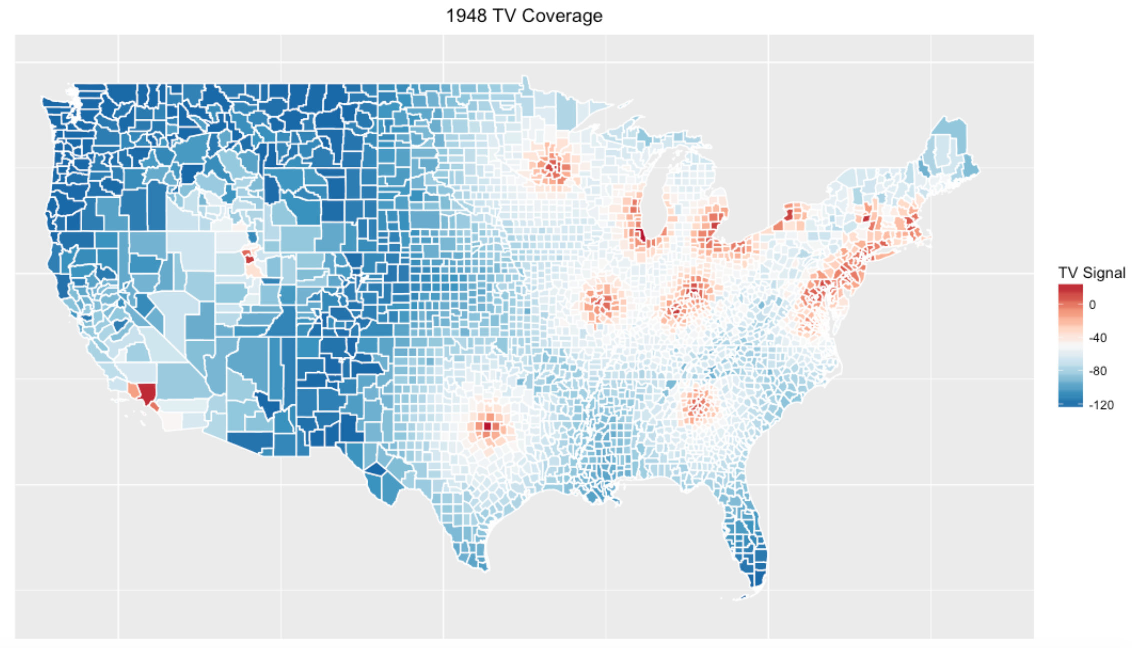


Figure 21: 1952 ITM Signal Strength

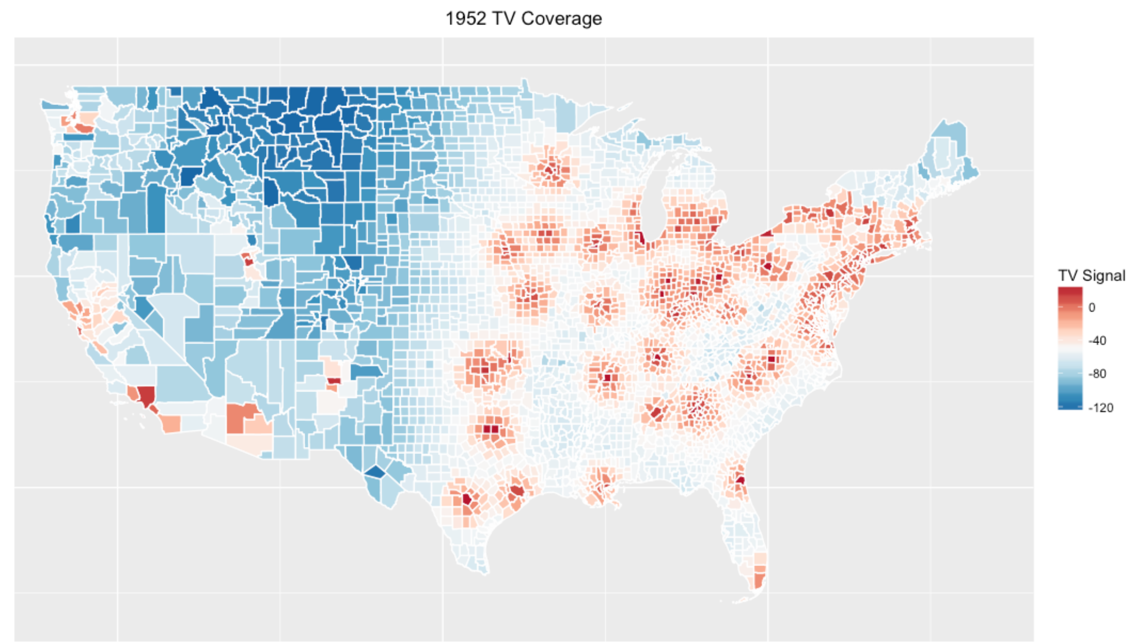


Figure 22: 1954 ITM Signal Strength

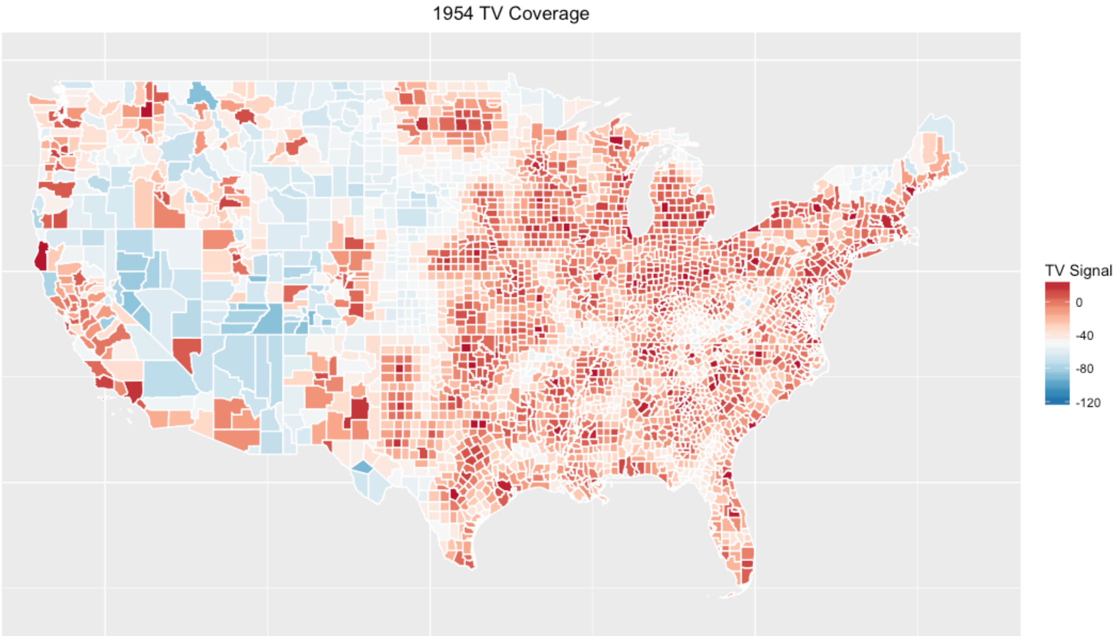


Figure 23: 1960 ITM Signal Strength

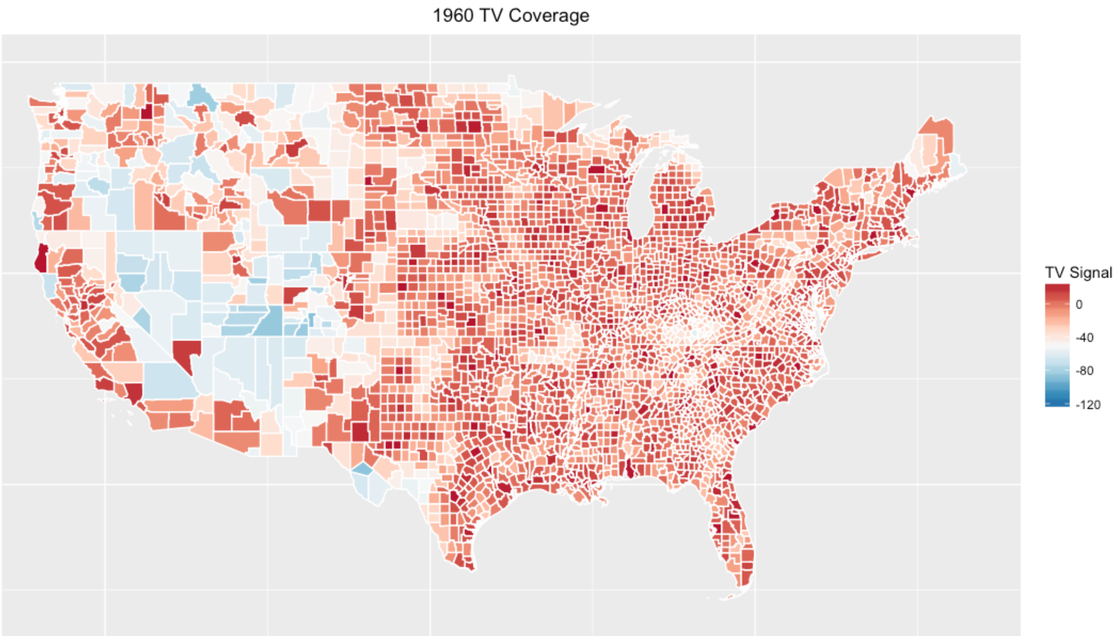


Figure 24: Effect of TV on Leisure: Percent Not Working

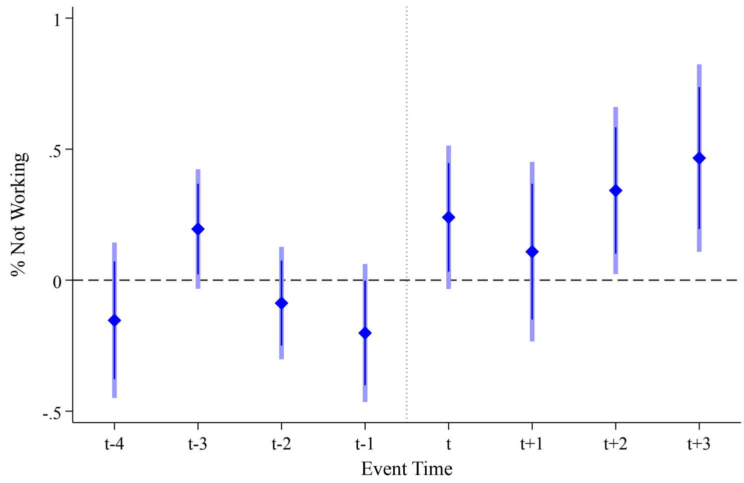
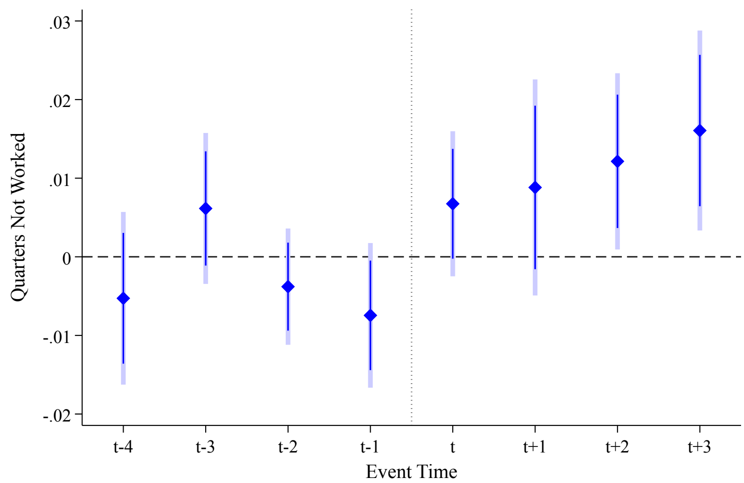
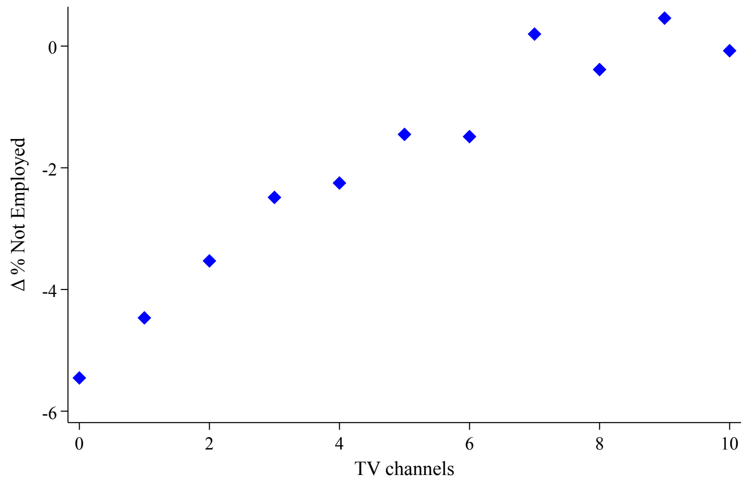


Figure 25: Effect of TV on Leisure: Quarters Worked



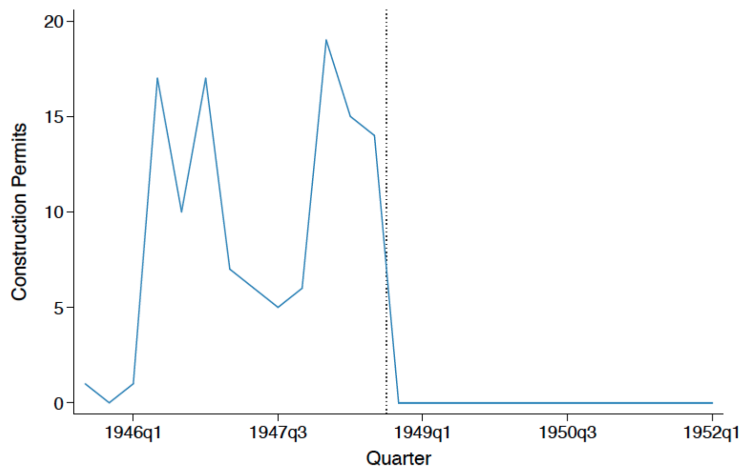
Notes: The figures plot changes in individual working careers around the time of TV channel launches. Coefficients are based on a dynamic difference in difference regression of respectively non-employment rate and quarters without work on TV channels, all regressions control for individual fixed effects. 95 and 99 percent confidence intervals are reported. Employment records come from social security records.

Figure 26: Leisure Response to more TV Channels



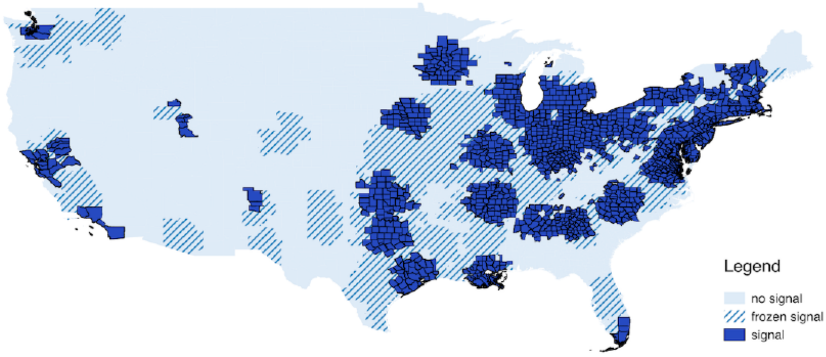
Notes: Figure shows change in non-employment rate between 1950 and 1960 at the county level by access to TV. TV channels measures the number of TV channels that can be watched in the county. Counties with more than 10 channels are grouped into the final bin.

Figure 27: Number of TV Station Construction Permits Issued



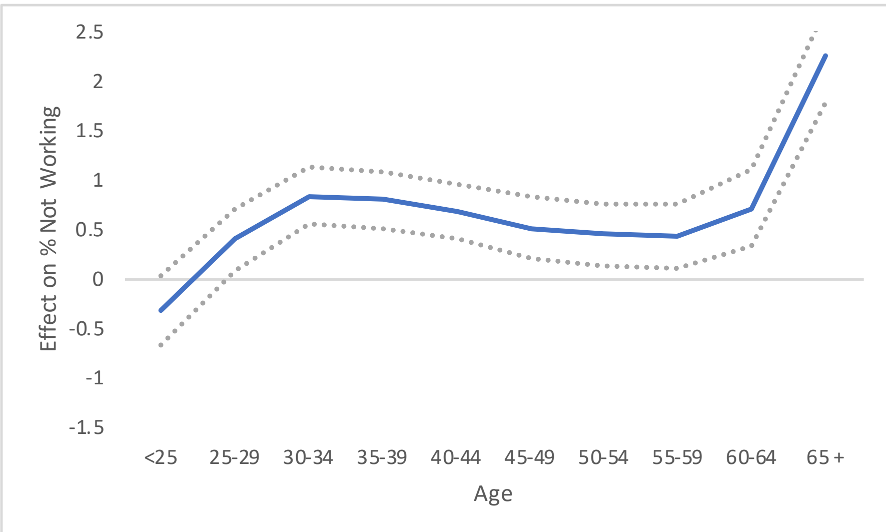
Notes: Construction permits issued by the FCC per quarter. When date of CP is unavailable, date is inferred based on station start date.

Figure 28: TV Signal of Licensed and Frozen Stations



Notes: 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 percent of the time at 90 percent of receivers at the county centroid.

Figure 29: Treatment Effect by Age Groups



Notes: Figure shows the effect of TV for different age groups and the 99 percent confidence interval.

15.C The DMA Approximation

Gentzkow (2006) and Gentzkow and Shapiro (2008b) approximate 1950s broadcast ranges with Nielsen media markets, or Designated Market Areas (DMAs), that are based on 2003 viewership. A DMA is a group of counties around a metropolitan area. The approximation takes the year in which the first station in a DMA began operation and assumes that each county in that DMA received a signal in that year. We find that 1960s coverage maps show differences between historical broadcast ranges and the 2003 DMAs. The DMA approximation sometimes underestimates and sometimes overestimates how far signals reached. The next two subsections give examples of each case. These are not representative, as we chose them specifically for exposition of the two types of problems with the DMA approximation.

An Example of DMA Underestimation

Proximal cities confound the DMA approximation of TV access. For example, panel (A) of figure 30 shows a coverage map of Kansas City from the 1967 *TV Factbook*. The blue line is the broadcast ring as defined by those counties that have over 50 percent coverage according to the map. Panel (B) overlays in red the Kansas City DMA. The DMA is too small–it excludes counties to the northwest that were likely covered. For a region with little variation in terrain, the irregular shape of the DMA also suggests that it cannot reflect the roughly circular true broadcast range.[95]

Let TVYEAR_i denote the year in which county i first had TV access. In panel (B), the DMA approximation assigns the highlighted counties between the two rings a TVYEAR of 1954. However, those counties fall well within the range of the Kansas City tower, and that tower started broadcasting in 1950. Therefore the true TVYEAR of the highlighted counties is likely 1950, not 1954. This misclassification owes to the nearby DMAs, Topeka and St. Joseph, whose broadcasts began in 1954. While it is true *today* that the highlighted counties are closest to the Topeka and St. Joseph signals, and are therefore not in the 2003 Kansas City DMA, those counties are close enough to Kansas City to have viewed Kansas City broadcasts in 1950.

The TV ownership data from Gentzkow and Shapiro (2008a) suggest that this is a case in which today's DMAs do not align with 1950s signals. The DMA data assign the highlighted counties in panel (B) as not receiving a TV signal until 1954, four years after the counties in the red Kansas City ring. If that were true, we ought to observe

[95] For two reasons, the *Factbook* maps ought to be taken only as suggestive regarding true 1950s signal reach. The first is that these maps were not published until the 1960s, and tower technology–power, height, etc.–improved substantially over time. The second is that the shading in the maps reflects surveys of viewership, not measures of signal strength. County coverage exceeding 50 percent for a station means that over 50 percent of households in the county watched that channel. Our measurement of signal reach will not rely on these maps.

the highlighted counties buying TVs well after the Kansas City counties. Panel (A) of figure 31 shows that in fact the timing of TV purchases is almost identical across the two groups, consistent with the hypothesis that Topeka and St. Joseph viewers received a 1950 signal from Kansas City. Substantial TV ownership in a county before that county's DMA-approximated TVYEAR is evidence of measurement error arising from signal overlap.

When signals overlap like this, DMAs underestimate coverage. The overlap between Kansas City and Topeka, for example, leads the DMA data to underestimate how many counties the Kansas City broadcast reached in the 1950s. Spot-checking coverage maps suggests that DMAs can also overestimate coverage.

An Example of DMA Overestimation

Today's DMAs sometimes extend further from city centers than historical signals did. Panel (C) of figure 30 shows a *Factbook* coverage map of Minneapolis-St. Paul. The blue line rings counties whose coverage exceeded 50 percent. Panel (D) adds the Minneapolis-St. Paul DMA in red. That DMA is too large, in that it includes the highlighted counties that were likely out of reach of the broadcast, which leads to overestimation of coverage. The highlighted counties have a DMA TVYEAR of 1948, since that is when the first Minneapolis station began operation. But many of those counties appear to be too far away from the tower to receive the early Minneapolis signals. Panel (B) of figure 31 shows that TV purchases in the highlighted counties–the group inside the DMA but outside the mapped broadcast range–lagged purchases in the counties inside the *Factbook* coverage area, consistent with the hypothesis that the DMA overestimates 1950s signal reach. That pattern remains after controlling for county characteristics like income and population that are associated with TV ownership.

Causes and Prevalence of Measurement Error

This section moves beyond examples to the causes of measurement error and evidence on the prevalence of those causes. To start with underestimation, the two conditions under which the signal overlap problem arises are: Neighboring DMA towers (1) are close enough for signals to overlap and (2) started broadcasts in different years[96]. The closer the towers and the further apart the initial broadcast years, the larger the potential measurement error. To find possible areas of overlap, we ranked pairs of DMAs by their distance apart. There are 166 unique pairs of DMAs whose towers are less than 100 miles apart (a typical broadcast radius) with broadcasts beginning in different years.

[96] Condition (2) is necessary because if two towers were close but started broadcasts in the same year, then all surrounding counties would get a signal in the same year, so proximity alone would not lead to misclassification. Terrain also matters–mountains could prevent overlap–and our measurement of TV access will account for variation in elevation.

Table 23 lists the first 40. Among them are the Kansas City, Topeka, and St. Joseph pairs. Other metropolitan areas such as Pittsburgh and Cleveland are close enough to smaller neighboring stations like Youngstown to create the same overlap issue.[97]

Overestimation, by contrast, can arise because of improvements in TV towers over time. In most cities, the 1950s saw expanded broadcast ranges through both upgrades to existing stations and also construction of new towers. The 2003 DMAs are therefore prone to overstate early 1950s signal reach, when towers were weaker. Figure 32 charts changes in height and power. The average height above ground of a commercial tower in 1948 was 483 feet, and already by 1960 that had increased to 629 feet. Some stations moved to higher ground, and tower height above average surrounding terrain rose from 721 to 992 feet. Average visual power jumped from 19 to 170 kilowatts over that period, and average aurul power increased from 11 to 87 kilowatts. While power does not map directly to broadcast reach, as higher frequency channels require more power to operate, the fixed DMAs do not capture whatever shifts in broadcast areas the tower upgrades did create.

More directly assessing the prevalence and magnitude of measurement error would require first knowing actual broadcast areas. We take the evidence presented here as reason to develop a more precise measure of TV access with the ITM. Figure 33 maps ITM-estimated signal strength for the Kansas City and Minneapolis-St. Paul regions. Panel (A) shows signal strength in the counties around Kansas City in 1950, the year the first tower there went up. In Panel (B), we add the Kansas City DMA, which confirms that the DMA excludes counties to the northwest of the city that in fact had TV access. Similarly, in Panel (C) we map coverage around Minneapolis-St. Paul when broadcasts began there in 1948. Panel (D) shows that this DMA includes a group of counties to the northwest that were out of range.

[97] Note also that in 1948 the FCC froze applications for new broadcast licenses in part because it realized it had allowed stations to be too close together.

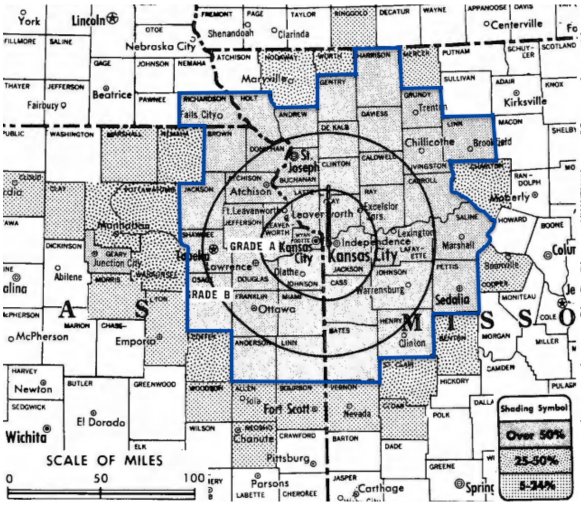
15.D Additional Tables and Figures

Table 23: Proximal Market Areas

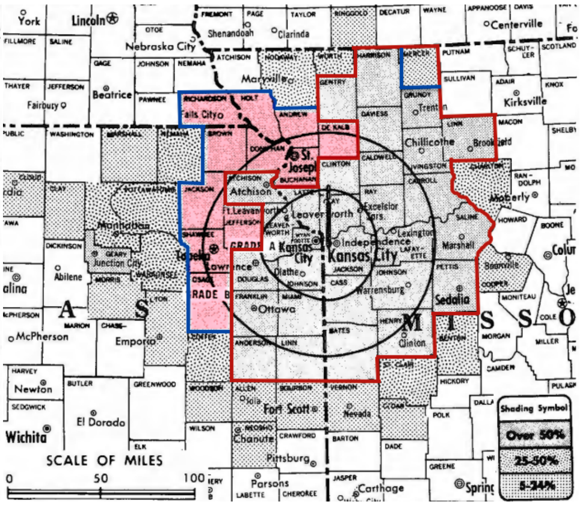
DMA 1	DMA 2	Miles Apart	Years Apart
Pittsburgh (PA) [1949]	Steubenville (OH) [1954]	32.79	5
Washington (DC) [1946]	Harrisburg (PA) [1949]	35.86	3
Harrisonburg (VA) [1954]	Charlottesville (VA) [1960]	36.04	6
Harrisburg (PA) [1949]	Johnstown (PA) [1950]	42.47	1
Cleveland (OH) [1948]	Youngstown (OH) [1953]	42.53	5
Grand Rapids (MI) [1949]	Lansing (MI) [1950]	45.46	1
Binghamton (NY) [1950]	Elmira (NY) [1953]	45.67	3
Syracuse (NY) [1949]	Utica (NY) [1950]	46.36	1
Kansas City (MO) [1950]	St. Joseph (MO) [1954]	48.35	4
Cincinnati (OH) [1948]	Dayton (OH) [1949]	48.48	1
Lake Charles (LA) [1954]	Beaumont (TX) [1955]	49.55	1
Youngstown (OH) [1953]	Steubenville (OH) [1954]	50.28	1
Columbus (OH) [1949]	Zanesville (OH) [1953]	52.28	4
Binghamton (NY) [1950]	Wilkes Barre (PA) [1953]	52.39	3
Zanesville (OH) [1953]	Parkersburg (WV) [1954]	52.44	1
Cleveland (OH) [1948]	Steubenville (OH) [1954]	52.49	6
Detroit (MI) [1947]	Toledo (OH) [1948]	53.08	1
San Francisco (CA) [1949]	Sacremento (CA) [1954]	54.15	5
Baton Rouge (LA) [1953]	Lafayette (LA) [1955]	54.94	2
Pittsburgh (PA) [1949]	Youngstown (OH) [1953]	57.01	4
Hartford (CT) [1948]	Springfield (MA) [1953]	57.39	5
Nashville (TN) [1951]	Bowling Green (KY) [1960]	58.19	9
Grand Rapids (MI) [1949]	South Bend (IN) [1953]	58.36	4
Indianapolis (IN) [1949]	Lafayette (IN) [1953]	58.74	4
Lima (OH) [1953]	Ft. Wayne (IN) [1954]	58.86	1
Kansas City (MO) [1950]	Topeka (KS) [1954]	59.70	4
South Bend (IN) [1953]	Ft. Wayne (IN) [1954]	60.10	1
Birmingham (AL) [1949]	Montgomery (AL) [1953]	60.13	4
Memphis (TN) [1949]	Jonesboro (AR) [1960]	60.48	11
Jacksonville (FL) [1950]	Gainesville (FL) [1960]	61.83	10
Roanoke (VA) [1953]	Charlottesville (VA) [1960]	62.10	7
Denver (CO) [1952]	Colorado Springs (CO) [1953]	63.65	1
Rochester (MN) [1953]	La Crosse (WI) [1954]	63.69	1
Richmond (VA) [1948]	Norkfolk (VA) [1950]	63.88	2
Washington (DC) [1946]	Baltimore (MD) [1948]	63.95	2
Champaign (IL) [1953]	Terre Haute (IN) [1954]	64.67	1
Syracuse (NY) [1949]	Watertown (NY) [1955]	65.18	6

Notes: In brackets is the year in which a broadcast began in each DMA. Some DMAs are abbreviated for brevity. For example, the Birmingham (AL) - Anniston (AL) - Tuscaloosa (AL) DMA is listed just as Birmingham (AL).

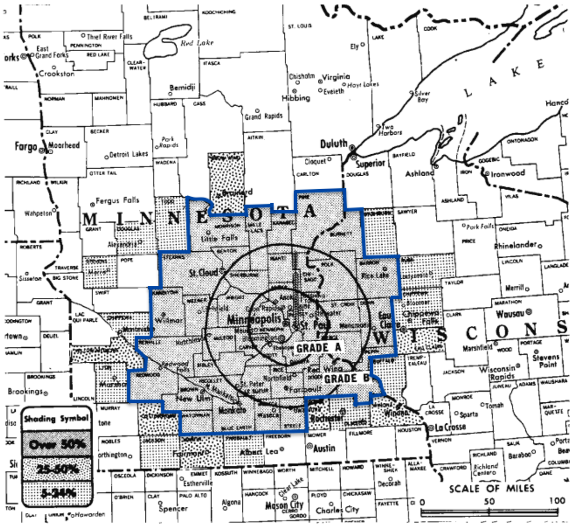
Figure 30: Coverage Maps and Designated Market Areas



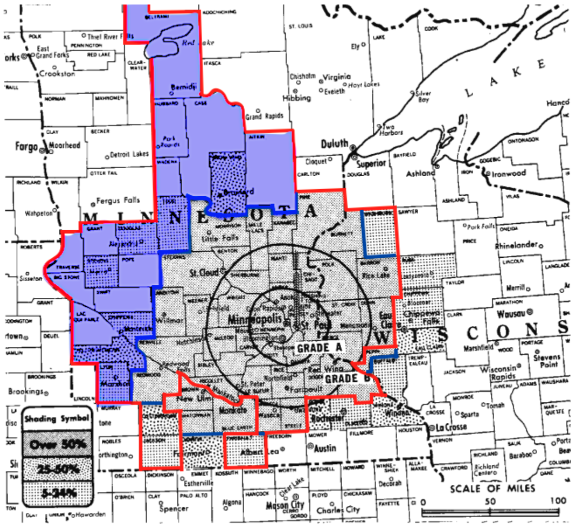
(A) Kansas City coverage map ring (in blue)



(B) Kansas City DMA ring (in red)

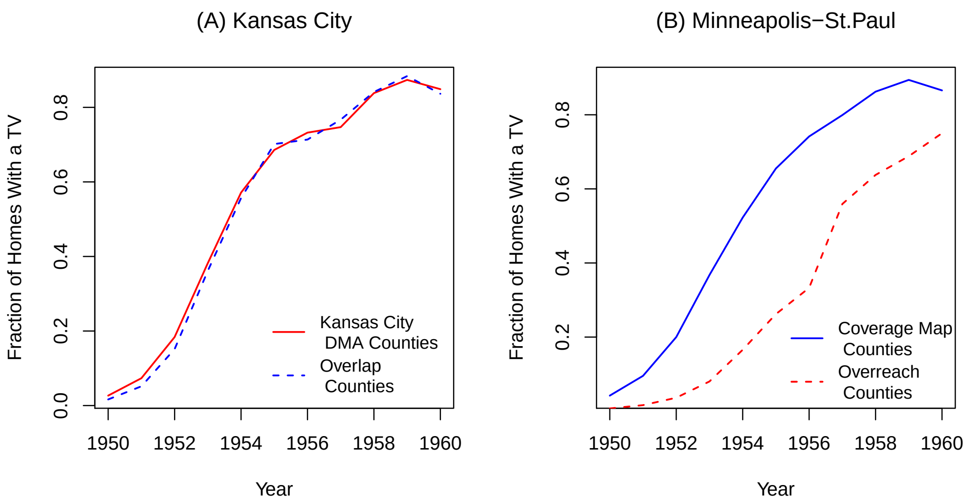


(C) Minneapolis coverage map ring (in blue)



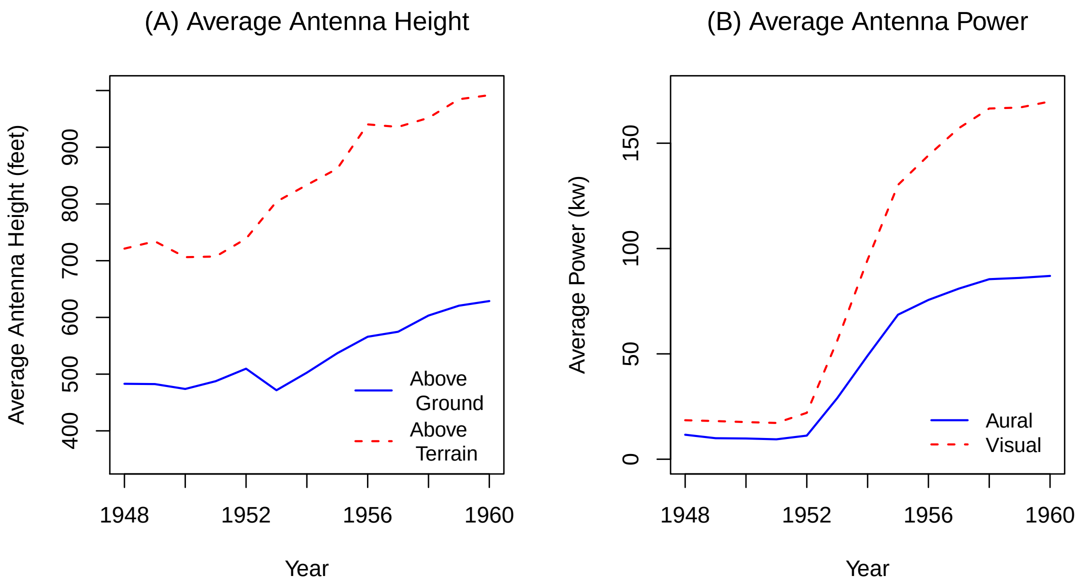
(D) Minneapolis DMA ring (in red)

Figure 31: TV Purchases Patterns



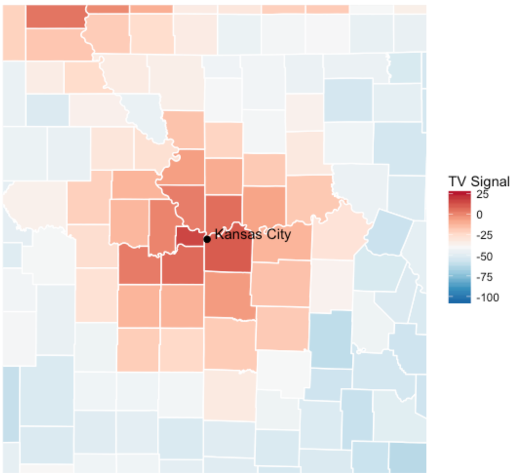
Notes: Panel (A) shows average TV ownership around Kansas City for counties in the groups indicated in the legend. "Overlap Counties" refers to those highlighted in Panel (B) of Figure 1. In Panel (B), for Minneapolis-St. Paul, "Coverage Map Counties" refers to those ringed in Panel (C) of Figure 1, whose coverage exceeds 50 percent according to *TV Factbook* coverage maps. "Overreach Counties" refers to those highlighted in Panel (D) of Figure 1, which fall inside the Minneapolis-St. Paul DMA but outside the *TV Factbook* broadcast range.

Figure 32: Broadcast Tower Improvements

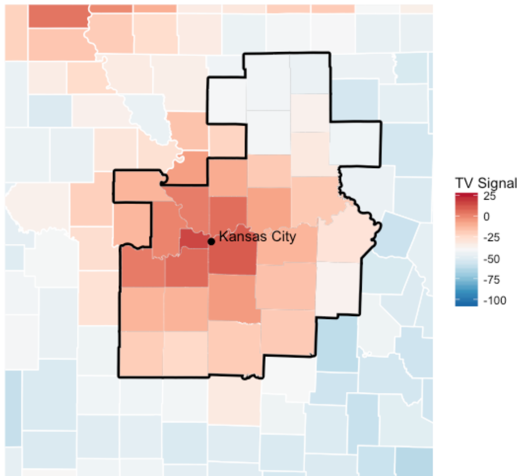


Notes: The figures show *TV Factbook* data on average tower height and power for all commercial stations in operation in each year from 1948 to 1960.

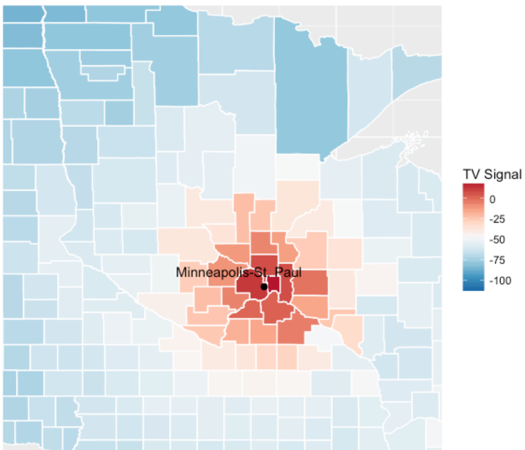
Figure 33: ITM Signal Strength, Kansas City and Minneapolis-St. Paul



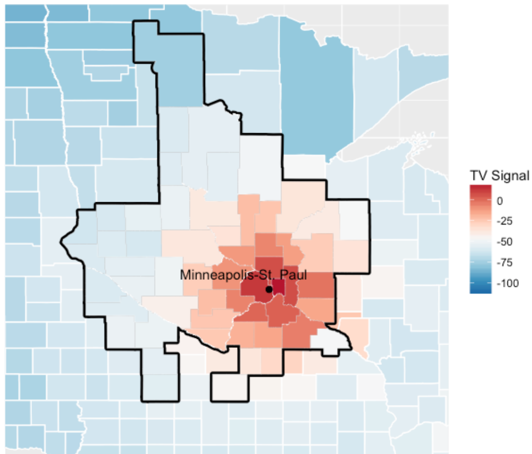
(A) 1950 Kansas City ITM signal strength



(B) Kansas City DMA (in black)



(C) 1948 Minneapolis ITM signal strength



(D) Minneapolis DMA (in black)

Part III
Reservation Wages and the Wage Flexibility Puzzle

16 Introduction

The currently dominant model of equilibrium unemployment – the search and matching framework developed by Diamond, Mortensen and Pissarides (see, Pissarides (2000), for an overview) – offers valuable insights into labour market dynamics. However, the canonical version of the DMP model struggles to quantitatively match the relatively large unemployment fluctuations and mild cyclicality of wages. This point was highlighted by Shimer (2005), who noted that the canonical model is unable to deliver the observed unemployment volatility in response to productivity shocks of plausible magnitudes. A rich strand of work has addressed the ensuing "Shimer" or "unemployment volatility puzzle" by emphasizing the role of wage rigidity in accounting for the volatility of unemployment and job vacancies. As noted by Hall and Milgrom (2008), in a large class of models with job search frictions "wage stickiness is the sole determinant of unemployment volatility" (p. 1657). Thus unemployment volatility and wage stickiness are two sides of the same coin, and the Shimer puzzle can be rephrased as "why are wages sticky?", which we refer to as the "wage flexibility puzzle".[98]

Empirical evidence indeed suggests that real wages are only mildly procyclical. Extensive work by Blanchflower and Oswald (1994) suggests that the elasticity of wages with respect to the unemployment rate is 0.1 in absolute value, and most existing estimates are not very far from this benchmark (see Card (1995), for a review of related work, and Nijkamp and Poot (2005), for a meta-analysis). The modest procyclicality of wages implies that shocks to labour demand have a much larger short-run impact on unemployment rather than wages.

This paper offers an alternative perspective on the Shimer and wage flexibility puzzles and proposes a novel solution – namely we explicitly consider the role of reservation wages and modify the canonical model by introducing backward-looking reference-dependence in their determination. Reference-dependent preferences have often featured in economic behaviour in general and labour supply modelling in particular (Farber, 2008; DellaVigna, 2009), with the aim of explaining observed deviations from the standard neoclassical model of individual decision making. In

[98] Appendix 22.C provides a more formal analysis of the link between unemployment volatility and wage stickiness for the model we use in this paper.

several contexts, reference points are determined by both past personal experiences and peer influences (Akerlof and Yellen, 1990; Blanchard and Katz, 1999).

In the canonical model, reservation wages are forward-looking, determined by current and future labour market conditions. Introducing reference dependence in job search, shaped for instance by one's previous employment history, generates less cyclical reservation wages than the canonical model if reference points are less cyclical than current labour market conditions. If a worker who lost her job at the start of a recession forms future wage aspirations based on her pre-recession earnings, she would set her reservation wage above the level implied by neoclassical – purely forward-looking – preferences. As a consequence, reservation wages may not fall in a recession as much as the canonical model predicts.[99] Related to this point, Falk et al. (2006) show that past minimum wages that are no longer in effect shape reservation wages, making them less cyclical than in the standard search model, and Della Vigna et al. (2017) show that a search model with reference points represented by recent income fares better than conventional models at explaining the pattern of unemployment exits around the time of benefit exhaustion.

A number of papers in the related literature have addressed the unemployment volatility puzzle by proposing alterations to wage determination in the canonical model, and our framework encompasses most of them. In particular, our model allows for weakly cyclical hiring costs (Pissarides, 2000), infrequent wage negotiations in ongoing job matches (Pissarides, 2000; Rudanko, 2009; Haefke et al., 2013; Kudlyak, 2014), and backward-looking elements in wage negotiations in new matches (Gertler et al., 2008; Gertler and Trigari, 2009, introduce both innovations to wage negotiations))[100]. Using these model elements, we derive a relationship between wages and unemployment (the "wage curve") that, under plausible assumptions, is not shifted by labour demand shocks. Demand shocks, independent of their source or magnitude, are associated with movements along the wage curve, and its slope determines the relative volatility of wages and unemployment over the business cycle. Our approach has a natural analogy in a perfectly competitive labour market model, in which the labour supply curve is not shifted by labour demand shocks. We argue that the wage curve is a useful modelling tool for assessing the relative merit of different theories in explaining the wage flexibility puzzle.

We show that, absent reference dependence, infrequent negotiations, and backward-looking elements in wage setting, the canonical model can only replicate the modest observed cyclicality of wages if replacement ratios are extremely high (see also Hagedorn and Manovskii, 2008). If we allow for infrequent renegotiation of wages, implying higher wage cyclicality on new, rather than continuing, matches, and introduce a backward-looking component in newly-negotiated wages, the model can only address

[99] See Genesove and Mayer (2001) for a similar application to the housing market.

[100] See also Shimer (2005) and Rogerson and Shimer (2011) for overviews.

the wage flexibility puzzle if unemployment persistence is implausibly low or the duration of wage contracts is implausibly long.

In addition, none of these ingredients within the canonical model are able to match the observed cyclicality of reservation wages, which this paper estimates using British and German data and finds to be very similar to the cyclicality in wages. The intuition is that, even if wages are completely acyclical (which might happen for a variety of reasons, see e.g. Shimer (2005); Hall (2005); Hall and Milgrom (2008); Michaillat (2012)), the reservation wage still exhibits a considerable amount of cyclicality because workers would be prepared to accept lower wages in a recession, when job opportunities are scarce, independent of the cyclicality of offered wage. Conversely, we show that introducing reference dependence in the determination of reservation wages can explain the low observed cyclicality in both wages and reservation wages for plausible values of the replacement ratio, the persistence in unemployment, and the length of labour contracts.

By shifting the focus of the wage flexibility puzzle onto reservation wages, this paper also makes a contribution to the empirical analysis of reservation wages. In recent years, rich longitudinal data on job search behaviour in the US, analysed by Krueger and Mueller (2011, 2012, 2016), have greatly added to knowledge on reservation wage determination, but they cover too short a time span to investigate their cyclicality. Le Barbanchon et al. (2018) use French administrative data on reservation wages to estimate their sensitivity to potential benefit duration. Our paper adds to the empirics of reservation wages by exploring their cyclical properties. We provide estimates of the cyclicality of wages and reservation wages for the UK and West Germany using micro data from the British Household Panel Survey (BHPS) and the German Socio Economic Panel (SOEP), respectively. These are the only two known sources of (publicly available) information on reservation wages, which cover at least one full business cycle. Our baseline estimates for the elasticity of wages and reservation wages to aggregate unemployment are about -0.17 and -0.15, respectively, for the UK, but markedly lower and only borderline significant for West Germany. All estimated elasticities are considerably lower than those predicted by the theoretical model without reference dependence.

One concern in our empirical analysis of reservation wages is about the quality of reservation wage data, which we address by showing that the correlation between reservation wages and job search outcomes has the sign predicted by search theory. Ceteris paribus, higher reservation wages lead to longer job search spells and higher entry wages upon job finding. We therefore argue that reservation wage data, though likely noisy, embody meaningful information about job search behaviour, and there is no evidence that their cyclicality is systematically under-estimated.

The paper is organized as follows. Section 2 lays out a job search model with infrequent wage negotiations and a backward-looking component in wage setting, allowing for reference-dependent reservation wages. Section 3 derives theoretical predictions for the cyclicality of newly-negotiated wages, wages in new jobs, average wages and reservation wages, illustrating cyclicality predictions of various models for plausible parameters values. Section 4 estimates wage and reservation wage curves for the U.K. and West Germany. Section 5 identifies reference dependence in reservation wages and proposes a quantitative solution to the wage flexibility puzzle. Section 6 concludes.

17 The model

This section lays out a search and matching model to derive implications for the cyclicality of wages. Our set-up allows for elements of wage rigidity proposed by previous work on the Shimer (2005) puzzle, namely acyclical hiring costs (Pissarides (2000)), infrequent wage negotiations in ongoing matches, and backward-looking elements in wage setting for new hires (Gertler and Trigari (2009); Gertler et al. (2008); Pissarides (2000); Rudanko (2009); Haefke et al. (2013); Kudlyak (2014)). In addition, we emphasize the role of reservation wages in wage cyclicality, and allow for reference dependence in their determination. A special case of this model is the classical DMP framework with continuous and forward-looking wage negotiation and without reference-dependence in reservation wages.

In the interest of simplicity, we assume away heterogeneity in workers or jobs, implying homogeneous wages and reservation wages in steady-state. But outside of steady-state, there is heterogeneity across wages set at different times (due to infrequent negotiations) and heterogeneity across wages set at the same time (due to heterogeneity in reservation wages, in turn driven by reference-dependence). However, means of relevant variables provide sufficient statistics for the working of our model, and it will not be necessary to take into account higher moments of the wage and reservation wage distribution. This property stems from the linearity of value functions. To ease the model presentation, we will present and discuss below its key building blocks, and provide derivations in Appendix 22.C.

17.1 Employers

Each firm has one job, which can be either filled and producing or vacant and searching. We denote by $J(w_i;w_{il},t)$ the value of a filled job paying a wage w_i to worker i, whose previous wage was w_{il}. Backward looking reference dependence has the effect of making the wage in a worker's previous job a state variable in the value of the current job, as previous wages influence reservation wages and, hence, wage negotiations. Wages are occasionally renegotiated, and renegotiation opportunities are assumed to arrive at an exogenous rate ϕ,[101] leading to a staggered wage setting process à la Calvo (1983). The parameter ϕ captures the extent to which wages on new and continuing jobs may differ. If the wage in an existing match is renegotiated, neither party has the option to continue the match at the previous wage, which has thus no influence on the outcome of the wage bargain, and any renegotiation results in a new wage $w_{ir}(w_{il}, t)$ where the notation allows the negotiated wage to depend on the previous wage. Based on these modelling elements, the value of a filled job that pays a wage w_i at time t is given by:

$$\begin{aligned} rJ(w_i; w_{il}, t) =& p(t) - w_i - s\left[J(w_i; w_{il}, t) - V(t)\right] \\ &+ \phi\left[J(w_{ir}(w_{il}, t); w_{il}, t) - J(w_i; w_{il}, t)\right] + E_t \frac{\partial J(w_i; w_{il}, t)}{\partial t} \end{aligned} \tag{34}$$

where $V(t)$ is the value of a vacant job at time t, $p(t)$, denotes the productivity of a job-worker pair, and is the ultimate source of shocks, and s is the constant rate at which jobs are destroyed.[102] The second term in square brackets represents the change in job value resulting from renegotiation. Note that, conditional on the current wage, the lagged wage only affects the value function through its potential impact on future wage renegotiations.

The value of a vacant job at time $t, V(t)$, is given by:

$$rV(t) = -c(t) + q(t)E_t[J(w_i; w_{il}, t) - V(t) - C(t)] + E_t \frac{\partial V(t)}{\partial t} \tag{35}$$

Following Pissarides (2000) and Silva and Toledo (2009), we allow the cost of a vacancy to include both a per-period cost, $c(t)$, and a fixed cost, $C(t)$, paid upon hiring. The related literature sometimes indexes vacancy costs to productivity shocks or wages (e.g. Pissarides (2000)), and we return to this issue later. For the moment we simply allow both components of vacancy costs to be time-varying, and assume that they are exogenous to the individual firm. Finally, $q(t)$ is the rate at which vacancies are filled. This rate varies over time via the impact of shocks on labour market tightness.

The first expectation term in equation 35 captures uncertainty about wages in future matches. When a firm and a worker match, we assume that they negotiate a wage with probability α, and such negotiated wage will depend on the worker's previous wage, while with probability $1 - \alpha$ a pre-existing ("old") wage is paid, randomly drawn from the existing cross-section of wages.[103] The extent of job creation at old wages

[101] We assume renegotiation opportunities arrive exogenously, not triggered by a threatened separation caused by a demand shock. This amounts to assuming that demand shocks never cause the surplus in continuing matches to become negative. Allowing for this possibility would induce an extra source of cyclicality as it implies more frequent renegotiation in recessions.

[102] We will consider later a simple extension with countercyclical job separations.

[103] We assume throughout that old wages generate some surplus to both parties. This is the case whenever there is sufficient surplus-sharing in steady-state and deviations from steady-state are small.

(represented by $1-\alpha$) is the source of the backward-looking component in wage setting.

17.2 Workers

Workers can be either unemployed and searching or employed and producing. The value of being employed at a wage w_i at time t when one's previous wage was w_{il} is denoted by $W(w_i; w_{il}, t)$ and given by:

$$\begin{aligned} rW(w_i; w_{il}, t) =& w_i + \phi[W(w_{ir}(w_{il}, t); w_{il}, t) - W(w_i; w_{il}, t)] \\ & - s\left[W(w_i; w_{il}, t) - W(\rho(w_{il}, t); w_{il}, t)\right] + E_t \frac{\partial W(w_i; w_{il}, t)}{\partial t}, \end{aligned} \tag{36}$$

where $\rho(w_{il}, t)$is the reservation wage at time t for a worker with a previous wage w_{il}, and $W(\rho(w_{il}, t); w_{il}, t)$ is the perceived value of being unemployed at time t for someone who has a previous wage of w_{il}.[104]

The value of being unemployed at time t , with a previous wage of w_{il}, is given by:

$$rU(w_{il}, t) = z + \lambda(t) E_t[W(w_i; w_{il}, t) - U(w_{il}, t)] + E_t \frac{\partial U(w_{il}, t)}{\partial t}, \tag{37}$$

where z is the flow utility when unemployed, assumed to be fixed in the short-run,[105] and $\lambda(t)$ is the rate at which the unemployed find jobs, which varies over time with labour market tightness.

17.3 Wage determination

We assume Nash bargaining, thus the wage negotiated at time $t, w_{ir}(w_{il}, t)$, is set to maximize the Nash maximand:

$$[W(w_{ir}; w_{il}, t) - W(\rho(w_{il}, t); w_{il}, t)]^{\beta} [J(w_i; w_{il}, t) - V(t)]^{1-\beta}, \tag{38}$$

where β denotes workers' relative bargaining power. Using value functions (34)-(37), the following result can be proved:

[104]Equation 36 assumes that, when thinking about the capital loss from losing a job, individuals continue to use their current previous wage as the reference point. An alternative would be to assume that they internalize the fact that, in the event of job loss, the current wage will become the new previous wage: we work through this case in the Appendix showing that it leads to the same formulae for the cyclicality of wages when expressed in terms of the replacement ratio.

[105]Chodorow-Reich and Karabarbounis (2016) argue that z is pro-cyclical. Allowing for pro-cyclical z makes wages more pro-cyclical in this model, making it even harder for other elements of the model to solve the wage flexibility puzzle.

Proposition 17.1. *Average newly-negotiated wages can be written as:*

$$w_r(t) = \rho(t) + \tilde{\beta}(r + \phi + s)\mu(t) - \tilde{\beta}(1 - \alpha)[w(t) - w_r(t)], \tag{39}$$

where $w_r(t)$ denotes the average newly-negotiated wage, $\rho(t)$ denotes the average reservation wage, $\tilde{\beta} \equiv \beta/(1-\beta)$, $\mu(t) = C(t) + c(t)/q(t)$ and $w(t)$ denotes the average wage.

Proof. See Appendix 22.C. □

Equation (39) gives an extremely simple result for the average negotiated wage, despite our general set-up with heterogeneous wages outside of steady state. This property stems from the linearity of value functions. The ability to provide a simple formula for average wages without having to be concerned about higher moments of the wage distribution comes from the fact that the value functions (34)-(37) are linear in individual wages.[106] Equation (39) states that average newly-negotiated wages are equal to reservation wages plus two further terms. The first term, $\tilde{\beta}(r + \phi + s)\mu(t)$, reflects workers' bargaining position and it depends on workers' relative bargaining power $\tilde{\beta}$ and expected hiring costs $\mu(t)$. This term determines the mark-up of wages over workers' outside options, as expected hiring costs measure the cost to the employer of replacing the current worker. The final term in $(w(t) - w_r(t))$ derives from ex-ante uncertainty about the wage to be paid in a new job, whether it will be drawn from the existing cross-section of wages or newly-negotiated. A higher average wage reduces the value of creating a new job, and, hence, must be offset by lower negotiated wages in equilibrium. One key feature of (39) is that variables productivity shocks or to the efficiency of the matching function (in turn affecting matching rates $\lambda(t)$ and $q(t)$) only impact wages only via the reservation wage (and/or the mark-up $\mu(t)$ whenever $c(t) > 0$). This result will become useful below.

17.4 The reservation wage

The wage equation 39 is conditional on the reservation wage, which is itself endogenous, and we consider alternative models for its determination. We consider first purely forward-looking behaviour, as implicit in the canonical model. In this case the reservation wage is the wage that makes a worker indifferent between work and unemployment. We call this the optimal reservation wage and denote it by $\rho_i^o(w_{il}, t)$, which is the solution to $W(\rho_i^o(t); w_{il}, t) = U(w_{il}, t)$. The following Proposition provides a differential equation for the optimal average reservation wage, $\rho^o(t)$.

[106] This result differs from that of Krusell et al. (2010), whose assumption of risk-averse individuals introduces non-linearities in value functions.

Proposition 17.2. *The average optimal reservation wage satisfies the following differential equation:*

$$\begin{aligned}[r+\lambda(t)+s]\,[\rho^o(t)-z] = \; & E_t\frac{d\rho^o(t)}{dt}+[\lambda(t)-\phi]\,[w_r(t)-z] \\ & +(1-\alpha)\lambda(t)\,[w(t)-w_r(t)]\end{aligned} \tag{40}$$

Proof. See Appendix 22.C. □

and it is a function of average and newly-negotiated wages and the tightness of the labour market as represented by $\lambda(t)$. Note that the validity of this expression does not hinge on Nash rent-sharing. In fact equation (40) is valid conditional on wages, however they are determined.

We contrast this reservation wage model with one that deviates from purely forward-looking behaviour by encompassing backward-looking reference points, which we assume to be (partly) determined by recent earnings. The idea of backward-looking reference dependence is consistent with prospect theory, in which outcomes are evaluated against a natural benchmark represented by the status quo (Kahneman and Tversky, 1979). It is also consistent with the concept of aspiration-based references points whenever individuals expect to maintain the status quo with some probability (Loomes and Sugden, 1986; Koszegi and Rabin, 2006). Kahneman and Tversky (1979) provide evidence of backward-looking reference points for the retail market, in which firms and customers use past prices as benchmark for judging the fairness of a transaction, and Genesove and Mayer (2001) offer similar evidence for the housing market. Most closely related to our set-up, Falk et al. (2006) and Della Vigna et al. (2017) show evidence on the role of past earnings as reference points during job search.

To introduce backward-looking reference points in out model, we assume that the deviation of the current reservation wage from its steady-state level ρ^* has two components. The first component is a function of the deviation of the optimal, forward looking reservation wage, $\rho_i^o(w_{il},t)$, from its steady-state level ρ^*. The second component is a function of the deviation of a reference wage, w_{il}, from its steady-state level, w^* (formulae for steady-state wages and reservation wages will be provided later). These assumptions lead to the following expression for the average reservation wage:

$$\rho_i(w_{il},t)-\rho^* = \alpha_\rho(\rho_i^o(w_{il},t)-\rho^*)+(1-\alpha_\rho)\alpha_l(w_{il}(t)-w^*), \tag{41}$$

where α_ρ captures the weight of forward-looking behaviour in reservation wages, w_l is the average, last observed wage before job loss, and captures its role in reference points. Lower α_ρ implies stronger reference dependence in reservation wages, and lower α_l implies lower cyclicality in reference points. The special case $\alpha_\rho = 1$ takes us back to the forward-looking model. Taking averages of (41) implies we can write the equation in terms of averages.

$$\rho(t) - \rho^* = \alpha_\rho(\rho^o(t) - \rho^*) + (1 - \alpha_\rho)\alpha_l(w_l(t) - w^*), \tag{42}$$

18 The Predicted Cyclicality of Wages

We derive model predictions for the cyclicality of wages and reservation wages, as measured by their respective elasticity with respect to the current unemployment rate. We start with the special case of a comparison of steady states, before considering the general case in which labour market conditions vary over time.

18.1 A comparison of steady states

In a steady-state when labour market conditions are constant, all wages, whether pre-existing or newly-negotiated will be the same as will be all reservation wages. From (39) steady-state wages can be written as:

$$w^* = \rho^* + \tilde{\beta}(r + \phi + s)\mu, \tag{43}$$

and, from (40), steady-state reservation wages can be written as:

$$\rho^* = z + \frac{\lambda^* - \phi}{r + \lambda^* + s}(w^* - z). \tag{44}$$

Combining equation (43) and equation 44 leads to the steady-state wage equation:

$$w^* = z + \tilde{\beta}(r + \lambda^* + s)\mu, \tag{45}$$

(45) can be used to compare how wages and unemployment co-vary when comparing steady-states. There are two reasons why wages may be procyclical. First, λ^* is higher when unemployment is lower as steady-state unemployment is given by $u^* = s/(s + \lambda^*)$. Second, hiring costs μ may vary with the cycle. First, using equation 67, if the cost of filling vacancies have a positive flow component ($c > 0$), hiring costs rise when unemployment is low, as vacancy durations rise (q falls). Secondly, vacancy costs themselves (c and C) may vary, and the literature often indexes them to productivity (Pissarides (2009)) or to the level of wages (Hagedorn and Manovskii (2008), do both). In either case hiring costs are pro-cyclical, in turn accentuating the pro-cyclicality of wages. As one of the aims of this paper is to show why it is hard for the canonical model to generate the modest observed cyclicality of wages, we assume in most of what follows that hiring costs are acyclical, which implies that the value of jobs to firms is also acyclical. Indeed most studies on the costs of filling jobs find the fixed cost component

to be more important than the variable cost, so we assume $c = 0$ and that C does not vary with short-term fluctuations in productivity and/or wages.[107] These assumptions imply constant μ. Replacing λ^* with steady-state unemployment, u^*, 45 can be wirtten as:

$$w^* = z + \tilde{\beta}\mu\left(r + \frac{s}{u^*}\right) \tag{46}$$

(46) gives a relationship between wages and the unemployment rate. This wage curve is conceptually akin to a labour supply curve in a competitive model, in the sense that productivity shocks and shocks to the matching function do not shift it, but drive movements along it[108]. The slope of equation 46 determines the relative response of wages and unemployment to shocks, independent of their source or size, allowing us to be agnostic about the nature of demand shocks, and to evaluate model predictions without measuring them. The robustness of our approach to a number of different sources of shocks is an advantage of our approach.

In what follows we focus on the elasticity of wages with respect to unemployment as the key cyclicality parameter. Differentiating equation 46 gives such elasticity across steady-states:

$$\epsilon_{w^*} = -\frac{\beta\mu}{1-\beta}\frac{s}{w^*u^*} = -\frac{w^*-z}{w^*}\frac{s}{ru^*+s} = -(1-\eta)\frac{s}{ru^*+s}, \tag{47}$$

where notation $\epsilon_x \equiv \partial \ln x / \partial \ln u$ is used to denote the unemployment elasticities of any generic variable x and $\eta \equiv z/w^*$ is the replacement ratio. As s is substantially larger than ru for conventional values of the interest rate, the $s/(ru^* + s)$ ratio is close to 1, implying that the unemployment elasticity of wages should be close to one minus the replacement ratio. Using the Blanchflower and Oswald (1994) benchmark of -0.1, equation 48 requires a replacement ratio of 0.9, a very close value to the 0.95 calibration used by Hagedorn and Manovskii (2008).

This value seems implausibly high. The OECD Social Policy Database[109] computes the proportion of net in-work income that is maintained when a worker becomes unemployed, by household composition and unemployment duration. In 2001, the overall average of this ratio across worker types in the UK and Germany was 0.60 and 0.66, respectively. These estimates do not assign a value to the increase in home time for the unemployed, and there is no definitive evidence on the size of this component. Krueger and Summers (1988) report that home production and leisure activities increase

[107] Of course, one has to assume that in the long run the vacancy cost is linked to productivity and/or wages as otherwise long-run growth would make the vacancy filling costs less and less important.

[108] Our approach here is different from much of the existing literature that focuses on measuring the relatively sensitivity of wages and unemployment to productivity shocks. Appendix 22.C shows how that approach leads to the same conclusions.

[109] http://www.oecd.org/els/soc/$NRR_Over5years_EN$.xlsx

during unemployment, but at the same time the unemployed enjoy these activities less than the employed.

We propose an approach to calibrate the replacement ratio that avoids the need to make an assumption about the value of home time. Rearranging equation 44, one can obtain the steady-state relationship between the replacement ratio and ρ^*/w^*, the ratio of reservation wages to wages:

$$1 - \frac{\rho^*}{w^*} = (1 - \eta)\frac{r + \phi + s}{r + \lambda + s}. \tag{48}$$

In the BHPS, unemployed workers are asked both for their reservation wages and expected wages on re-employment and the answers to these questions can be used to estimate ρ^*/w^*, whose median value during our sample period (1991-2009) is 0.80. As the duration of a wage contract $(1/\phi)$ is typically longer than the duration of a spell of unemployment $(1/\lambda)$, equation 49 implies an upper bound for the replacement ratio of 0.80. And for realistic values of λ and ϕ it will be markedly lower. To get a sense of magnitudes, if wages are renegotiated on average once a year ($\phi = 0.083$ on monthly data) and the job finding rate is set to its average level observed during our sample period for the UK ($\lambda = 0.139$ monthly), the replacement ratio equals 0.69. This value is somewhat above the benefit replacement ratio estimated from the OECD data,[110] but well below the level required to match the estimated elasticity of the wage curve.

In this comparison of steady-states we have assumed, as in much of the related literature, that variation in unemployment rates is associated with variation in job-finding rates, at constant job separation rates. However, countercyclical separations would amplify the impact of shocks on unemployment, as in a recession unemployment increases both because it is harder to find a job and it is easier to be made redundant (see, among others, Fujita and Ramey (2009), and Elsby and Michaels (2013)). This point is also made in the context of the unemployment volatility puzzle by Robin (2011). If we differentiate equation (46) for the general case with countercyclical separations we obtain:

$$\epsilon_{w^*} = -(1 - \eta)\left(\frac{s}{ru + s} - \frac{su}{r + su}\epsilon_s\right) \tag{49}$$

implying that a lower replacement ratio is necessary to match a given elasticity of wages to unemployment, as $\epsilon_s > 0$. However, the effect is quantitatively very small. Using the estimate of Elsby and Michaels (2013), ϵ_s is about 0.17 in the UK and 0.47 in Germany. The required replacement ratio to match a -0.1 elasticity of wages to unemployment would be very close to 0.89 in both countries.

We next turn to the cyclicality of the reservation wage. Using equation 39 and the

[110] This is consistent with a positive value of leisure utility, net of search costs, during unemployment.

assumptions of acyclical hiring costs and steady state ($w_r = w_a$) we obtain:

$$\epsilon_\rho = \frac{w^*}{\rho^*}\epsilon_w < \epsilon_w \tag{50}$$

i.e. reservation wages are more strongly cyclical than wages, and the ratio between the respective elasticities is given by $\rho^*/w^* = 0.80$. Thus reservation wages are expected to be about 25% more cyclical than wages. The existing literature does not provide estimates for the cyclicality of reservation wages, but later in the paper we provide such estimates for the UK and West Germany.

18.2 The General Case

In the general case, the economy can be out of steady state. We are interested in the predictions of the alternative models discussed for the elasticity of current wages with respect to the current unemployment rate. As our model is highly non-linear, we linearize it around its steady state, and we derive linear projections of all relevant variables (whether past, current or expected future) on the current unemployment rate. These linear predictions can then be used to obtain the elasticity of wages with respect to the unemployment rate predicted by the model. This methodology is described in general terms in more detail in Appendix 22.C: the approach has the advantage that it leads to a closed-form solution for the elasticity of interest as a function of model parameters, making the working of the model more transparent than with alternative approaches.

Out of steady state, workers need to form expectations about the dynamics of labour market conditions because the reservation wage is forward-looking according to equation 40 and equation 41. A key parameter underlying the formation of expectations is the persistence of shocks to labour market conditions. As $\lambda(t)$ summarizes labour market conditions in our model, it is convenient express all variables of interest in terms of the persistence in $\lambda(t)$. Based on linearization and the linear projections, the dynamic process for $\lambda(t)$ can be represented as:[111]

$$E_t\left(\frac{d\lambda(t)}{dt}\right) = -\xi\left(\lambda(t) - \lambda^*\right) \tag{51}$$

This implies that $\lambda(t)$ can be modelled 'as if' it followed the continuous-time version of an AR(1) process, where ξ represents its rate of convergence to steady state, with lower ξ implying higher persistence. This 'as if' result comes from the methodology of linearization and linear projection and is not an assumption that $\lambda(t)$ actually follows an AR(1) process. The persistence parameter ξ is also not unique to $\lambda(t)$. The discussion

[111] See Appendix 22.C for derivation.

of our methodology in Appendix 22.C shows that expressing changes in relevant variables in term of deviations of $\lambda(t)$ from its steady state is simply a convenient normalization, and normalizations based on other variables like unemployment or productivity yield identical results; and the evolution of these other variables would be characterised by the same persistence parameter ξ.

Our approach seems to differ from the one adopted in some of the related work, where it is more common to make assumptions about the dynamic process for productivity, which is typically modelled as AR(1), and in Appendix 22.C we present a version of our model in which the primitive model shocks are directly embodied in productivity. This alternative model version leads to equivalent results about cyclicality as our main model, but we argue that conditioning on $\lambda(t)$ rather than productivity allows us to be more agnostic about the nature of labour demand shocks.

Given these premises, we can prove the following Proposition about the cyclicality in wages and reservation wages:

Proposition 18.1.

1. The cyclicality of *newly-negotiated wages*, conditional on the cyclicality of reservation wages, is:

$$\epsilon_{w_r}(t) = \Gamma_r \left[\frac{\rho^*}{w^*} \epsilon_\rho(t) + \left(1 - \frac{\rho^*}{w^*}\right) \epsilon_\mu(t) \right], \tag{52}$$

where:

$$\Gamma_r = \frac{\alpha s + \phi + \xi}{(\alpha s + \phi + \xi) - \tilde{\beta}(1-\alpha)\xi} \tag{53}$$

2. The cyclicality of *reservation wages*, conditional on the cyclicality of newly-negotiated wages, is:

$$\epsilon_\rho(t) = -\frac{\alpha_\rho(\lambda^* + s + \xi)}{r + \lambda^* + s + \xi} \left(\frac{w^*}{\rho^*} - 1 \right) + \frac{w^*}{\rho^*} \Gamma_\rho \epsilon_{w_r}(t) \tag{54}$$

where

$$\Gamma_\rho = \frac{\alpha_\rho \left[(\alpha s + \phi)(\lambda^* - \phi) + \xi(\alpha\lambda - \phi) \right]}{(r + \lambda^* + s + \xi)(\alpha s + \phi + \xi)} + \frac{(1-\alpha_\rho)\alpha_l \lambda^* (\alpha s + \phi)}{(\lambda^* + \xi)(\alpha s + \phi + \xi)} \tag{55}$$

3. The cyclicality of *average wages* is:

$$\epsilon_w(t) = \frac{\alpha s + \phi}{\alpha s + \phi + \xi} \epsilon_{w_r}(t) \tag{56}$$

4. The cyclicality of *wages in new jobs* is:

$$\epsilon_{w_n}(t) = \frac{\alpha s + \phi + \alpha\xi}{\alpha s + \phi + \xi}\epsilon_{w_r}(t) = \frac{\alpha s + \phi + \alpha\xi}{\alpha s + \phi}\epsilon_{w}(t). \tag{57}$$

5. The cyclicality of *reference wages* is:

$$\epsilon_{w_l}(t) = \frac{\alpha s + \phi}{\alpha s + \phi + \xi}\frac{\lambda}{\lambda + \xi}\epsilon_{w_r}(t) \tag{58}$$

Proof: See Appendix 22.C.

These expressions provide insight on how key elements of the model – infrequent wage renegotiations, unemployment persistence, a backward-looking component in wage setting and reference dependence in reservation wages – affect the cyclicality of wages (average, in new jobs and newly-negotiated), reservation wages and reference wages. One of the advantages of our approach based on the wage curve is that one can provide closed-form expressions for all elasticities of interest.

Equations 56-58 express the cyclicality of average wages, wages in new jobs and reference wages as a function of the cyclicality of newly-negotiated wages. The relative magnitude of elasticities depends on model parameters and the closed form solution allows us to illustrates how several recent papers have addressed issues relating to wage cyclicality in the model.

First, equations 56-58 show that newly-negotiated wages are more cyclical than all other wages. This is in line with empirical evidence that wages in new jobs are more cyclical than average wages, as presented in Devereux and Hart (2006). This insight can help model predictions become closer to the data, as it is the cyclicality of wages on new hires that matters for hiring decisions and the cyclical behaviour of unemployment and vacancies (Hall (2005), Pissarides (2009), Haefke et al. (2013)). Quantitatively, the predicted difference in wage cyclicality between new and continuing jobs widens with the forward-looking component in wages (rising α) and with the length of labour contracts (falling ϕ), and shrinks with unemployment persistence (falling ξ). But, if unemployment is highly persistent, wages in new and continuing jobs tend to display similar degrees of cyclicality, independent of α and ϕ: although wages negotiated at different points in time reflect labour market conditions at different points in time, these are strongly serially correlated, and a regression of wages negotiated in the past on current unemployment would detect a significant relationship, with an elasticity not very different from that detected for new jobs. In the limiting case $\xi \to 0$, newly-negotiated wages, average wages, wages in new jobs and reference wages are all equally cyclical. This will turn out to be very important because our benchmark estimated values of ξ are 0.003 and 0.004 for the UK and Germany, respectively.

Consider next the predicted cyclicality of newly-negotiated wages and reservation wages. Equation 54 expresses the cyclicality of newly-negotiated wages as a function of the cyclicality of reservation wages and hiring costs. This result follows from the solution to the Nash sharing problem for wage setting equation 39 and is independent of the specific model of reservation wages used. Similarly, equation 54 expresses the cyclicality of reservation wages as a function of the cyclicality of newly-negotiated wages and an extra term that, over and above wage cyclicality, predicts reservation wages to fall in a recession: when unemployment rises and the chances of re-employment fall (falling λ), workers are willing to accept lower wage offers. This result follows from the reservation wage model Equation 42 and its validity is independent of the specific model of wage determination used.

This characterization of the wage and reservation wage elasticity is useful to shed light into potential solutions to the wage flexibility puzzle. Consider a case in which wages are only very weakly cyclical (as might happen if outside options have only a limited influence during wage bargaining, for the reasons argued by Hall and Milgrom (2008)). Even if wages are completely acyclical, expression equation 54 states that reservation wages would still be cyclical, as reemployment chances fall in recessions. In particular, if reservation wages are equal to optimal reservation wages ($\alpha_\rho = 1$) and the interest rate is low relative to the job-finding rate (as indeed it is), the predicted elasticity of reservation wages with respect to unemployment is about $-(w^*/\rho^* - 1)$. In the BHPS data this is estimated to be about -0.25, so reservation wages would be indeed quite cyclical. However, a value of α_ρ below 1, capturing reference dependence in reservation wages, would reduce the predicted cyclicality of reservation wages.

Similarly, equation 52 predicts that, if unemployment is very persistent ($\xi \to 0$), newly-negotiated wages are less cyclical than reservation wages, unless hiring costs are pro-cyclical. But pro-cyclical hiring costs obviously tend to raise the overall cyclicality of wages, making it harder for the canonical model to address the wage flexibility puzzle. If both reservation wages and hiring costs are not very cyclical, equation 52 predicts that newly-negotiated wages cannot be very cyclical either.

It is useful to derive a closed-form expression for the cyclicality of average wages in terms of all model parameters. Combining equation 52, 54 and 56, we derive the following Proposition:

Proposition 18.2.

The cyclicality of *average wages* can be written as:

$$\epsilon_w(t) = -(1-\eta^*)\frac{as+\phi}{as+\phi+\xi}\frac{\Gamma_r}{1-\Gamma_r\Gamma_\rho}\frac{r+\phi+s}{r+\lambda+s}\left[\frac{\alpha_\rho(\lambda+s+\xi)}{r+\lambda+s+\xi} - \frac{\partial \ln \mu(t)}{\partial \ln u(t)}\right] \tag{59}$$

Proof: Solving equation 52, 54 and 56, and then using equation 48 to eliminate ρ^*/w^* leads to equation 59.

Equation 59 highlights the role of key model parameters in wage cyclicality. However, due to its complexity, equation 59 is hard to visualise in its most general form, and we will develop its understanding in steps.

We first consider three special cases without reference dependence in reservation wages ($\alpha_\rho = 1$): the first case features continuous wage renegotiation ($\phi = \infty$); in the second case we turn off the backward-looking behaviour in wage setting ($\alpha = 1$); and in the third case we allow for both infrequent negotiations and backward-looking behaviour in wage setting ($\phi < \infty$ and $\alpha < 1$). We show that a model with continuous wage renegotiation cannot solve the wage flexibility puzzle, for any degree of backward-looking behaviour in wage setting, unless the replacement ratio is implausibly high. The model without backward-looking behaviour in wage determination is also unable to solve the puzzle unless unemployment has low persistence. However, we show that there are combinations of ϕ and α that do predict plausible values of wage cyclicality. But this solution to the puzzle comes with two major drawbacks. First, for plausible values of unemployment persistence a very low, unrealistic, value of ϕ is necessary to deliver low wage cyclicality, implying wage contracts longer than 4 years on average. Second, all these versions of the canonical model still predict high cyclicality in reservation wages. This leads us to the introduction of reference-dependent reservation wages in the canonical model ($\alpha_\rho < 1$), and we show that the modified model is capable of solving the wage flexibility puzzle for plausible values of all other parameters.

To give a sense of magnitudes involved, we will evaluate equation 59 and the role of its components at benchmark parameter values, unless otherwise stated. The parameter calibration used is described below.

18.3 Benchmark parameters

For the UK, we use the Quarterly Labour Force Survey (LFS) to obtain estimates of the average unemployment rate and monthly separation rate over the same sample period used in Section 4 (1991-2009). This gives $u = 0.067$ and $s = 0.010$, implying $\lambda = s(1-u)/u = 0.139$. For West Germany, we obtain $u = 0.078$ and $s = 0.012$ on the Socio Economic Panel (SOEP) for 1984-2010, yielding $\lambda = 0.142$.

The persistence parameter ξ is obtained by estimating an AR(1) model on the monthly, seasonally adjusted, time series for the unemployment rate. This series is available for the UK from the Office for National Statistics for 1971 onwards, and we use data for 1971-2014. We estimate $\xi = 0.003$ on the raw series, and $\xi = 0.007$ on HP filtered series. Virtually identical results are obtained by fitting the AR(1) model to the log – as opposed to the level – of the unemployment rate, or when using quarterly series. For West Germany, a harmonised, seasonally adjusted, series for the unemployment rate is available from the Bundesbank, from 1991 onwards, and we use data for 1991-

2014. We estimate $\xi = 0.004$ on the raw monthly series, whether in levels or logs, and obtain slightly higher estimates of 0.018-0.019 on the filtered series. Estimates on the quarterly series are very similar to those obtained on the monthly series.[112] We use as benchmark values the estimates obtained on the raw series (0.003 for the UK, 0.004 for West Germany) but will show predictions for higher values of ξ. As shown in Appendices A and B, one could have equivalently calibrated ξ to the persistence of productivity, rather than unemployment. When we do so, we obtain $\xi = 0.001$ and $\xi = 0.005$ for the UK and Germany, respectively. These values are extremely close to those directly estimated on unemployment and would produce virtually identical results for the cyclicality of wages.

We assume an expected contract length of 12 months in both countries, corresponding to $\phi = 0.083$. This seems to be the mode in most medium and large firms according to the review of wage setting practices in the US by Taylor (1999) (Section 2.2.1). Gottschalk (2005) estimates that in the US the hazard of a change in wages peaks 12 months after the previous change and Fabiani et al. (2010) find that 60% of firms in a number of European countries change base wages once a year. While we pick $\phi = 0.083$ as our benchmark value, we will show predictions for a range of values of ϕ.

We set the bargaining power of workers at $\beta = 0.05$ (see estimates reported by Manning (2003), Table 27), and note incidentally that β has limited importance whenever α is high, or newly-negotiated wages are close enough to average wages. We consider a monthly interest rate $r = 0.003$.

For pinning down the value of the replacement ratio, we use condition equation 48 to combine BHPS data on reservation wages and expected wages ($\rho/w \simeq 0.80$) and benchmark values for other parameters, which yield $\eta = 0.69$ in the UK. For West Germany, there is no available information on expected wages during unemployment, thus we calibrate the replacement ratio assuming that it exceeds the unemployment benefit ratio by the same amount as in the UK, i.e. 9 percentage points. This is equivalent to assuming that the extra utility of leisure enjoyed during unemployment is the same in both countries, giving $\eta = 0.75$ in West Germany. Our benchmark parameter values are summarized in Table 24.

[112] We use conventional smoothing parameters (129600 on monthly data; 1600 on quarterly data) in the HP filter but note that the resulting estimated trend component of unemployment for West Germany retains some degree of cyclicality. The trend becomes less cyclical with higher smoothing parameters, and the estimated persistence on the resulting filtered series is also then higher. Labor market flows relative to employment and unemployment stocks are lower in the UK and Germany than in the US, and unemployment persistence is higher, thus our estimates for somewhat differ from those often used in the literature for the analysis of US data.

18.4 Case I: Continuous Wage Negotiation

We start our analysis of equation 59 within the canonical model ($\alpha_\rho = 1$), imposing continuous renegotiation of wages, $\phi = \infty$.[113] We also assume that hiring costs are acyclical, as this makes it easier for the model to deliver low wage cyclicality.

Proposition 18.3.

If $\epsilon_\mu(t) = 0$ and $\alpha_\rho = 1$, then $\phi = \infty$ implies

$$\epsilon_w(t) = -(1-\eta^*)\frac{\lambda^* + s + \xi}{r + \lambda^* + s} = -(1-\eta^*)\frac{\xi u^* + s}{ru^* + s} \tag{60}$$

Proof: See Appendix 22.C

This expression coincides with the steady-state result equation 90, except for the ξu^* term in the numerator, which is quantitatively small as unemployment is highly persistent. At UK benchmark parameter values, the predicted elasticity of wages to unemployment in and out of steady state is -0.30 and -0.31, respectively. Using values for Germany, we obtain -0.24 and -0.25, respectively. From equation 60 it follows that – as in steady-state – a model with continuous wage negotiation can only deliver relatively low wage cyclicality if the replacement rate is higher than our calibrated value.

18.5 Case II: No backward-looking component in wage determination

We now assume that wages are newly-negotiated every time a new match is formed ($\alpha = 1$), while keeping the assumptions of no reference-dependence and acyclical hiring costs.

Proposition 18.4.

If $\epsilon_\mu(t) = 0$ and $\alpha_\rho = 1$, then $\alpha = 1$ implies

$$\epsilon_w(t) = -(1-\eta^*)\frac{\xi u^* + s}{ru^* + s}\frac{s + \phi}{s + \phi + \xi}\frac{r + \phi + s}{r + \phi + s + \xi}. \tag{61}$$

Proof: See Appendix 22.C

[113] Note incidentally that a model with continuous wage re-negotiation delivers counter-factual predictions about the level of reservation wages, as equation 40 and equation 44 imply $\rho^0, \rho^* \to -\infty$ for $\phi = \infty$. In other words, workers are willing to accept a job at any wage because they expect any accepted wage to be immediately revised, thanks to continuous renegotiation.

This special case differs from the previous one in equation 60 by the final two terms in equation 61, which are both below 1, thus a model with infrequent wage negotiations and $\alpha = 1$ delivers less cyclical wages than one with continuous negotiations. But again this difference is quantitatively small for small enough ξ, and the two models predict equally cyclical wages for $\xi = 0$. At benchmark parameter values, the predicted wage elasticity in equation 61 is -0.29 for the UK and -0.23 for Germany. The intuition is that, while infrequent negotiations imply that a large share of ongoing wage contracts were negotiated in the past and thus reflect past labour market conditions, current unemployment is strongly correlated with past unemployment because of its high persistence.

18.6 Case III: Infrequent Wage Negotiation and Backward-Looking Wages

While the canonical model may not solve the wage flexibility puzzle with either continuous wage renegotiation or fully forward-looking wage determination (unless unemployment has implausibly low persistence), there exist combinations of α and ϕ that do predict mild wage cyclicality even in the presence of high unemployment persistence. In the extreme case in which $\alpha, \phi \to 0$, equation 59 implies that the average wage elasticity also goes to zero. By continuity, there must exist values of α and ϕ that deliver a sufficiently low value of the wage elasticity.

Below we discuss the plausibility of such values. Panel A in 34b presents the combinations of ϕ and α that predict elasticities of average wages in the UK of $-0.05, -0.1, -0.15$ and -0.2 in turn, in correspondence of benchmark values of other parameter. As ϕ and α decrease, predicted wage cyclicality also falls, but the steep slope of the "isoquant" curves imply that ϕ plays a much more important role than α in determining wage cyclicality. But the values of ϕ required to match observed wage cyclicality are much lower than estimates. For example, even if all new job matches are paid old wages ($\alpha = 0$), a value of $\phi = 0.011$ is needed to predict a wage elasticity of -0.1, and $\phi = 0.019$ to predict a wage elasticity of -0.15, which is close to our estimates on BHPS data in the next Section. As $1/\phi$ is the expected duration of a wage contract in months, $\phi = 0.019$ implies that wage contracts are only negotiated every 4 years on average. If some share of newly-hired workers do negotiate their wages ($\alpha > 0$) the implied value of ϕ would be even lower. For example, if $\alpha = 0.5$, a wage elasticity of -0.10 (respectively, -0.15) requires that wages are renegotiated roughly every 24 (respectively, 9) years. These are implausibly long contract durations when compared to available evidence – e.g. Taylor (1999), Gottschalk (2005) and Fabiani et al. (2010) imply that most wage contracts are renegotiated once a year.

In Panel (b) we show that, if unemployment is less persistent, there exist lower and realistic values of ϕ that would bring wage cyclicality down to a level close to existing estimates. For example, if $\xi = 0.1$, the range of values consistent with a -0.1 wage elasticity contains 0.083, in line with yearly wage negotiations.

Figure 1 also shows the corresponding reservation wage predictions. Panel (a) shows that the canonical model with $\xi = 0.003$ would only be able to match a reservation wage elasticity of -0.15 (which we estimate in the next Section) with values of ϕ below 0.02. More importantly, even if we let ξ rise to 0.1, a very low value of ϕ is still needed to reconcile model predictions with the observed cyclicality of reservation wages (Panel (b)).

Predictions for West Germany are reported in 35. In Panel A the values of ϕ needed to deliver a given wage or reservation wage elasticity for $\xi = 0.004$ are slightly higher than in the UK (and in Panel (b) a higher, counterfactual, value $\xi = 0.1$ further raises the required values of ϕ). However, as we show in the next Section, the estimated wage and reservation wage elasticities are markedly lower in West Germany than in the UK, in most cases between 0 and -0.05. Reconciling these elasticities with model predictions in Panel (a) and (b) would still require implausibly low values of ϕ.

In summary, this analysis has shown that the canonical model can only match the observed cyclicality of wages in both the UK and West Germany under either an implausibly long duration of wage contracts, or an implausibly low value of unemployment persistence. For given values of ϕ and ξ, the canonical model fares much worse at predicting reservation wage cyclicality than wage cyclicality. In other words, a clear drawback to solving the wage flexibility puzzle via low ϕ and high ξ is that the canonical model still predicts considerable "excess" cyclicality in reservation wages, as implied by equation 54.

18.7 Reference-dependence in Reservation Wages

We next show that a model with reference-dependent reservation wages may fare better in delivering weakly cyclical wages and reservation wages, even in the presence of high unemployment persistence. Some insight is gained by considering equation equation 59 for the limiting case $\xi \to 0$, having imposed acyclical hiring costs.

Proposition 18.5.

If $\partial \ln \mu(t)/\partial \ln u(t) = 0$, then $\xi \to 0$ implies:

$$\epsilon_w(t) \to -\alpha_\rho(1-\eta^*)\frac{\lambda+s}{r+\lambda+s}\frac{r+\phi+s}{(r+s)\left[1-(1-\alpha_\rho)\alpha_l\right]+(1-\alpha_\rho)(1-\alpha_l)\lambda+\alpha_\rho\phi} \tag{62}$$

Proof: See Appendix 22.C

Note first that the extent of backward-looking behaviour in wage setting plays no role in equation 62 because, as unemployment is fully persistent, current and lagged wages are perfectly correlated; thus making wages backward-looking has no effect on cyclicalities. Second, equation 62 shows that the parameters measuring reference-dependence (α_ρ) and backward-looking behaviour in the determination of reservation wages (α_l) do play a role in wage cyclicality. In particular, as $\alpha_\rho \to 0$ predicted wage cyclicality goes to zero and, conditional on α_ρ, cyclicality rises with α_l, measuring the role of the last observed wage in reference dependence. Turning to reservation wage cyclicality, equation 54 implies that lower α_ρ reduces the cyclicality of reservation wages both directly, via reference dependence, and indirectly, via lower wage cyclicality.

Below we show graphically the relative importance of reference dependence in reservation wages (α_ρ) and backward looking behaviour in wage setting (α) in addressing the wage cyclicality puzzle. Panel (a) in 36 plots the elasticity of average wages in the UK, as a function of $1 - \alpha$ and $1 - \alpha_\rho$ in turn, in correspondence of benchmark values of other parameters. The top dotted line plots the relationship between the elasticity of average wages and $1 - \alpha$, having imposed $\alpha_\rho = 1$. This is the curve denoted "Model 1", delivering a relatively high wage cyclicality, which is not very responsive to the actual level of backward looking behaviour in wage setting. This is a different way of restating the point made in Panel A of Figure 1, namely that the canonical model cannot replicate the observed wage cyclicality for plausible values of ϕ and ξ, whereby α does not have any strong impact.

We next consider the role of reference points in reservation wages, by plotting the relationship between $1 - \alpha_\rho$ and the cyclicality in average wages, having ruled out backward looking behaviour in wage setting ($\alpha = 1$). We consider two cases, denoted as "Model 2" and "Model 3", respectively. In Model 2, the reference wage is completely acyclical ($\alpha_l = 0$), and in Model 3 the reference wage is as cyclical as average wages ($\alpha_l = 1$). The main result is that both reservation wage models generate less cyclical wages than a model that introduces the same level of backward-looking behaviour in wage setting. The reduction in cyclicality is clearly greater when the reference point in reservation wages is assumed to be completely acyclical ($\alpha_l = 0$). Panel (b) in 35 gives a very similar picture for the predicted cyclicality of reservation wages: while the elasticity of reservation wages to unemployment hardly responds to backward-looking behaviour in wage setting, it declines much faster with reference dependence in reservation wages. Panels (a) and (b) in 37 give the corresponding predictions for Germany, and allow us to draw exactly the same conclusions about the relative roles of α and α_ρ.

This section has shown that, when reservation wages have a reference-dependent component, our model can produce markedly less-cyclical wages and reservation wages for plausible benchmark parameter values. Once we allow for reference dependence in reservation wages there is no need to alter the wage setting process to make wages more rigid than in the canonical model. While the empirical literature has established that wages are only mildly cyclical, there is no corresponding evidence for reservation wages. We turn to the empirical analysis of wages and reservation wages in the next section.

19 Empirical wage and reservation wage curves.

19.1 Estimates of the wage curve

This Section provides estimates of wage and reservation wage cyclicality for the UK and West Germany, based on data from the BHPS and the SOEP, respectively. Both are longitudinal studies: the BHPS runs from 1991 to 2009, and the SOEP runs from 1984 onwards. The main advantage of these data sets is that they contain information on reservation wages over a long period of time.

We first provide wage curve estimates, and focus on the elasticity of (log) hourly wages to unemployment.[114] Our empirical specification for the wage equation is in line with the wage bargaining model of Section 2, and controls for the usual demographics that influence wages, as well as a measure of the unemployment rate. Wage curves estimated for the US typically use state-level unemployment as the measure of the cycle, and include both year and state fixed effects, identifying the unemployment elasticity of wages from within-region deviations in unemployment from aggregate trends (Blanchard and Katz, 1992; Hines et al., 2001; Krueger and Summers, 1988). However, this strategy is not empirically feasible for the UK and West Germany, where regional unemployment differentials are highly persistent, making it hard to identify any cyclicality in wages over and above unrestricted time and region effects. As a result, our baseline specifications use national unemployment as a business cycle indicator, and we model underlying productivity growth by linear or quadratic trends. We also present estimates based on regional unemployment, which typically deliver lower wage cyclicality, though the estimates are imprecise.

[114] Blanchflower and Oswald (1994) provide estimates of the wage curve for several OECD countries, and suggest a remarkably stable elasticity of wages to unemployment of -0.1. Their work has been extended to cover more recent US evidence by Devereux and Hart (2006), Hines et al. (2001) and Blanchflower and Oswald (2005). For the UK, Bell et al. (2002) obtain a short-run elasticity of wages to unemployment in the UK around -0.03, and long-run elasticities varying between -0.05 and -0.13. Further work has found that the sensitivity of wages to unemployment in the UK has increased over recent decades (Faggio and Nickell (2005), and Gregg et al. (2014)), and that job movers' wages are more procyclical than stayers' (Devereux and Hart (2006)). For Germany, Blanchflower and Oswald (1994) provide estimates between -0.01 and -0.02 using data from the International Social Survey Programme, and Wagner (1994) finds elasticities between 0 and -0.09 on the SOEP, and slightly higher estimates up to -0.13 on data from the Institute for Employment Research (IAB). Baltagi et al. (2009) estimate dynamic specifications on IAB data and find elasticities consistently lower than -0.1. Ammermüller et al. (2009) use data from the German micro census and suggest a -0.03 upper bound for the elasticity in empirical specifications close to ours.

Our sample period is 1991-2009 for the UK and 1984-2010 for West Germany. The sample includes employees aged 16-65, with non-missing wage information, Descriptive statistics for our wage samples are reported in Table 32 in the Appendix for both the BHPS and the SOEP. Regression results for the UK are presented in Table 25.[115] The dependent variable is the log gross hourly wage, deflated by the aggregate consumer price index. All specifications control for individual characteristics (gender, age, education, job tenure and household composition) and region fixed-effects, and standard errors are clustered at the annual level. Column 1 includes the (log of the) aggregate unemployment rate and a linear trend, and delivers an insignificant impact of unemployment on wages. The unemployment effect becomes significant in column 2, which includes a quadratic trend. This better absorbs non linearities in aggregate productivity growth, while cyclical wage fluctuations are now captured by the unemployment rate, with an elasticity of -0.165. Column 3 introduces individual fixed-effects, and the unemployment elasticity stays virtually unchanged.

Columns 4 and 5 distinguish between wages on new and continuing jobs, by including an interaction term between the unemployment rate and an indicator for the current job having started within the past year. In column 4 the coefficient on the interaction term implies that newly-negotiated wages are 50% more cyclical than wages on continuing jobs, in line with the hypothesis that wages are only infrequently renegotiated. Note, however, that even wages on continuing jobs significantly respond to the state of the business cycle, consistent with some degree of on-the-job renegotiation. But when job fixed effects are included in column 5, the difference in cyclicality between old and continuing wages is much lower and borderline significant. As the excess cyclicality in column 5 is identified by unemployment fluctuations within a job spell, and unemployment is highly persistent, we likely lack power to identify the effect of interest within job spells, which are on average only observed over 2.6 waves. The alternative explanation is that the (permanent) quality of newly-created jobs is procyclical, and when such cyclicality is captured by job fixed-effects the excess cyclicality in newly-negotiated wages is much reduced (see Gertler and Trigari (2009); Gertler et al. (2016) for a similar result for the US). A similar degree of cyclicality in new and continuing jobs is consistent with very high unemployment persistence.

If wages are infrequently renegotiated, the unemployment rate at the start of a job is expected to have a long-lasting impact on the wage while on the same job, over and above the impact of current unemployment. This is tested in column 6, which shows that both starting and current unemployment have a significant impact on wages. Columns 7 and 8 control for lagged unemployment, with or without its current value, and column 9 controls for the lagged dependent variable. In virtually all specifications the wage elasticity to unemployment is negative and significant, and does not fall below -0.17.

[115] The full set of coeficients, including the control variables included in specification of column 2 in Table 24, are reported in Table 33 of the Appendix.

When controlling for regional rather than aggregate unemployment, specifications that also include a quadratic trend deliver a negative and significant unemployment elasticity, but its magnitude in all specifications stays above -0.1, as illustrated in Table 34 in the Appendix. Similarly as for aggregate wage curves, we do find evidence of excess cyclicality of wages on new jobs (column 5), but this falls when job fixed-effects are introduced (column 6).

The corresponding results for West Germany are presented in Table 26. The dependent variable is the log monthly wage, deflated by the consumer price index, and all regressions control for the log of monthly hours worked. The use of monthly, as opposed to hourly, wages is motivated by comparability with the reservation wage regressions presented in the next subsection, as information on reservation wages is only available at the monthly level. The unemployment elasticity of wages in all jobs is markedly lower than in UK estimates, in line with previous estimates for West Germany, and is only significant for new matches (column 4) or when lagged unemployment is used (columns 7 and 8). A clear similarity between West Germany and the UK is that the unemployment elasticity of wages is higher for new hires than for continuing jobs, but such difference becomes not significant when controlling for job fixed-effects (column 5). Estimates based on regional unemployment (Table 35 in the Appendix) are qualitatively similar to those reported in Table 34, but with smaller elasticities throughout.

To summarise evidence on the wage equation, our analysis delivers elasticities of wages with respect to unemployment between -0.1 and -0.17 for the UK, and markedly lower values (often non statistically significant) for West Germany.

19.2 Estimates of the reservation wage curve

The role of reservation wages in business cycle fluctuations is underexplored, and there exists no empirical work on their cyclicality. An obvious reason for this gap in the literature is the scarcity of reservation wage data. For the US, a few studies analyse reservation wage data occasionally collected (Feldstein and Poterba (1984); Holzer (1986b,a); Petterson (1998); Ryscavage (1988)). In recent years the Survey of Unemployed Workers in New Jersey has substantially advanced the empirical study of reservation wages (Krueger and Mueller (2011, 2016, 2012), Hall and Mueller (2018)) but these only cover a span of 24 weeks. Early work on reservation wages for the UK has used cross-section survey data (Lancaster and Chesher (1983), Jones (1988)). Le Barbanchon et al. (2018) investigate the empirical determinants of reservation wages in France, and find that reservation wages do not significantly respond to benefit duration. In the US, no data source has collected reservation wage information on a regular basis for a long period of time, but this is available in both the BHPS and the SOEP.

In the BHPS respondents in each wave 1991-2009 are asked about the lowest weekly take-home pay that they would consider accepting for a job, and about the hours they would expect to work for this amount. Using answers to these questions we construct a measure of the hourly net reservation wage, and deflate it using the aggregate consumer price index. A similar question is asked of SOEP respondents in all waves since 1987, except 1990, 1991 and 1995. The reservation wage information is elicited in monthly terms[116] and is not supplemented by information on expected hours, thus specifications for Germany use monthly reservation wages as the dependent variable, and control for whether an individual is looking for a full-time or part-time job, or a job of any duration.

The working sample includes all individuals with information on reservation wages. In the BHPS the question on reservation wages is asked of all individuals who are out of work in the survey week and are actively seeking work or, if not actively seeking, would like to have a regular job. In the SOEP the same question is asked of all individuals who are currently out of work but contemplate going back to work in the future. Descriptive statistics for the reservation wage samples are reported in Table 32.

Theory implies that reservation wages should respond to three sets of variables. First, as the reservation wage depends on expected wage offers, reservation wage equations should control for factors featuring in wage curves, namely gender, human capital components, regional and aggregate effects, as well as a measure of workers' outside options, proxied by the unemployment rate. As the duration of unemployment affects workers' employability, this should also be controlled for in reservation wage equations. Second, the reservation wage responds to the probability of receiving a wage offer, and therefore to the unemployment rate. Cyclical factors, as captured by the unemployment rate, thus affect the reservation wage via both the probability of receiving an offer and the expected wage offer. Third, the reservation wage depends on the level of utility enjoyed while out of work, which we proxy using available measures of unemployment benefits and family composition.

The estimates for the UK reservation wage equation are reported in Table 27. The dependent variable is the log of the real hourly reservation wage. All specifications control for the same set of individual characteristics as wage equations, having replaced job tenure with the elapsed duration of a jobless spell, and for the amount of benefit income received. In column 1 the state of the business cycle is captured by the (log) national unemployment rate and a linear trend is included. The unemployment coefficient is equal to -0.095 and is significant at the 5% level. Such elasticity rises to -0.175 when a quadratic trend is included in column 2, and slightly declines to -0.146 when individual fixed-effects are introduced in column 3. Columns 4 and 5 control for lagged unemployment, and the associated elasticity is somewhat smaller than the elasticities in specification that only control for current unemployment.

The results from regional reservation wage equations are reported in Table 35 and show that only when one controls for a quadratic trend is the unemployment elasticity

[116]The actual question in German is "Wie hoch müsste der Nettoverdienst mindestens sein, damit Sie eine angebotene Stelle annehmen würden? (im Monat)".

significant. Overall, the elasticity of reservation wages to regional unemployment is markedly lower than the elasticity with respect to aggregate unemployment.

We estimate similar reservation wage specifications for West Germany,[117] and the results are reported in Table 36. While the elasticity of reservation wages with respect to current unemployment is wrongly signed, the elasticity of reservation wages with respect to lagged unemployment has the expected sign and is significant. This result is also replicated on estimates based on regional unemployment (see Table 35)

From estimates of this Subsection we conclude that there is fairly limited cyclicality in reservation wages. Specifications that control for individual fixed-effects deliver a reservation wage elasticity of -0.146 in the UK, and about zero in West Germany (or -0.082 when using lagged unemployment). Such elasticities are very close to the corresponding wage elasticities (respectively: -0.169, about zero, and -0.065, respectively). These estimates are not consistent with the predictions of the canonical model for two reasons. First, the canonical model predicts a reservation wage elasticity close to -0.3 both in the UK and West Germany, and this figure is far outside the range of estimates obtained. Second, the canonical model predicts reservation wages to be more cyclical than wages, while we have empirically established that they display very similar degrees of cyclicality.

19.3 The quality of reservation wage data

One concern in the empirical analysis of reservation wages is that the self-reported reservation wage information used may be of low quality, hence the lack of a strong response to cyclical fluctuations. However, it should be noted that the impact of most covariates considered on reservation wages (e.g. age, education and gender) has the expected sign and is precisely estimated, as shown in Table 33. We further address concerns about the quality of reservation wage data by investigating whether the correlation between reservation wages and job search outcomes has the sign predicted by search theory. Ceteris paribus, a higher reservation wage should cause a longer remaining duration in unemployment and higher entry wages upon job finding.

Table 29 illustrates the effect of reservation wages on each outcome for the UK. Column 1 simply regresses an indicator of whether a worker has found a job in the past year on the reservation wage recorded at the beginning of that year and a set of year and region dummies. The impact of the reservation wage is virtually zero. This estimate

[117] In Germany the duration of unemployment compensation is a nonlinear function of age and previous social security contributions, which are potentially correlated to individual characteristics that also determine wages. We thus exploit nonlinearites in entitlement rules to obtain the number of months to benefit expiry, which is used as an instrument for unemployment benefits in Table 28 (in which age and months of social security contributions feature linearly in all regressions). No instruments are required for the UK reservation wage equations as the duration of benefits in the UK is determined by job search behaviour rather than previous employment history.

is likely to be upward biased due to omitted controls for worker ability, as more able workers have both higher reservation wages and are more likely to find employment. Column 2 controls for the usual individual covariates and the national unemployment rate, and indeed shows that, conditional on such factors, workers with higher reservation wages tend to experience significantly longer unemployment spells. Column 3 shows that this results is robust to the introduction of individual fixed-effects.

Columns 4-6 show the impact of reservation wages on wages for those who find jobs. In column 4, which does not control for individual characteristics, the estimated elasticity of reemployment wages with respect to reservation wages is positive and highly significant, but likely to be upward biased by unobserved individual factors that are associated to both higher reservation wages and higher reemployment wages. Such elasticity falls by about a quarter in column 5, which controls for individual characteristics, and is further halved in column 6, which controls for individual fixed-effects, but remains statistically significant. Similar results for West Germany are presented in Table 30, and they are in line with the UK results, with the qualification that the negative impact of reservation wages on job-finding rates is stronger for West Germany than for the UK. The conclusion from this analysis is that the reservation wage data, though undoubtedly noisy, embody meaningful information about job search behaviour, and there is no particular reason to think that their cyclicality is systematically under-estimated.

20 Reference dependence in reservation wages

We next aim to identify the presence of reference dependence in reservation wages. If past wages shape reference points, which in turn influence reservation wages, we should observe a significant correlation between past wages and reservation wages. While such correlation is consistent with the existence of reference points, it is clearly also consistent with alternative mechanisms. One possible confounding factor is any direct link between unemployment benefits and past wages, as unemployment income is a key component of reservation wages in the canonical model. This is the case for West Germany, where benefit entitlement is a function of age and previous social security contributions, which are in turn directly linked to past wages, implying a positive correlation between past and reservation wages, over and above the role of reference points. By contrast, in the UK, unemployment compensation only varies (quite coarsely) with family composition, and is not directly linked to previous wages, making the UK an ideal case study for reference points in reservation wages. We thus restrict the analysis that follows to the UK.

The second confounding factor is represented by unobserved productivity components of past wages, which are reflected in reservation wages in the canonical model via their effect on the wage offer distribution. Our approach consists in isolating the component of past wages that can be reasonably interpreted as rents – as opposed to productivity – and observe its correlation with reservation wages. A rational worker would not use past rents in forming their current reservation wage (absent wealth effects, which we do not find to be important), whereas a worker who uses past wages as a reference point might do so.

Let's consider a simple empirical model for the reservation wage:

$$\ln \rho_{it} = \beta_1 X_{it} + \beta_2 w_i^* + \beta_3 R_{it-d} + \varepsilon_{it}, \tag{63}$$

where X_{it} denotes observable characteristics, w_i^* denotes worker ability, and R_{it-d} denotes the level of rents in the last job observed (d periods ago). The coefficient of interest is β_3, indicating whether rents lost with past jobs influence current reservation wages.

Assume the following model for the last observed wage:

$$\ln w_{it-d} = \gamma_1 X_{it-d} + w_i^* + R_{it-d} + u_{it-d}. \tag{64}$$

If one regresses the reservation wage on the last observed wage as in:

$$\ln \rho_{it} = \delta_1 X_{it} + \delta_2 \ln w_{it-d} + \varepsilon_{it}, \tag{65}$$

the OLS estimate for δ_2 would capture the effect of both unobserved heterogeneity and rents on the reservation wage, and is possibly attenuated by the presence of measurement error in past wages. Identification of the effect of interest would require an instrument that represents a significant component of past rents, while being orthogonal to worker ability.

As a proxy for the size of rents in a given job we use industry affiliation, in line with a long-established literature concluding that part of inter-industry wage differentials represent rents (see the classic papers, Krueger and Summers (1988), and Gibbons and Katz (1992), and Benito (2000), and Carruth et al. (2004), for British evidence). Specifically, we use as an instrument for previous wages the predicted, inter-industry wage differential obtained on an administrative dataset, the Annual Survey of Hours and Earnings (ASHE), whose sample size allows us to control for industry affiliation at the 4-digit level. We estimate a log wage equation for 1982-2009 on ASHE, controlling for 4-digit industry effects, unrestricted age effects, region, and individual fixed effects. The inclusion of individual fixed effects allows us to capture the component of inter-industry wage differentials that is uncorrelated to individual unobservables, and is thus key to justify our exclusion restriction. We then match the estimated industry effects to individual records in the BHPS, and use them as an instrument for last observed wages in reservation wage regressions.

Having controlled for unobserved heterogeneity in the construction of our instrument, the exclusion restriction would still be violated in the presence of wealth effects in job search behaviour (see for example Shimer and Werning (2007), for a model of job search with asset accumulation). Rents received in previous jobs would have an impact on asset accumulation, in turn affecting worker utility during unemployment and reservation wages. This does not seem to be a major issue in our working sample, in which more than three quarters of unemployed workers have no capital income, and another 11% have capital income below £100 per year. But in order to control for wealth effects, if any, we include indicators for household assets and housing tenure in the estimated reservation wage equations.

Past wages can be obtained for currently unemployed respondents who had previous employment spells over the BHPS sample period. For those who are observed in employment at any of the previous interview dates, we use contemporaneous information on their last observed job. For those who are not observed in employment at any interview date, but had between-interview employment spells, we use the most recent retrospective information on previous jobs. Retrospective employment information is typically more limited than contemporaneous information, and in particular it does not cover working hours. The analysis that follows is thus entirely based on monthly wages and reservation wages.

Our results are reported in Table 31. Column 1 reports OLS estimates of the reservation wage equation, controlling for the last observed wage in the BHPS panel. The sample is smaller than the original sample of Table 27, as for about 45% of the reservation wage sample we do not observe any previous job in the BHPS panel. The coefficient on the wage in the last job is, unsurprisingly, positive and highly significant. The specification in column 2 introduces individual fixed-effects, and the coefficient on the lagged wage is markedly reduced, as part of the observed association between current reservation wages and past wages is driven by unobserved worker quality. Column 3 allows for some gradual decay of the influence of past wages on reservation wages, controlling for the interaction between the past wage and the number of years since it was observed. The coefficient on the interaction term implies that the influence of previous wage realisations on current reservation wages should vanish about 4 years after job loss, although this effect is only significant at the 10% level.

Column 4 instruments the previous wage with its rent component, as proxied by the 4-digit industry level differential, and shows that this has a positive and significant impact on the reservation wage, consistent with a model in which previous rents affect workers' reference points during job search. The IV coefficient on the past wage is higher than the OLS coefficient, due to the presence of transitory components, (classical) measurement error, and unobserved compensating differentials in the last observed wage (see also Manning (2003), chapter 6). The specification in Column 5 introduces individual fixed-effects, and the coefficient of interest is now identified by the sub-sample of individuals with multiple unemployment spells originating from different 4-digit industries. Unlike in the OLS model, the coefficient on the lagged wage remains very close to the one obtained without fixed-effects in column 4. Once lagged wages are instrumented by inter-industry wage differentials, their impact on current wage aspirations is no longer confounded by unobserved ability. Indirectly, this signals that unobserved ability is not driving the (very disaggregate) industry allocation of individuals, confirming the validity of the instrument. Column 6 allows for changes in reference points over time, but the decay effect is no longer significant. In summary, the finding that rents in previous jobs affect reservation wages is not consistent with the determination of reservation wages in the canonical model, but is instead consistent with a model in which reference wages influence reservation wages, and these reference wages are, in part, influenced by past wages.

20.1 Quantitative predictions of reference dependence

Using estimates from the previous subsections, we consider whether there exists a combination of backward-looking behaviour in wage setting and reservation wages, summarized by a triple of parameter values $(\alpha, \alpha_\rho, \alpha_l)$, that yields quantitative predictions close to our empirical findings. The data moments we use to nail down the values of these three parameters are: (i) the coefficient on lagged wages in the determination of reservation wages (0.15, from column 6 in Table 31); (ii) the elasticity of wages with respect to unemployment (-0.17, from column 3 of Table 25); (iii) the elasticity of reservation wages with respect to unemployment (-0.15, from column 3 of Table 27). Specifically, we impose $(1-\alpha_\rho)\alpha_l = 0.15$, as $(1-\alpha_\rho)\alpha_l$ is the coefficient on lagged wages in the reservation wage equation 42, and then select combinations of (α, α_ρ) that produce, in correspondence of baseline parameters used in Section 3, an elasticity of wages and reservation wages with respect to unemployment within 0.02 of -0.17 and -0.15, respectively. Figure 36 plots values of (α, α_ρ) that satisfy these criteria. Two clear points emerge. First, only values of α_ρ in the range 0.40-0.55 meet the above criteria and, second, once α_ρ lies in this range, almost any value of α meets the criteria.

This reinforces our earlier point that the degree of backward-looking behaviour in wage-setting has virtually no bite on the predicted cyclicality of wages, while we note here that this is instead quite sensitive to the extent of reference dependence in reservation wages. A model in which between 45% and 60% of the variation in reservation wages is driven by backward-looking reference points is able to match well the observed cyclicality of average wages and reservation wages and address wage (and reservation wage) flexibility puzzles

To conclude, we discuss alternative mechanisms that could potentially reduce the cyclicality of reservation wages and thus address the wage flexibility puzzle, even in the absence of reference dependence. In Appendix 22.E we consider two further alternatives to the canonical search model, based on on-the-job search and hyperbolic discounting, respectively, and find that neither option adequately addresses the wage flexibility puzzle. In particular, a model with on-the-job search would deliver even more strongly cyclical reservation wages than the canonical model in correspondence of plausible parameter values, while a model with hyperbolic discounting would deliver less cyclical reservation wages, but at the cost of making them insensitive to the expected wage, while Table 33 shows that wages and reservation wages respond in very similar ways to most covariates considered.

Learning could be an alternative source of backward-looking behaviour in reservation wages, which may have observationally equivalent consequences to our idea of reference-dependence. In a random search model, forward-looking agents would interpret rents in previous jobs as a purely random component, unrelated to their prospective wage offer distribution, and the impact of such rents on the reservation wage is precisely the mechanism on which hinge our conclusions regarding reference dependence. But if workers are trying to learn about where jobs with high rents can be found and can direct their job search towards high-rent industries, wages in previous jobs may convey useful information about future job opportunities and hence influence reservation wages. In this case, both reference-dependence and learning would introduce backward-looking components in reservation wages, and this is the element that we emphasize as central for addressing the wage flexibility puzzle, while being more agnostic about the exact source of such backward-looking behaviour.

21 Conclusions

We propose a search model with infrequent wage negotiations and reference dependence in reservation wages to derive a relationship between wages and unemployment – the wage curve – which is unaffected by demand shocks. The slope of this curve is an estimate of the relative variability of wages and unemployment in response to demand shocks. Absent reference dependence, we show that the model can only explain the modest pro-cyclicality of wages if replacement ratios are implausibly high, unemployment persistence implausibly low or labour contracts implausibly long. A further model prediction is that reservation wages should be more strongly cyclical than wages, because they embody cyclicality from both expected wage offers and the probability of receiving an offer. We next show next that the introduction of reference dependence in reservation wages – based on backward-looking reference points – can deliver mildly cyclical wages and reservation wages for plausible value of other model parameters.

We turn to individual data for the UK and West Germany and find that within each country wages and reservation wages display very similar degrees of cyclicality, substantially lower than the one predicted by the canonical model without reference dependence. We provide evidence that reservation wages significantly respond to backward-looking reference points, as proxied by rents earned in previous jobs.

In a model calibration we show that backward-looking reference dependence in reservation wages markedly reduces the predicted cyclicality of both wages and reservation wages and can reconcile theoretical predictions of search models with the observed cyclicality of wages and reservation wages.

22 Appendix III: Reservation Wages and the Wage Flexibility Puzzle

22.A Tables

Table 24: Benchmark Parameters for the UK and West Germany

Symbol	Interpretation	UK	Germany	Source
s	Separation rate	0.010	0.012	Quarterly LFS (UK), SOEP data (Germany)
u	Unemployment rate	0.067	0.078	Official unemployment rate
λ	Job-finding rate	0.139	0.142	Separation rate and unemployment rate: $\lambda = s(1-u)/u$
ξ	Persistence in unemployment	0.003	0.004	AR(1) estimates on monthly series for unemployment rate
ϕ	Frequency of wage renegotiations	0.083	0.083	Annual frequency (Taylor (1999), Gottshalk 2015, Fabiani et al. (2010))
r	Interest rate	0.003	0.003	Conventional value
η	Replacement rate	0.690	0.754	For UK: equation 49, using $\rho^*/w^* = 0.80$ (from BHPS). For Germany: benefit replacement ratio + extra utility of leisure during unemployment as implied by UK estimates.
β	Bargaining power of workers	0.050	0.050	Manning (2011) – Table 4

Notes. s, λ, ξ, ϕ and r are expressed in monthly terms.

Table 25: Estimates of a Wage Equation for the UK, 1991-2009.

	Dependent variable: log hourly wage								
	1	2	3	4	5	6	7	8	9
Log wage, lagged									0.102**
									(0.046)
Log unemployment rate	-0.022	-0.165***	-0.169***	-0.147***	-0.109***	-0.137***	-0.022		-0.150***
	(0.032)	(0.044)	(0.044)	(0.042)	(0.029)	(0.039)	(0.070)		(0.009)
Log unemployment rate * new job				-0.075***	-0.019				
				(0.013)	(0.011)				
Log unemployment rate, at start of job						-0.069***			
						(0.013)			
Log unemployment rate, lagged							-0.113**	-0.126**	
							(0.050)	(0.032)	
Trend	linear	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic
Individual fixed effects			✓	✓		✓	✓	✓	✓
Job fixed effects					✓				
Observations	96,270	96,270	92,380	92,380	77,854	91,712	92,380	92,380	53,054
R-squared	0.397	0.397	0.810	0.810	0.889	0.810	0.778	0.778	

Notes. The sample includes employees aged 18-65 with non-missing wage information. The dependent variable is the log gross hourly wage, deflated by the aggregate consumer price index. Estimation method: OLS in columns 1-8; Arellano Bond (1991) estimator for dynamic panel data models in column 9. The unemployment concept is national. All regressions include a gender dummy, age and its square, three education dummies, a cubic trend in job tenure, a dummy for married, the number of children in the household, and eleven region dummies. Regressions in columns 4 and 5 also include a dummy for the job having started in the previous 12 months. Standard errors are clustered at the year level in columns 1 and 2, and using 2-way cluster-robust variance (Cameron and Miller (2015)) in columns 3-9. Source: BHPS.

Table 26: Estimates of a Wage Equation for the West Germany, 1984-2010

	Dependent variable: Log monthly wage								
	1	2	3	4	5	6	7	8	9
Log wage, lagged									0.390***
									(0.027)
Log unemployment rate	0.082	0.002	-0.028	-0.015	-0.005	-0.023	0.070**		-0.015
	(0.048)	(0.025)	(0.019)	(0.018)	(0.015)	(0.019)	(0.025)		(0.030)
Log unemployment rate * new job				-0.096***	0.034				
				(0.026)	(0.022)				
Log unemployment rate, at start of job						-0.011			
						(0.008)			
Log unemployment rate, lagged							-0.120***	-0.065***	
							(0.024)	(0.018)	
Trend	linear	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic
Individual fixed effects			✓	✓		✓	✓	✓	✓
Job fixed effects					✓				
Observations	166,614	166,614	161,075	160,865	149,617	161,075	161,075	161,075	101,526
R-squared	0.649	0.651	0.415	0.415	0.199	0.415	0.415	0.415	

Notes: The sample includes employees aged 18-65 with non-missing wage information. The dependent variable is the log gross monthly wage, deflated by the aggregate consumer price index. Estimation method: OLS in columns 1-8; Arellano Bond (1991) estimator for dynamic panel data models in column 9. The unemployment concept is national. All regressions include a gender dummy, age and its square, three education dummies, a cubic trend in job tenure, a dummy for married, the number of children in the household, and eleven region dummies. Regressions in columns 4 and 5 also include a dummy for the job having started in the previous 12 months. Standard errors are clustered at the year level in columns 1 and 2, and using 2-way cluster-robust variance (Cameron and Miller (2015)) in columns 3-9. Source: SOEP.

Table 27: Estimates of a Reservation Wage Equation for the UK, 1991-2009.

Dependent variable: log hourly reservation wage					
	1	2	3	4	5
Log unemployment rate	-0.095**	-0.175***	-0.146***	0.010	
	(0.046)	(0.058)	(0.042)	(0.146)	
Log unemployment rate, lagged				-0.119	-0.112***
				(0.096)	(0.026)
Trend	linear	quadratic	quadratic	quadratic	quadratic
Individual fixed effects			✓	✓	✓
Observations	14,874	14,874	10,774	10,774	10,774
R-squared	0.248	0.249	0.614	0.614	0.614

*Notes:*The sample includes unemployed jobseekers with non-missing reservation wage information. The dependent variable is the log net hourly reservation wage, deflated by the aggregate consumer price index. Estimation method: OLS. The unemployment concept is national. All regressions also include a gender dummy, age and its square, three education dummies, a cubic trend in unemployment duration, a dummy for married, the number of children in the household, the log of unemployment benefits, a dummy for receipt of housing benefits, and eleven region dummies. Standard errors are clustered at the year level in columns 1 and 2, and using 2-way cluster-robust variance (Cameron and Miller (2015)) in columns 3-5. Source: BHPS.

Table 28: Estimates of a Reservation Wage Equation for West Germany, 1987-2010.

Dependent variable: log monthly reservation wage					
	1	2	3	4	5
Log unemployment rate	0.173**	0.001	0.038	0.175**	
	(0.070)	(0.065)	(0.054)	(0.070)	
Log unemployment rate, lagged				-0.196***	-0.082*
				(0.064)	(0.045)
Trend	linear	quadratic	quadratic	quadratic	quadratic
Individual fixed effects			✓	✓	✓
Observations	11,221	11,221	7,911	7,911	7,911
R-squared	0.414	0.418	0.123	0.125	0.123

Notes: See notes to Table 32 for the sample used. The dependent variable is the log net monthly reservation wage, deflated by the aggregate consumer price index. Estimation method: IV. All regressions also include a gender dummy, age and its square, three education dummies, a cubic trend in unemployment duration, a dummy for married, the number of children in the household, the log of unemployment benefits, a dummy for receipt of housing benefits, controls for whether an individual looks for full-time, part-time or any job (the omitted category being "unsure about preferences"), months of social insurance contributions and eleven region dummies. Unemployment benefits are instrumented by months to benefit expiry. These are obtained by exploiting benefit entitlement rules, based on (nonlinear) functions of age and previous social security contributions. Standard errors are clustered at the year level in columns 1 and 2, and using 2-way cluster-robust variance (Cameron and Miller (2015)) in columns 3-5. Source: SOEP.

Table 29: Reservation Wages, Post-Unemployment Wages and Job Finding Probabilities in the UK, 1991-2009

	Whether found job			Log post-unemployment wage		
	1	2	3	4	5	6
Log reservation wage	-0.001	-0.020**	-0.020**	0.436***	0.312***	0.157**
	(0.008)	(0.008)	(0.011)	(0.021)	(0.036)	(0.080)
Year dummies	✓			✓		
Trend	no	quadratic	quadratic	no	quadratic	quadratic
Further controls		✓	✓		✓	✓
Individual fixed-effects			✓			✓
Observations	15,278	14,701	10,642	2,685	2,594	602
R-squared	0.018	0.078	0.039	0.217	0.299	0.290

Notes: See notes to Table 27 for the sample used. Estimation method: OLS. All specifications include eleven region dummies. Further controls in columns 2, 3, 5 and 6 are a gender dummy, age and its square, three education dummies, a cubic trend in unemployment duration, a dummy for married and the number of children in the household. Standard errors are clustered at the year level in columns 1, 2, 4 and 5; and using 2-way cluster-robust variance (Cameron and Miller (2015)) in columns 3 and 6. Source: BHPS.

Table 30: Reservation Wages, Post-Unemployment Wages and Job Finding Probabilities in West Germany, 1987-2010

	Whether found job			Log post-unemployment wage		
	1	2	3	4	5	6
Log reservation wage	0.033***	-0.081***	-0.100***	0.737***	0.391***	0.123
	(0.007)	(0.011)	(0.016)	(0.023)	(0.034)	(0.106)
Year dummies	✓			✓		
Trend	no	quadratic	quadratic	no	quadratic	quadratic
Further controls		✓	✓		✓	✓
Individual fixed-effects			✓			✓
Observations	11,534	11,534	8,156	2,984	2,984	755
R-squared	0.007	0.071	0.033	0.244	0.348	0.127

Notes: See notes to Table 28 for the sample used. Estimation method: OLS. All specifications include eleven region dummies. Further controls in columns 2,3 5 and 6 are a gender dummy, age and its square, three education dummies, a cubic trend in unemployment duration, a dummy for married, the number of children in the household, whether an individual looks for a full-time, part-time or any job (the omitted category is "unsure about preferences"). Standard errors are clustered at the year level in columns 1, 2, 4 and 5; and using 2-way cluster-robust variance (Cameron and Miller (2015)) in columns 3 and 6. Source: SOEP.

Table 31: Reservation wages and rents in previous jobs: UK, 1991-2009.

Dependent variable: log monthly reservation wage						
	1	2	3	4	5	6
Estimation method	OLS	OLS	OLS	IV	IV	IV
Last observed log wage	0.083***	0.033***	0.042***	0.133***	0.149**	0.153***
	(0.005)	(0.010)	(0.011)	(0.018)	(0.063)	(0.067)
Last observed log wage			-0.011*			-0.002
* years since observed			(0.006)			(0.009)
Log unemployment rate	-0.183***	-0.173***	-0.174***	-0.159*	-0.177**	-0.166*
	(0.081)	(0.075)	(0.075)	(0.084)	(0.067)	(0.078)
Trend	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic
Individual fixed effects		✓	✓		✓	✓
Observations	8,091	5,737	5,737	7,732	5,520	5,520
R-squared	0.284	0.098	0.099			
First stage, F-test$^{(a)}$				908.9	908.9	53.7
First stage, F-test$^{(b)}$						64.2

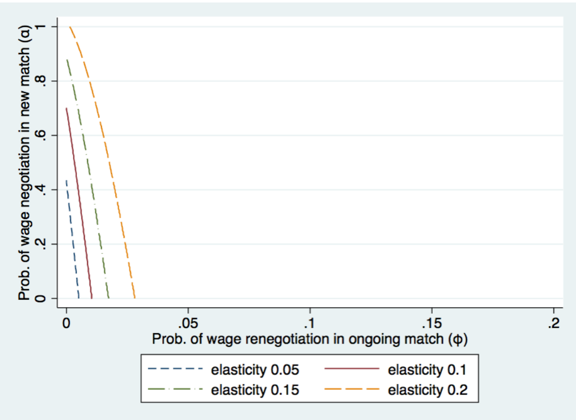
Notes: See notes to Table 27 for the sample used. All regressions also include a gender dummy, age and its square, three education dummies, a cubic trend in the number of years since the last job was observed, a dummy for married, the number of children in the household, the log of unemployment benefits, three dummies for capital income (0,<100£, 100£+ per year, where the excluded category is "don't know"), three dummies for housing tenure (owned with mortgage, local authority rented, other rented, where the excluded category is outright owned) and eleven region dummies. Instruments used: predicted industry wage (4-digit) for previous job (columns 4 and 5); predicted industry wage (4-digit) for previous job and its interaction with years since previous job (column 6). (a) denotes Sanderson and Windmeijer (2016) first-stage F-statistic for the first equation (last observed log wage) and (b) denotes the corresponding statistic for the second equation (last observed log wage*years since observed). Standard errors are clustered at the year level in columns 1-2 and 4-5, and using 2-way cluster-robust variance (Cameron and Miller (2015)) in columns 3 and 6. Source: BHPS.

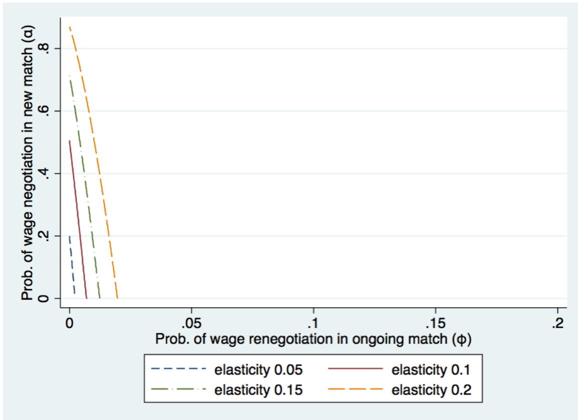
22.B Figures

Figure 34: The role of backward looking behaviour in wage setting $(1-\alpha)$ and frequency of wage renegotiations (ϕ) in wage and reservation wage cyclicality. UK parameter values.

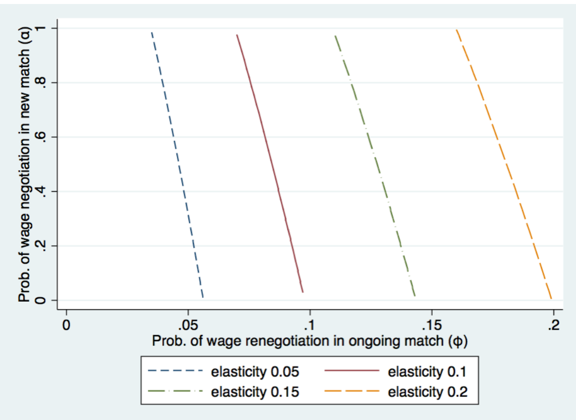
Wage Elasticity **Reservation Wage Elasticity**

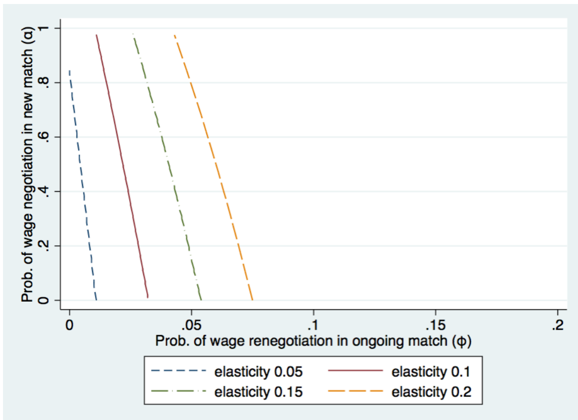
(a) $\xi = 0.004(estimated\ value)$





(b) $\xi = 0.01(counterfactual\ persistence)$



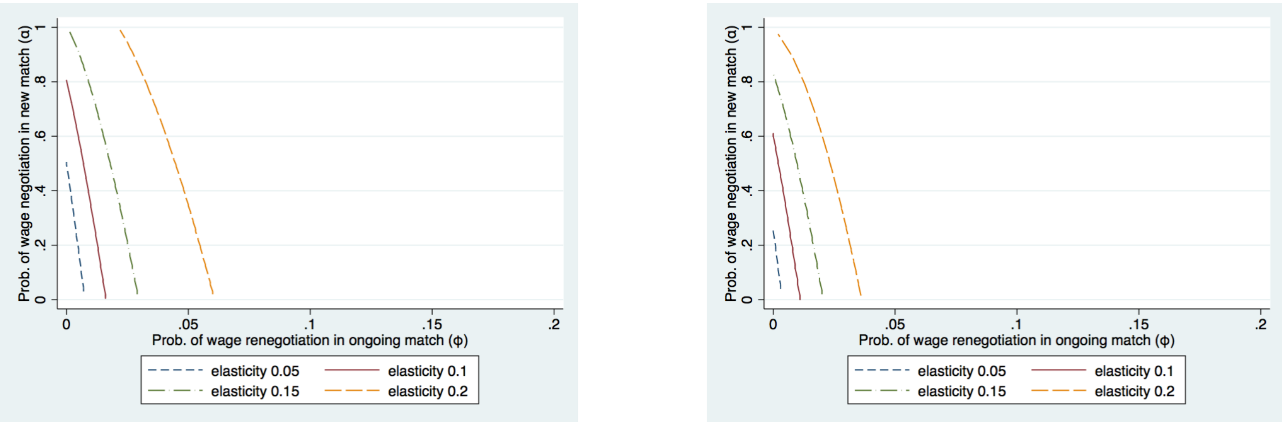


Notes: The curves represent combinations of ϕ and α that deliver indicated values of the elasticity of average wages or reservation wages to unemployment, in correspondence of alternative values of unemployment persistence.

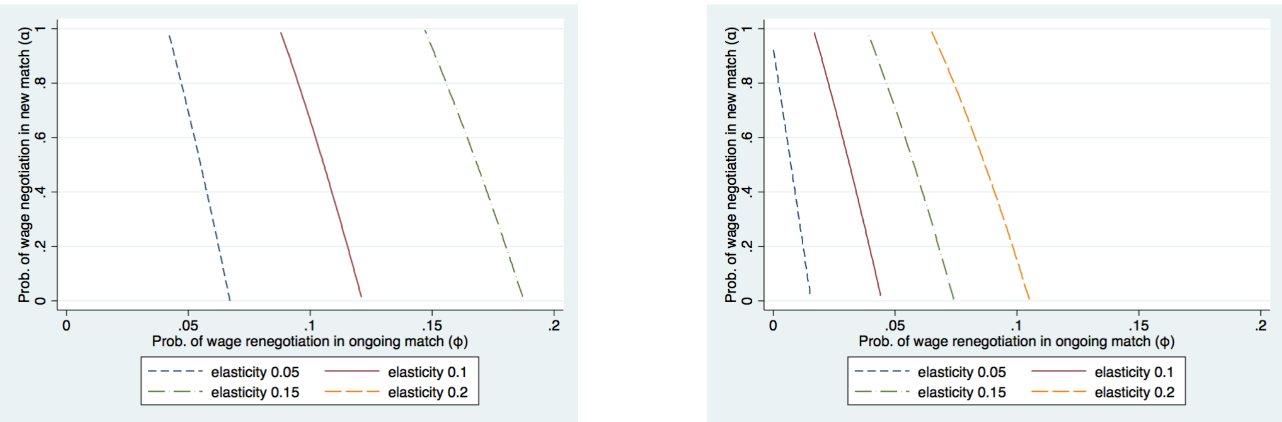
Figure 35: The role of backward looking behaviour in wage setting $(1-\alpha)$ and frequency of wage renegotiations (ϕ) in wage and reservation wage cyclicality. West Germany parameter values.

Wage Elasticity **Reservation Wage Elasticity**

(a) $\xi = 0.004(estimated\ value)$



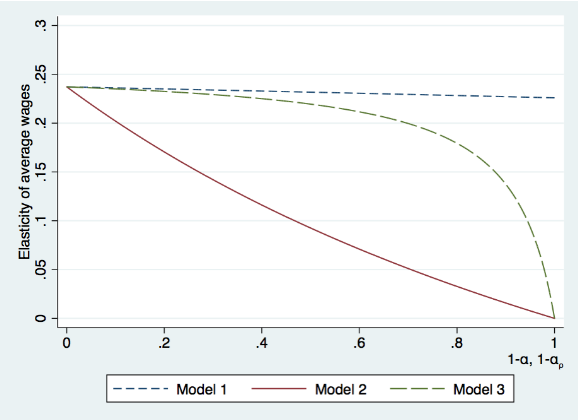
(b) $\xi = 0.01(counterfactual\ persistence)$



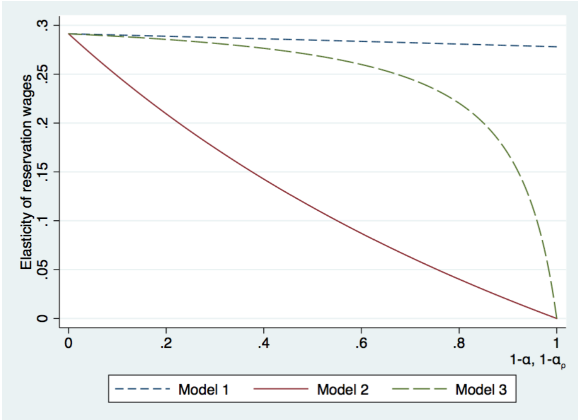
Notes: The curves represent combinations of ϕ and α that deliver indicated values of the elasticity of average wages or reservation wages to unemployment, in correspondence of alternative values of unemployment persistence.

Figure 36: The role of reference dependence in reservation wages ($1 - \alpha_\rho$) on wage and reservation wage cyclicality. UK parameter values.

(a)



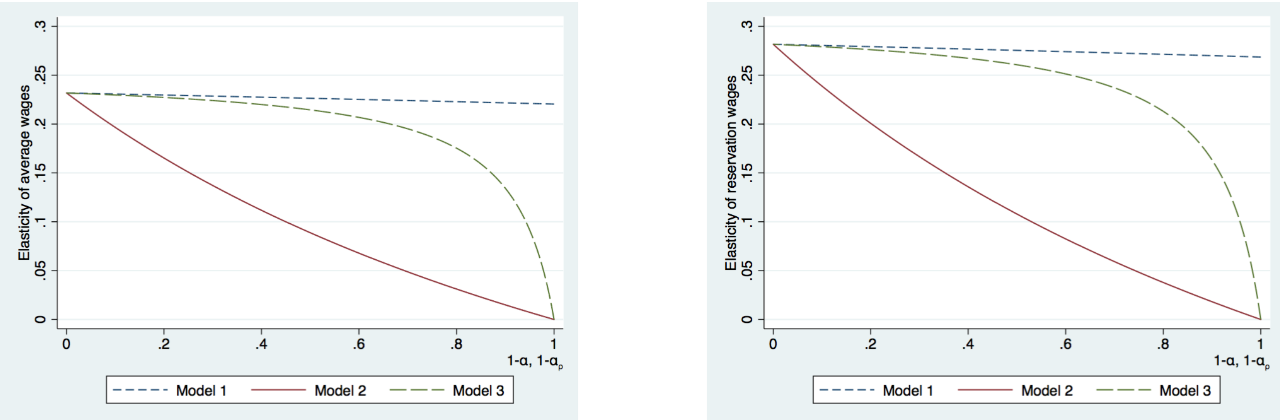
(b)



Notes: The curves represent predicted elasticities of wages and reservation wages with respect to α and α_ρ. **Model 1** is a special case without reference dependence in reservation wages ($\alpha_\rho = 1$), and the running variable on the horizontal axis is $1 - \alpha$, measuring backward looking behaviour in wage setting. **Model 2** is a special case without backward looking behaviour in wage setting ($\alpha = 1$), and the running variable on the horizontal axis is $1 - \alpha_\rho$, measuring reference dependence in reservation wages, under the assumption that the reference wage is completely acyclical ($a_l = 0$). **Model 3** is a special case without backward looking behaviour in wage setting ($\alpha = 1$), and the running variable on the horizontal axis is again $1 - \alpha_\rho$, under the assumption that the reference wage is as cyclical as the average wage ($\alpha_l = 1$). All other parameter values used are described in Section 3C and Table 24.

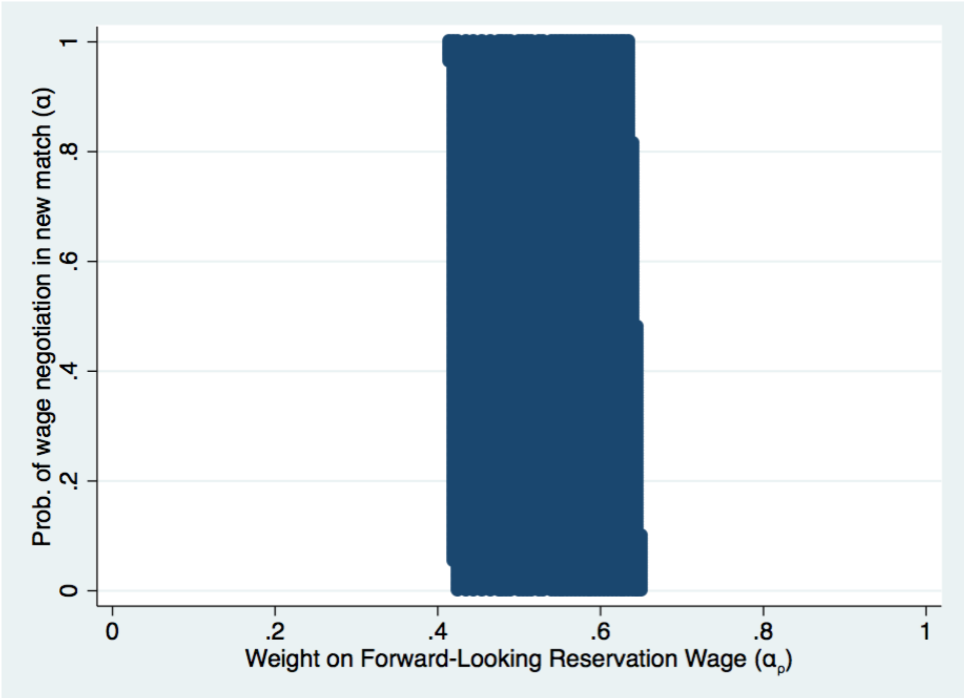
Figure 37: The role of reference dependence in reservation wages ($1 - \alpha_\rho$) on wage and reservation wage cyclicality.West Germany parameter values.

(a) (b)



Notes: The curves represent predicted elasticities of wages and reservation wages with respect to α and α_ρ. **Model 1** is a special case without reference dependence in reservation wages ($\alpha_\rho = 1$), and the running variable on the horizontal axis is $1 - \alpha$, measuring backward looking behaviour in wage setting. **Model 2** is a special case without backward looking behaviour in wage setting ($\alpha = 1$), and the running variable on the horizontal axis is $1 - \alpha_\rho$, measuring reference dependence in reservation wages, under the assumption that the reference wage is completely acyclical ($a_l = 0$). **Model 3** is a special case without backward looking behaviour in wage setting ($\alpha = 1$), and the running variable on the horizontal axis is again $1 - \alpha_\rho$, under the assumption that the reference wage is as cyclical as the average wage ($\alpha_l = 1$). All other parameter values used are described in Section 3C and Table 24.

Figure 38: Parameter Values that Explain Observed Cyclicality



Notes: The shaded region shows the combinations of α, the probability of negotiating a wage on a new match, and α_ρ, the weight on the forward-looking reservation wage, that predict a wage elasticity and a reservation wage elasticity within 0.02 of 0.17 and 0.15, respectively, for a predicted sensitivity of the reservation wage to the lagged wage $(1-\alpha_\rho)\alpha_l$, equal to 0.15 (taken from column 5 of Table 31). All other parameter are set at baseline values.

22.C Derivation of model results

Derivation of the wage curve

Employers and workers

As the value function (34) is linear in wages, the first expectation term in the value function of a vacant job (35) can be replaced by averages, i.e.:

$$
\begin{aligned}
rV(t) &= -c(t) + q(t)[\alpha J(w_r(t); w_l(t), t) \\
&+ \big(1-\alpha\big) J(w(t); w_l(t), t) - V(t) - C(t)] + E_t \frac{\partial V(t)}{\partial t}
\end{aligned} \tag{66}
$$

Free entry of vacancies ensures $V(t) = 0$, so equation (66) can be rearranged to give:

$$
\alpha J(w_r(t); w_l(t), t) + (1-\alpha) J(w(t); w_l(t), t) = C(t) + \frac{c(t)}{q(t)}, \tag{67}
$$

i.e. the expected value of a newly-filled job equals the total expected cost of filling a vacancy, including the fixed cost $C(t)$ and the per period cost $c(t)$, multiplied by the expected duration of a vacancy, $1/q(t)$. Using derivatives in equation 34 and equation 69, equation 67 can be rewritten as:

$$
\begin{aligned}
J(w_r(t); w_l(t), t) &= C(t) + \frac{c(t)}{q(t)} - \frac{(1-\alpha)\big(w(t) - w_r(t)\big)}{r + \phi + s} \\
&= \mu(t) - \frac{(1-\alpha)\big(w(t) - w_r(t)\big)}{r + \phi + s},
\end{aligned} \tag{68}
$$

where $\mu(t) \equiv c(t)/q(t) + C(t)$.

Equations (34) and (36) imply

$$
\frac{\partial W(w_i; w_{il}, t)}{\partial w_i} = -\frac{\partial J(w_i; w_{il}, t)}{\partial w_i} = \frac{1}{r + \phi + s}. \tag{69}
$$

As (r, ϕ, s) are constant, equation 69 implies that $W(w_i; w_{il}, t)$ and $J(w_i; w_{il}, t)$ are separable in w_i and (w_{il}, t).[118]

[118] At this point we have not proved additive separability, but assume it will lead to no contradiction in later analysis, so this can be thought of as a “guess and verify” approach.

Finally, equation (34) further implies:[119]

$$\frac{\partial J(w_i;w_{il},t)}{\partial w_{il}} = \frac{\phi}{(r+s)}\frac{\partial J(w_{ir}(w_{il},t);w_{il},t)}{\partial w_i}\frac{\partial w_{ir}}{\partial w_{il}} = -\frac{\phi}{(r+s)(r+\phi+s)}\frac{\partial w_{ir}}{\partial w_{il}}. \tag{70}$$

Wage determination

The Nash maximand (38) implies

$$\begin{aligned}(1-\beta)&\frac{\partial J(w_{ir}(w_{il},t);w_{il},t)}{\partial w_{ir}}[W(w_{ir}(w_{il},t);w_{il},t)-W(\rho(w_{il},t);w_{il},t)]\\ +\beta&\frac{\partial W(w_{ir}(w_{il},t);w_{il},t)}{\partial w_{ir}}[J(w_{ir}(w_{il},t);w_{il},t)-V(t)]=0.\end{aligned} \tag{71}$$

Using equation (69), (71) can be rewritten as:

$$\begin{aligned}&[W(w_{ir}(w_{il},t);w_{il},t)-W(\rho(w_{il},t);w_{il},t)]\\ &=\tilde{\beta}[J(w_{ir}(w_{il},t);w_{il},t)-J(w_r(t);w_l(t),t)+J(w_r(t);w_l(t),t)],\end{aligned} \tag{72}$$

where $\tilde{\beta}\equiv\beta/(1-\beta)$. Using (69), (70) and (68), equation (72) can be written as:

$$\begin{aligned}&w_{ir}(w_{il},t)=\rho(w_{il},t)+\\ &\tilde{\beta}\left[w_r(w_l,t)-w_{ir}(t)-\frac{\phi(w_{il}(t)-w_l(t))}{r+s}\frac{\partial w_{ir}}{\partial w_{il}}+(r+\phi+s)\mu(t)-(1-\alpha)(w(t)-w_r(t))\right].\end{aligned} \tag{73}$$

Equation 73 implies that newly-negotiated wages depend on past wages if and only if reservation wages depend on them. Past wages are heterogeneous outside steady state if there is infrequent wage renegotiation, thus reference-dependence introduces some heterogeneity in newly-negotiated wages out of steady state. But linearity of value functions implies that means of all variables of interest are sufficient statistics for the working of our model, and one does not need to keep track of heterogeneity out of steady state. Taking means of equation 73 leads to the wage equation (39).

[119] This expression includes $\partial w_{ir}/\partial w_{il}$ which will turn out to be a constant – see equations (73) and (77) below.

The reservation wage

Exploiting the linearity of (37) in wages, we can again replace expectations with averages and obtain

$$rU(w_{il},t) = z + \lambda(t)\left[\alpha W(w_r(w_l,t);w_l,t) + (1-\alpha)W(w(t);w_l,t) - U(w_{il},t)\right] + E_t\frac{\partial U(w_{il},t)}{\partial t}. \tag{74}$$

Evaluating equations (74) and (36) at $\rho_i^o(w_{il},t)$ yields, after some rearrangement, the following expression for the optimal, forward-looking reservation wage:

$$\begin{aligned}\rho_i^o(w_{il},t) &= z + [\lambda(t)-\phi]\left[W(w_r(w_l,t);w_l,t) - W(\rho_i^o(w_{il},t);w_{il},t)\right]\\ &+ \lambda(t)\,[1-\alpha]\left[W(w(t);w_l,t) - W(w_{ir}(w_{il},t);w_{il},t)\right] + E_t\left[\frac{\partial U(w_{il},t)}{\partial t} - \frac{\partial W(\rho_i^o(w_{il},t);w_{il},t)}{\partial t}\right].\end{aligned} \tag{75}$$

$W(\rho_i^o(w_{il},t);w_{il},t) = U(w_{il},t)$ further implies:

$$\frac{\partial U(w_{il},t)}{\partial t} = \frac{\partial W(\rho_i^o(w_{il},t);w_{il},t)}{\partial t} + \frac{\partial W(\rho_i^o(w_{il},t);w_{il},t)}{\partial w_i}\frac{\partial \rho_i^o(w_{il},t)}{\partial t} \tag{76}$$

Using equations 69 and (76), equation (75) can be rearranged to give equation (40).

The assumption about reference dependence outlined in the text implies that the deviation of the reservation wage $\rho(w_{il},t)$ from its steady-state value ρ^* can be expressed as

$$\rho(w_{il},t) - \rho^* = \alpha_\rho(\rho_i^o(w_{il},t) - \rho^*) + (1-\alpha_\rho)\alpha_l(w_{il} - w^*), \tag{77}$$

Taking averages of equation (77) leads to equation (41) for the average reservation wage.

General Methodology

Our approach to deriving the sensitivity of wages with respect to unemployment can be understood in a general way. Let's denote the vector of all relevant variables by x. Our model of the economy can, after linearization, be represented by a first-order differential equation:

$$\Omega[x(t) - x^*] + \Phi\frac{dx(t)}{dt} = 0 \tag{78}$$

for some matrices (Ω, Φ), where x^* denotes steady-state values (whose time derivative is zero in steady-state). Typically, some of the derivative terms will be forward-looking 'jump' variables (thus the derivative is an expected future change) and some will be backward-looking 'state' variables (thus the derivative is a realized change): for simplicity of notation we do not make this distinction explicit.

Let's now consider the linear projection of $x(t)$ on X(t) (where X might be one element of x), and denote it by:

$$x(t) - x^* = \theta_x (X(t) - X^*) \tag{79}$$

And, similarly, denote the linear projections of $dx(t)/dt$ and $dX(t)/dt$ on X by:

$$\frac{dx(t)}{dt} = \theta_{dx}(X(t) - X^*) \tag{80}$$

and

$$\frac{dX(t)}{dt} = \theta_{dX}(X(t) - X^*) \tag{81}$$

respectively. Differentiating equation 79 leads to

$$\frac{dx(t)}{dt} = \theta_x \frac{dX(t)}{dt} \tag{82}$$

and, using equation 81, gives

$$\frac{dx(t)}{dt} = \theta_x \theta_{dX}(X(t) - X^*), \tag{83}$$

which finally gives

$$\frac{\theta_{dx}}{\theta_x} = \theta_{dx} \tag{84}$$

Equation 85 implies that θ_{dx}/θ_x must be the same for every variable in the system. One corollary of this is that the change in every variable in the system can be written as the same function of its own deviation from the steady-state equilibrium. For a forward-looking variable this would take the form:

$$\frac{dx(t)}{dt} = \theta_{dX}(X(t) - X^*) = \frac{\theta_{dx}}{\theta_x}(x(t) - x^*) \equiv -\xi(x(t) - x^*) \tag{85}$$

where we define ξ as the persistence parameter, which will be the same for all variables in the system. equation 84 implies that variables look 'as if' they have an AR(1) structure. If one of the variables is exogenous, the assumption about its dynamic process determines the value of ξ. But because this parameter is identical for all variables, one does not need to calibrate this parameter using an exogenous variable: any variable will do.

As the value of ξ does not depend on the variable X chosen, one can normalize on any convenient variable and then convert to any other. In our case, it is convenient to normalize on λ, which is the only determinant of unemployment in the presence of exogenous separations.

Combining equation 79 and equation 80, equation 78 can be written as:

$$\Omega\theta_x[X(t) - X^*] + \Phi\theta_{dx}[X(t) - X^*] = 0, \tag{86}$$

which simplifies to

$$\Omega\theta_x + \Phi\theta_{dx} = 0 \tag{87}$$

This could be solved using conventional methods for solving linear differential equations with 'jump' and 'state' variables. We now turn to solving this equation for our particular model.

Proof of Proposition 18.1:

We prove Proposition 18.1 in 7 steps, summarised in the following results:

*Result 2:*The sensitivity of the (log) unemployment rate to $\lambda(t)$ is given by:

$$\frac{d\ln u(t)}{d\lambda(t)} = -\frac{1}{\lambda^* + s + \xi}. \tag{88}$$

Proof: Unemployment follows a differential equation $du(t)/dt = s\,[1 - u(t)] - \lambda(t)u(t)$, which can be linearized around the steady-state to yield:

$$\frac{du(t)}{dt} = -(\lambda^* + s)(u(t) - u^*) - u^*(\lambda(t) - \lambda^*). \tag{89}$$

equation 89 can be written as $\theta_u = -u^*/(\lambda^* + s + \xi)$, as unemployment is a backward-looking variable, which implies equation 88. This result is useful because we are ultimately elasticity of wages with respect to unemployment. Using equation 88 we can convert the sensitivity of any model variable to $\lambda(t)$ into an unemployment elasticity.

Result 2: The cyclicality of average wages

As wages are renegotiated at rate ϕ, the law of motion for average wages can be written as:

$$\begin{aligned}\frac{dw}{dt} &= \frac{\lambda(t)u(t)}{1-u(t)}\alpha\,[w_r - w] + \phi\,[w_r - w] \\ &= \left[\alpha\frac{\lambda(t)u(t)}{1-u(t)} + \phi\right][w_r - w]\end{aligned} \tag{90}$$

The first term in equation 90 reflects wage changes from the inflow of new jobs (equal to the separation rate s in steady-state) multiplied by the share of negotiations in new jobs (α), and the second term reflects wage changes from renegotiations in existing jobs. Equation 90 implies that the average wage rises whenever newly-renegotiated wages are higher than average wages.

Linearizing equation 90 around steady-state gives:

$$\frac{dw}{dt} = -(\alpha s + \phi)\left[(w - w^*) - (w_r - w_r^*)\right] + \frac{\alpha u^*(w_r^* - w_a^*)}{1 - u^*}(\lambda(t) - \lambda^*), \tag{91}$$

where the second term is zero because re-negotiated wages are equal to average wages in steady state. As the average wage is a backward-looking variable we the can then derive:

$$\theta_w = \frac{\alpha s + \phi}{\alpha s + \phi + \xi}\theta_{w_r} \tag{92}$$

i.e. average wages are less cyclical than newly-negotiated wages unless there is continual wage re-negotiation ($\phi = \infty$) or unemployment is fully persistent ($\xi = 0$). This gives Proposition 18.1.3. As wages in new jobs are equal to w_r with probability α and w with probability $1 - \alpha$ this also proves Proposition 18.1.4.

<u>Result 3:</u> The cyclicality of reference wages

If both those leaving and those entering unemployment are drawn at random from the respective pools (as we assume), the last observed wage follows the law of motion:

$$\frac{dw_l}{dt} = \frac{s(1 - u(t))}{u(t)}(w - w_l) \tag{93}$$

The reference wage rises (falls) whenever average wages are above (below) previous wages, and it does so at a rate given by the share of the unemployed who have just lost their jobs. In steady-state this rate is equal to the job-finding rate λ.

Linearizing equation 93 around the steady-state gives:

$$\frac{dw_l}{dt} = -\lambda^*\left[(w - w^*) - (w_l - w_l^*)\right], \tag{94}$$

using the fact that reference wages are equal to average wages in steady state. As the reference wage is a backward-looking variable we the can then derive:

$$\theta_{w_l} = \frac{\lambda^*}{\lambda^* + \xi}\theta_{w_a} \tag{95}$$

i.e. reference wages are less cyclical than average wages unless unemployment is fully persistent ($\xi = 0$). Combining equation 95 and equation 92 gives Proposition 18.1.5.

<u>Result 4:</u> The cyclicality of newly-negotiated wages

The linearized version of equation 39 can be written as:

$$\begin{aligned}\theta_{w_r} &= \theta_\rho + \tilde{\beta}(r+\phi+s)\theta_\mu - \tilde{\beta}(1-\alpha)(\theta_w - \theta_{w_r}) \\ &= \theta_\rho + \frac{w^* - \rho^*}{\mu^*}\theta_\mu - \tilde{\beta}(1-\alpha)(\theta_w - \theta_{w_r}),\end{aligned} \tag{96}$$

which allows for cyclicality in hiring costs. Using equation 92 we can eliminate θ_w from equation 96 and re-arrange to obtain:

$$\left(1 - \frac{\tilde{\beta}(1-\alpha)\xi}{\alpha s + \phi + \xi}\right)\theta_{w_r} = \theta_\rho + \frac{w^* - \rho^*}{\mu^*}\theta_\mu \tag{97}$$

which can be expressed in elasticity form:

$$\left(1 - \frac{\tilde{\beta}(1-\alpha)\xi}{\alpha s + \phi + \xi}\right)\frac{\partial \ln w_r(t)}{\partial \ln u(t)} = \frac{\rho^*}{w^*}\frac{\partial \ln \rho(t)}{\partial \ln u(t)} + \left(1 - \frac{\rho^*}{w^*}\right)\frac{\partial \ln \mu(t)}{\partial \ln u(t)} \tag{98}$$

<u>Result 5:</u> The cyclicality of optimal reservation wages
The linearized version of equation 40 can be written as:

$$\begin{aligned}E_t\frac{d\rho^o(t)}{dt} &= (r+\lambda^*+s)(\rho^o(t)-\rho^*) - (\alpha\lambda^* - \phi)(w_r(t) - w_r^*) \\ &\quad - (1-\alpha)\lambda^*(w(t)-w^*) + (\rho^* - w^*)(\lambda(t) - \lambda^*).\end{aligned} \tag{99}$$

As the reservation wage is forward-looking, from equation 99 we can derive:

$$(r+\lambda^*+s+\xi)\theta_{\rho^o} = (\alpha\lambda^* - \phi)\theta_{w_r} + (1-\alpha)\lambda^*\theta_w + (w^* - \rho^*)\theta_\lambda \tag{100}$$

Finally, using equation 92 to eliminate θ_w from equation 100 yields:

$$(r+\lambda^*+s+\xi)\theta_{p^o} = \left[\lambda^* - \phi - \frac{(1-\alpha)\lambda^*\xi}{\alpha s + \phi + \xi}\right]\theta_{w_r} + (w^* - \rho^*)\theta_\lambda \tag{101}$$

<u>Result 6:</u> The cyclicality of optimal reservation wages
From equation 77 we have that:

$$\theta_\rho = \alpha_\rho\theta_{\rho^o} + (1-\alpha_\rho)\alpha_l\theta_{wl}, \tag{102}$$

Which, using equation 95 and equation 92 can be written as:

$$\theta_\rho = \alpha_\rho \theta_{\rho^o} + \frac{(1-\alpha_\rho)\alpha_l(\alpha s + \phi)\lambda}{(\alpha s + \phi + \xi)(\lambda + \xi)} \theta_{w_r}, \tag{103}$$

Substituting equation 101 into equation 103 and converting to an elasticity with respect to unemployment proves Proposition 18.1.2.

Result 7: The cyclicality of newly-negotiated wages.
Solving equation 97, equation 101 and equation 103 leads to:

$$\begin{aligned} &\theta_{w_r} \\ &\quad = \chi(w^* - \rho^*)(\alpha s + \phi + \xi)\left[(r + \lambda + s + \xi)\frac{\theta_\mu}{\mu^*} + \alpha_\rho\right] \end{aligned} \tag{104}$$

with

$$\begin{aligned} 1/\chi =& (\alpha s + \phi + \xi)\big(r + \lambda(1-\alpha_\rho) + \phi\alpha_\rho + s + \xi\big) \\ & - (1-\alpha)\xi\left[\tilde{\beta}(r + \lambda + s + \xi) - \alpha_\rho\lambda\right] - (1-\alpha_\rho)\alpha_l(r + \lambda + s + \xi)\frac{\lambda}{\lambda + \xi} \end{aligned} \tag{105}$$

Using equation 49, this can be expressed in elasticity form:

$$\begin{aligned} &\frac{\partial \ln w_r(t)}{\partial \ln u(t)} \\ &\quad = \chi(1-\eta^*)\frac{r+\phi+s}{r+\lambda+s}(\alpha s + \phi + \xi)\left[(\lambda + \alpha s + \xi)\alpha_\rho - (r + \lambda + s + \xi)\frac{\partial \ln \mu(t)}{\partial \ln u(t)}\right] \end{aligned} \tag{106}$$

Setting $\theta_\mu = 0$ yields equation 52.

Proof of Proposition 18.3:

If $\alpha_\rho = 1$, then as $\phi \to \infty$, equation 53 shows that $\Gamma_r \to 1$. equation 55 shows that we also have:

$$\frac{\Gamma_\rho}{r + \phi + s} \to \frac{1}{r + \lambda + s + \xi} \tag{107}$$

Substituting into equation 59 and setting the hiring cost cyclicality to zero we have that:

$$\frac{\partial \ln w(t)}{\partial \ln u(t)} \rightarrow -(1-\eta^*)\frac{\lambda^* + s + \xi}{r + \lambda^* + s} = -(1-\eta^*)\frac{\xi u^* + s}{ru^* + s} \tag{108}$$

which gives equation 60.

Proof of Proposition 18.4:

If $\alpha_\rho = 1$, and $\alpha = 1$, equation 53 shows that $\Gamma_r \rightarrow 1$. equation 55 shows that we also have:

$$\Gamma_\rho = \frac{\lambda - \phi}{r + \lambda + s + \xi} \tag{109}$$

Substituting into equation 59 and setting the hiring cost cyclicality to zero we have that:

$$\frac{\partial \ln w(t)}{\partial \ln u(t)} = -(1-\eta^*)\frac{\lambda^* + s + \xi}{r + \lambda^* + s}\frac{s + \phi}{s + \phi + \xi}\frac{r + s + \phi}{r + s + \phi + \xi} \tag{110}$$

which gives equation 61.

Proof of Proposition 18.5:

As $\xi \rightarrow 0$ we have that $\Gamma_r \rightarrow 1$ and:

$$\Gamma_\rho \rightarrow \frac{\alpha_\rho(\lambda - \phi)}{r + \lambda + s} + (1 - \alpha_\rho)\alpha_l \tag{111}$$

Substituting into equation 59, setting the hiring cost cyclicality to zero, and re-arranging leads to equation 62

The relationship between our approach and the existing literature

Our approach is built on the derivation of a wage curve, i.e. a relationship between wages and unemployment, which is not affected by shocks to labour demand (e.g. productivity shocks). Labour demand shocks are associated with movements along the wage curve, but they do not alter its position. The shocks are modelled as innovations to the arrival rate of job offers (or, equivalently, to the unemployment rate, which is only affected by the arrival rate in our model). We believe this approach has the advantage of being agnostic about the source and nature of labour demand shocks, and it does not require to measure productivity shocks, but it's important to relate our methodology and results to those of the existing literature in the three following areas:

1. The relationship between our model and one in which shocks are modelled as innovations to productivity, which is assumed to follow an exogenous stochastic process, and the arrival rate of job offers is endogenous;
2. The implications of our model about the response of unemployment to productivity shocks;
3. The use of wage and reservation wage elasticities with respect to current unemployment (our outcome of interest) as opposed to their relative standard deviations, which are the focus of much of the related literature.

A Model with Productivity Shocks

This section shows how our model can be thought of as equivalent to one in which the primitive shock is a productivity shock, assumed to follow an exogenous stochastic process.

Based on the approach for proving Proposition 18.1 in Appendix 22.C, we define θ_p as the relationship between the deviation of productivity $p(t)$ from steady state, and the deviation of $\lambda(t)$ from steady state. As explained in Appendix 22.C, this relationship does not make a statement about the respective exogeneity/endogeneity of $p(t)$ and $\lambda(t)$, as it can be inverted to express the deviation of $\lambda(t)$ from steady state as a function of the deviation of $p(t)$ from steady state. If one is interested in the sensitivity of any outcome of interest, $x(t)$, to productivity rather than unemployment, one would simply divide θ_x by θ_p. Finally, as explained in Appendix 22.C, the parameter ξ can also be treated as the persistence in the productivity shock, i.e we assume

$$E_t\left(\frac{dp(t)}{dt}\right) = -\xi(p(t) - p^*). \tag{112}$$

Define $\tilde{J} = J(w_r(t); w_l(t), t)$ so that 68 can be written as:

$$\tilde{J}(t) = \mu(t) - \frac{(1-\alpha)(w(t) - w_r(t))}{r + \phi + s} \tag{113}$$

Taking averages of equation 34 and using the fact that $V(t) = 0$ at all times gives:

$$\begin{aligned}(r+s)\tilde{J}(t) &= p(t) - w_r(t) + E_t \frac{\partial J(w_r(t); w_l(t), t)}{\partial t} \\ &= p(t) - w_{ra}(t) + E_t\left[\frac{d\tilde{J}(t)}{dt} - \frac{\partial J}{\partial w}\frac{dw_{r(t)}dt}{-}\frac{\partial J}{\partial w_l}\frac{dw_l(t)}{dt}\right]\end{aligned} \tag{114}$$

Using equation 69 and equation 70, equation 114 can be written as:

$$(r+s)\tilde{J}(t) = p(t) - w_r(t) + E_t \\ \left[\frac{d\tilde{J}(t)}{dt} - \frac{1}{r+\phi+s}\frac{dw_r(t)}{dt} + \frac{\phi}{(r+s)(r+\phi+s)}\frac{dw_l(t)}{dt}\right] \tag{115}$$

Using the techniques in Appendix 22.C we can linearize equation 113 and equation 115 as:

$$\theta_{\tilde{J}} = \theta_\mu - \frac{(1-\alpha)(\theta_w - \theta_{w_l})}{r+\phi+s} \tag{116}$$

and

$$(r = s+\xi)\theta_{\tilde{J}} = \theta_p - \theta_{w_r} - \frac{\theta_{w_r}}{r+\phi+s} + \frac{\phi\theta_{w_l}}{(r+s)(r+\phi+s)} \tag{117}$$

respectively. Equation 116 can be used to substitute for $\theta_{\tilde{J}}$ in equation 117, giving an expression for the relationship between θ_p and wage dynamics (as captured by θ_{w_r}, θ_w, θ_{w_l} and θ_{w_p}). This could be then used to express θ_{w_p} as a function of the exogenous parameters of the model (which determine θ_{w_r}, θ_w, θ_{w_l} and θ_{w_p} as shown in Appendix 22.C). If one wants to relate some variable x to productivity one can then derive the implied relationship by dividing θ_x by θ_p.

The above argument is based on the assumption that θ_μ is exogenous. We extend it below to the case of endogenous θ_μ , noting that the dynamics of the mark-up $\mu(t) \equiv c(t)/q(t) + C(t)$ is potentially affected by the impact of productivity on running and lump-sum hiring costs, $c(t)$ and $C(t)$ respectively, and by the time it takes to fill a vacancy, $1/q(t)$. As we are simply interested in illustrating the case of endogenous θ_μ here, we assume constant $c(t)$ and $C(t)$ but allow $\mu(t)$ to vary with $q(t)$, namely:

$$\theta_\mu = -\theta_q \frac{c}{q^2} \tag{118}$$

θ_q is in turn determined by the matching function, which we assume to have constant returns in unemployment and vacancies and delivers the following relationship between λ and q:

$$\lambda = qv(q) \tag{119}$$

where $v(q)$ is the relationship between the unemployment-to-vacancy ratio and the vacancy-filling rate q. Linearizing equation 119 yields an expression for θ_q.

In summary, we have shown that one can build a model in which the response of all variables of interest to productivity can be expressed as a function of the exogenous model parameters. One of those parameters ξ, which we have modelled as the persistence in $\lambda(t)$ is endogenous if the exogenous source of shocks is represented by innovations to productivity. But, as explained in the first part of Appendix 22.C, in the model the value of ξ would be identical if it denoted instead the persistence in productivity. Empirically, the estimate of ξ may differ according to the variable used to calibrate it but, as discussed in the main text, this makes little or no quantitative difference whether one calibrates ξ to the arrival rate of job offers or productivity.

The economic intuition for why modelling shocks to productivity is equivalent to directly modelling shocks to the job finding rate is that, in search models, productivity determines the vacancy filling rate, which then determines the job-finding rate. Thus the dynamics in productivity must be mirrored in the dynamics in the job-finding rate.

The Impact of Changes in Productivity on Unemployment

For simplicity we limit our discussion to the steady-state, in which the wage curve is given by equation 46. This is a special case of the model above. In steady state, the labour-demand (or vacancy-creation) curve is given by:

$$(p - w) = (r + s)\mu(u), \tag{120}$$

which allows hiring costs to depend on unemployment. Differentiating equation 120 with respect to productivity, and taking into account the dependence of wages on unemployment through the wage curve gives:

$$1 - \frac{w}{p}\frac{\partial \ln w}{\partial \ln u}\frac{\partial \ln u}{\partial \ln p} = \frac{(r+s)\mu}{p}\frac{\partial \ln \mu}{\partial \ln p} = \left(1 - \frac{w}{p}\right)\frac{\partial \ln \mu}{\partial \ln u}\frac{\partial \ln u}{\partial \ln p} \tag{121}$$

Re-arranging equation 121 leads to:

$$\frac{\partial \ln u}{\partial \ln p} = \frac{1}{\frac{w}{p}\frac{\partial \ln w}{\partial \ln u} + \left(1 - \frac{w}{p}\right)\frac{\partial \ln \mu}{\partial \ln u}} \tag{122}$$

If hiring costs are acyclical, the elasticity of unemployment with respect to productivity is simply the inverse of the elasticity of wages with respect to unemployment, rescaled by the ratio of wages to productivity, which is related to the size of hiring costs relative to productivity. If hiring costs are cyclical, unemployment becomes less sensitive to productivity, and wages become more strongly procyclical.

Note that, given the assumption of constant returns to scale in this model, TFP shocks are identical to average labour productivity so that the assumption of exogenous TFP shocks also implies that the measured average product of labour is exogenous. But, as pointed out by Rogerson and Shimer (2011), a different assumption about the production function might lead to different conclusions. For example, a Cobb-Douglas production function with decreasing returns to labour would always deliver proportionality between average labour productivity and the wage, though causation may run from the latter to the former. In this case the correlation between wages and productivity is uninformative about anything other than the production function. Our approach allows us to be more agnostic about the form of the production function although one should recognize that different production functions might lead to different wage curves (see also Elsby and Michaels (2013), for the derivation of a very similar wage curve).

Elasticity versus Relative Standard Deviations

Our main parameter of interest is the elasticity of log wages with respect to log unemployment as estimated from wage and reservation wage curves. Most of the existing literature uses relative standard deviations as the parameter of interest. Suppose the relationship between log wages and log unemployment is written as:

$$\ln w = \beta_0 + \beta \ln u + \varepsilon$$

where for simplicity we assume there are no other regressors and β is a regression coefficient. If the correlation coefficient between $\ln w$ and $\ln u$ is r, the following relationships hold:

$$\beta = \frac{cov(\ln w, \ln u)}{var(\ln u)} = r\frac{stdev(\ln w)}{stdev(\ln u)}$$

Thus the regression coefficient (our elasticity) and the ratio of standard deviations are identical whenever $r = -1$ or, equivalently, the R^2 from the regression equals 1. Many models in the literature implicitly assume this because they are one-factor models in which TFP alone causes variation in unemployment and wages. If variation in unemployment is driven by TFP shocks, one could also convert this into the relative standard deviation of wages and productivity. If $r \neq 1$, the elasticity and relative standard deviations differ, and the elasticity is preferable for our purposes as we are interested in the variation in wages driven by unemployment, not their total variation. Mortensen and Nagypál (2007) provide a very clear discussion of these issues.

The Impact of Changes in Unemployment Benefits on Unemployment

The parameter configuration of Hagedorn and Manovskii (2008) has been criticized by Costain and Reiter (2008) for making unemployment excessively sensitive to changes in unemployment benefits. Here we relate this claim to the prediction of our model with reference-dependence.

In the steady-state, changes in the value of nonemployment z shift wages one-for-one (see equation 46), so that the elasticity of wages with respect to z is given by:

$$\frac{\partial \ln w}{\partial \ln z} = \eta \tag{123}$$

If hiring costs are acyclical, the elasticity of unemployment to z is given by:

$$\frac{\partial \ln u}{\partial \ln z} = \frac{\partial \ln w}{\partial \ln z} / \frac{\partial \ln w}{\partial \ln u} = \eta / \frac{\partial \ln w}{\partial \ln u} \tag{124}$$

However, Costain and Reiter (2008) assume that unemployment benefits (b) are the sole determinant of z. This may not always be the case, especially in models with reference-dependence. If one allows a different link between benefits and z we have that:

$$\frac{\partial \ln u}{\partial \ln b} = \frac{\partial \ln u}{\partial \ln z}\frac{\partial \ln z}{\partial \ln b} = \eta \frac{\partial \ln z}{\partial \ln b} / \frac{\partial \ln w}{\partial \ln u} \tag{125}$$

One can then explain a small impact of benefits on aggregate unemployment via a small impact of benefits on z and, hence, reservation wages. This is consistent with empirical evidence on the determinants of reservation wages, reported in columns 2 and 4 of Table 33: our specifications do control for benefits, and we find benefits to have only a small effect on reservation wages.

An alternative Model for Updating the Previous Wage

In our baseline case workers are assumed to believe they will keep their "current" previous wage if the current job ends, even though the current job generally pays a wage different from the current previous wage. This section considers an alternative model in which workers internalize that the current wage will become their previous wage in the event of job loss. This means that value of being employed at a wage at time when the previous wage was is denoted by $W(w; w_l, t)$ needs to be modified from equation 36 to:

$$\begin{aligned} rW(w; w_l, t) &= w + \phi[W(w_r(w_l, t); w_l, t) - W(w; w_l, t)] \\ &- s[W(w; w_l, t) - W(\rho(w, t); w, t)] + E_t \frac{\partial W(t; w)}{\partial t}, \end{aligned} \tag{126}$$

This change implies that equation 69 needs to be modified to:

$$(r + \phi + s)\frac{\partial W}{\partial w} = 1 + s\frac{\partial W}{\partial w}\frac{\partial \rho}{\partial w_l} + s\frac{\partial W}{\partial w_l} \tag{127}$$

Differentiating equation 126 gives:

$$(r + \phi + s)\frac{\partial W}{\partial w_l} = \phi\frac{\partial W}{\partial w}\frac{\partial w_r}{\partial w_l} = \phi\frac{\partial W}{\partial w}\frac{\partial w_r}{\partial \rho}\frac{\partial \rho}{\partial w_l} \tag{128}$$

Combining equation 128 and equation 127 and using $\partial w_r/\partial \rho = 1$ and $\partial \rho/\partial w_l = (1 - \alpha_r)\alpha_l$, gives:

$$(r + \phi + s)\frac{\partial W}{\partial w} = 1 + s\frac{\partial W}{\partial w}\frac{\partial \rho}{\partial w_l} + \frac{s\phi}{(r + \phi + s)}\frac{\partial W}{\partial w}\frac{\partial \rho}{\partial w_l} \tag{129}$$

which can be re-arranged to yield:

$$\frac{\partial W}{\partial w} = \frac{1}{r + \phi + s\left[1 - \frac{\partial \rho}{\partial w_l}\right] - \frac{s\phi}{(r + \phi + s)}\frac{\partial \rho}{\partial w_l}} \tag{130}$$

This means that when it comes to wage determination $\partial W/\partial w = -\partial J/\partial w$ no longer holds. Instead, the wage determination equation 73 can be written in the same form, but with the bargaining power parameter modified to:

$$\tilde{\beta}\frac{-\frac{\partial J}{\partial w}}{\frac{\partial W}{\partial w}} = \tilde{\beta}\frac{r + \phi + s\left(1 - \frac{\partial \rho}{\partial w_l}\right) - \frac{s\phi}{(r + \phi + s)}\frac{\partial \rho}{\partial w_l}}{r + \phi + s} \tag{131}$$

All other results follow equivalently because when we eliminate $\tilde{\beta}$ to convert to the replacement ratio we get the same expression as in the main text.

22.D Additional Tables and Figures

Table 32: Descriptive statistics

	United Kingdom				West Germany			
	Wage sample		Reservation wage sample		Wage sample		Reservation wage sample	
Variables:	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.
Reservation wage			5.226	6.206			1180.366	703.219
Wage	9.866	6.203			2387.666	1898.023		
Female	0.526	0.500	0.546	0.498	0.430	0.495	0.616	0.486
Age	38.106	11.691	34.666	14.024	39.039	11.644	33.289	11.316
Higher education	0.117	0.321	0.247	0.431	0.254	0.435	0.143	0.350
Upper secondary education	0.269	0.443	0.353	0.478	0.528	0.499	0.549	0.498
Lower secondary education	0.405	0.491	0.314	0.464	0.178	0.382	0.211	0.408
No qualifications	0.209	0.407	0.085	0.280	0.040	0.040	0.097	0.086
Married	0.717	0.451	0.514	0.500	0.657	0.475	0.559	0.497
No. Kids	0.686	0.965	0.917	1.168	0.730	0.990	1.027	1.120
Duration in current status (years)	4.880	5.969	4.387	5.748	10.464	9.653	2.962	3.902
Benefits			276.414	318.201			255.835	448.710
Looking for full-time work							0.482	0.500
Looking for part-time work							0.382	0.486
Looking for either							0.109	0.312
Unsure about working hours							0.027	0.161
Social insurance contributions (months)							5.242	6.878
Months to benefit expiry							1.109	3.679
Entitled to unemployment benefits							0.196	0.397
Hours worked					38.495	12.680		
Number of observations	96,270		14,874		166,614		11,221	

Notes: Samples include employees aged 16-65 with non-missing wage information (wage sample), and unemployed jobseekers aged 18-65 with non-missing reservation wage information (reservation wage sample). Source: BHPS 1991-2009 and SOEP 1984-2010.

Table 33: Detailed results on wage and reservation wage equations for the UK and West Germany

	United Kingdom		West Germany	
	1	2	3	4
Dependent variable:	Log wage	Log res wage	Log wage	Log res wage
Log aggregate unemployment rate	-0.165***	-0.175***	0.002	0.001
	(0.044)	(0.058)	(0.025)	(0.065)
Female	-0.263***	-0.102***	-0.265***	-0.188***
	(0.009)	(0.011)	(0.015)	(0.018)
Age	0.073***	0.033***	0.082***	0.018***
	(0.002)	(0.002)	(0.002)	(0.003)
Age^2 (/100)	-0.084***	-0.034***	-0.009***	-0.003***
	(0.002)	(0.002)	(0.000)	(0.000)
Lower secondary qualification	0.193***	0.068***	0.023**	-0.016
	(0.008)	(0.009)	(0.011)	(0.024)
Upper secondary qualification	0.361***	0.157***	0.230***	0.093***
	(0.007)	(0.011)	(0.015)	(0.023)
Higher education	0.710***	0.352***	0.562***	0.276***
	(0.004)	(0.013)	(0.019)	(0.029)
Married	0.092***	0.042***	0.032***	-0.038***
	(0.006)	(0.006)	(0.003)	(0.010)
No. kids in household	-0.019***	0.018***	-0.020***	-0.006
	(0.003)	(0.004)	(0.004)	(0.005)
Duration in current status (years)	0.018***	-0.002	0.037***	0.013**
	(0.001)	(0.002)	(0.002)	(0.005)
Duration in current $status^2$ (years/10)	-0.010***	-0.001	-0.012***	-0.014**
	(0.001)	(0.002)	(0.001)	(0.006)
Duration in current $status^3$ (years/100)	0.017***	0.003	0.002***	0.003**
	(0.002)	(0.003)	(0.000)	(0.002)
Log(Unemp benefits + 1)		0.004**		0.004
		(0.001)		(0.003)
Receives housing benefits		0.017**		-0.075***
		(0.008)		(0.026)
Social insurance contributions (years)				0.005***
				(0.001)
Looking for full-time work				0.151***
				(0.036)
Looking for part-time work				-0.507***
				(0.033)
Looking for any hours				-0.051*
				(0.031)
Log hours worked			0.912***	
			(0.042)	
Year	-0.009	0.004	0.022***	0.027***
	(0.007)	(0.007)	(0.002)	(0.008)
$(Year-1990)^2$	0.001***	0.001**	-0.696***	-1.003***
	(0.000)	(0.000)	(0.078)	(0.253)
Observations	96,270	14,847	166,614	11,221
R-squared	0.397	0.249	0.605	0.359

Notes: See notes to Table 32 for sample used. The wage measure is hourly for the UK and monthly for West Germany. All regressions include region dummies. Standard errors are clustered at the year level. Source: BHPS 1991-2009 and SOEP 1984-2010.

Table 34: Estimates of a Wage Equation for the UK, 1991-2009. Further estimates with regional controls.

	Dependent variable: Log monthly wage									
	1	2	3	4	5	6	7	8	9	10
Log hourly wage, lagged										0.073
										(0.052)
Log regional unemployment rate	0.010	-0.009	-0.036*	-0.053***	-0.044**	-0.042***	-0.039**	-0.010		-0.058***
	(0.010)	(0.022)	(0.019)	(0.017)	(0.017)	(0.010)	(0.016)	(0.022)		(0.022)
Log regional unemployment rate * new job					-0.032***	-0.011**				
					(0.005)	(0.006)				
Log regional unemployment rate, at start of job							-0.084***			
							(0.015)			
Log regional unemployment rate, lagged								-0.060***	-0.065***	
								(0.017)	(0.013)	
Trend	no	linear	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic
Year dummies	✓									
Individual fixed effects				✓	✓		✓	✓	✓	✓
Job fixed effects						✓				
Observations	96,269	96,269	96,269	92,380	92,380	77,854	91,712	92,380	92,380	53,054
R-squared	0.399	0.397	0.397	0.809	0.810	0.889	0.810	0.810	0.810	

Notes: See notes to Table 32 for sample. The dependent variable is the log gross hourly wage, deflated by the aggregate consumer price index. Estimation method: OLS in columns 1-9; Arellano Bond (1991) estimator for dynamic panel data models in column 10. All regressions include a gender dummy, age and its square, three education dummies, a cubic trend in job tenure, a dummy for married, the number of children in the household and eleven region dummies. Standard errors are clustered at the region*year level in column 1; at the year level in columns 2 and 3; and using 2-way cluster-robust variance (Cameron and Miller (2015)) in columns 4-10. Source: BHPS.

Table 35: Estimates of a Wage Equation for West Germany, 1984-2010. Further estimates with regional controls

	Dependent variable: Log monthly wage									
	1	2	3	4	5	6	7	8	9	10
Log monthly wage, lagged										0.390***
										(0.027)
Log regional unemployment rate	-0.033**	0.015	0.006	-0.008	-0.003	0.001	-0.009	0.049***		-0.004
	(0.016)	(0.026)	(0.023)	(0.015)	(0.014)	(0.011)	(0.015)	(0.016)		(0.016)
Log regional unemployment rate * new job					-0.039***	-0.011				
					(0.013)	(0.011)				
Log regional unemployment rate, at start of job							-0.013			
							(0.009)			
Log regional unemployment rate, lagged								-0.079***	-0.044***	
								(0.013)	(0.014)	
Trend	no	linear	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic	quadratic
Year dummies	✓									
Individual fixed effects				✓	✓		✓	✓	✓	✓
Job fixed effects						✓				
Observations	166,614	166,614	166,614	161,075	160,865	149,617	157,241	161,075	161,075	101,526
R-squared	0.652	0.649	0.651	0.414	0.415	0.199	0.422	0.415	0.415	

Notes: See notes to Table 32 for sample. The dependent variable is the log gross monthly wage, deflated by the aggregate consumer price index. Estimation method: OLS in columns 1-9; Arellano Bond (1991) estimator for dynamic panel data models in column 10. All regressions include a gender dummy, age and its square, three education dummies, a cubic trend in job tenure, a dummy for married, the number of children in the household and eleven region dummies. Standard errors are clustered at the region*year level in column 1; at the year level in columns 2 and 3; and using 2-way cluster-robust variance (Cameron and Miller (2015)) in columns 4-10. Source: SOEP.

Table 36: Estimates of a Reservation Wage Equation for the UK, 1991-2009. Further estimates with regional controls.

	Dependent variable: Log hourly reservation wage					
	1	2	3	4	5	6
Log regional unemployment rate	0.007	-0.047	-0.054*	-0.034	0.048	
	(0.025)	(0.031)	(0.028)	(0.030)	(0.037)	
Log regional unemployment rate, lagged					-0.106***	-0.078***
					(0.030)	(0.024)
Trend	no	linear	quadratic	quadratic	quadratic	quadratic
Year dummies	✓					
Individual fixed effects				✓	✓	✓
Observations	14,873	14,873	14,873	10,774	10,774	10,774
R-squared	0.252	0.247	0.247	0.613	0.614	0.614

Notes: See notes to Table 32 for sample. The dependent variable is the log net hourly reservation wage, deflated by the aggregate consumer price index. Estimation method: OLS. All regressions also include a gender dummy, age and its square, three education dummies, a cubic trend in unemployment duration, a dummy for married, the number of children in the household, the log of unemployment benefits, a dummy for receipt of housing benefits, and eleven region dummies. Standard errors are clustered at the year*region level in column 1, at the year level in columns 2 and 3, and using 2-way cluster-robust variance (Cameron and Miller (2015)) in columns 4-6. Source: BHPS.

Table 37: Estimates of a Reservation Wage Equation for West Germany. Further estimates with regional controls

	Dependent variable: Log monthly reservation wage					
	1	2	3	4	5	6
Log regional unemployment rate	-0.079*	0.028	0.018	0.034	0.116***	
	(0.043)	(0.044)	(0.028)	(0.023)	(0.029)	
Log regional unemployment rate, lagged					-0.113***	-0.031
					(0.032)	(0.023)
Trend	no	linear	quadratic	quadratic	quadratic	quadratic
Year dummies	✓					
Individual fixed effects				✓	✓	✓
Observations	11,221	11,221	11,221	7,911	7,911	7,911
R-squared	0.421	0.413	0.418	0.124	0.125	0.123

Notes: See notes to Table 32 for sample. The dependent variable is the log net monthly reservation wage, deflated by the aggregate consumer price index. Estimation method: IV. All regressions also include a gender dummy, age and its square, three education dummies, a cubic trend in unemployment duration, a dummy for married, the number of children in the household, the log of unemployment benefits, a dummy for receipt of housing benefits, controls for whether an individual looks for full-time, part-time or any job (the omitted category being "unsure about preferences"), months of social insurance contributions and eleven region dummies. Unemployment benefits are instrumented, see notes to Table 28. Standard errors are clustered at the year) level in columns 1-3, and using 2-way cluster-robust variance (Cameron and Miller (2015)) in columns 4-6. Source: SOEP.

22.E Alternative models for the reservation wage

The reservation wage with on-the-job search

Our baseline model assumes that only the unemployed search for jobs, while a fraction close to half of new jobs are taken by workers currently employed (Manning, 2003). This subsection considers how the reservation wage is altered when both the unemployed and the employed search for jobs. The analysis is conditional on expected wages, without need to specify the process for wage determination.

For simplicity, we assume that the economy is in steady-state, so wages and job offer arrival rates for employed and unemployed jobseekers are constant, and they will be denoted by λ^e and λ^u, respectively. The corresponding value functions are given by:

$$rW(w) = w - s[W(w) - U] + \lambda^e \int_w [W(x) - W(w)]dF(x) \tag{132}$$

and

$$rU = z + \lambda^u \int_\rho [W(x) - U]dF(w), \tag{133}$$

respectively. The reservation wage satisfies $W(\rho) = U$, and can be expressed as:

$$\begin{aligned}\rho &= z + (\lambda^u - \lambda^e) \int_\rho [W(w) - U]dF(w) \\ &= z + (\lambda^u - \lambda^e) \int_\rho \frac{1 - F(w)}{r + s + \lambda^e[1 - F(w)]}dw,\end{aligned} \tag{134}$$

where the second equality follows from integration by parts, given $W'(w) = \{r + s + \lambda^e[1 - F(w)]\}^{-1}$. The possibility of search on-the-job implies that the distribution of wages across workers, $G(w)$ differs from the distribution of wage offers $F(w)$ and it can be shown (see Burdett and Mortensen (1998)) that the two are related by:

$$1 - G(w) = (s + \lambda^e)\frac{1 - F(w)}{s + \lambda^e[1 - F(w)]}, \tag{135}$$

having imposed $F(\rho) = 0$ as no firm chooses to offer a wage below the reservation wage in equilibrium. Using equation 100 and the approximation $r \approx 0$, equation 134 can be written as:

$$\rho \approx z + \frac{\lambda^u - \lambda^e}{s + \lambda^e} \int_\rho [1 - G(w)]dw = z + \frac{\lambda^u - \lambda^e}{s + \lambda^e}(w_a - \rho) \tag{136}$$

Re-arranging gives:

$$\rho \approx \frac{(s+\lambda^e)z + (\lambda^u - \lambda^e)\bar{w}}{s+\lambda^u} \tag{137}$$

Unemployment is given by $u = s/(s+\lambda^u)$, and substituting this in equation 137 gives

$$\rho \approx z + (1-u)\left(1 - \frac{\lambda^e}{\lambda^u}\right)(w_a - z), \tag{138}$$

where w_a denotes the average wage.

According to equation 138, reservation wages are acyclical whenever the job arrival rates for employed and unemployed workers are equal, $\lambda^e = \lambda^u$, as in this case the reservation wage equals the flow of unemployment income, $\rho = z$ (Burdett and Mortensen (1998)). Intuitively, taking or leaving a job offer has no consequences for future job opportunities when arrival rates are independent of one's employment status, and the optimal search strategy consists in accepting the first offer that provides a higher flow utility than that enjoyed while unemployed. If z is not cyclical, neither is the reservation wage.

While this seems an attractive path to reduce the cyclicality of reservation wages, it has the less desirable consequence that the reservation wage is independent of factors that influence the distribution of wages. This prediction is strongly rejected by the data, as high-wage workers tend to have relatively higher reservation wages. Detailed results reported in Table 33 show that gender, age and education affect wages and reservation wages in the same direction, thus the reservation wage is positively related to the wage that workers expect to earn. Taken to equation 138, this result implies that off-the-job search is more effective than on-the-job search, a conclusion that is also in line with structural estimates of labour market transition rates.

In general, using equation 138, the reservation wage embodies the cyclicality in wages, plus a further cyclical component represented by $(1-u)(1-\lambda^e/\lambda^u)$. The term $1-u$ is clearly pro-cyclical. To determine the cyclicality of cyclicality of the term in λ^e/λ^u, we show that this ratio is positively related to the fraction of new jobs filled by previously employed workers, which can be directly measured on data on labour market transitions. The two measures are related as the more effective on-the-job search, the higher the fraction of jobs that are filled by someone already employed.

To see this, denote by f the position of a firm in the wage offer distribution. The fraction of workers employed in firms at or below position f satisfies:

$$[s + \lambda^e(1-f)]G(f)(1-u) = \lambda^u u f, \tag{139}$$

which simply equates flows into and out of firms paying f or below. Re-arranging and using $u = s/(s+\lambda^u)$ gives:

$$G(f) = \frac{sf}{s+\lambda^e(1-f)} \tag{140}$$

Total recruits to a firm at position f, $R(f)$, are given by:

$$R(f) = \lambda^u u + \lambda^e(1-u)G(f) = \frac{s\lambda^u}{s+\lambda^u}\frac{s+\lambda^e}{s+\lambda^e(1-f)} \tag{141}$$

and total recruits in the economy are given by:

$$R = \int_0^1 R(f)df = \frac{s\lambda^u}{s+\lambda^u}\frac{s+\lambda^e}{s+\lambda^e(1-f)} = \frac{s\lambda^u}{s+\lambda^u}\frac{s+\lambda^e}{\lambda^e}\ln\left(\frac{s+\lambda^e}{s}\right) \tag{142}$$

As the total recruits from unemployment are given by λ^u/u, this implies that the fraction of recruits from employment, which we will denote by ζ, is given by:

$$\zeta = 1 - \frac{\lambda^e}{(s+\lambda^e)\ln\left(\frac{s+\lambda^e}{s}\right)} = \frac{\lambda^e/\lambda^u}{\left(\frac{u}{1-u}+\frac{\lambda^e}{\lambda^u}\right)\ln\left(1+\frac{\lambda^e}{\lambda^u}\frac{1-u}{u}\right)} \tag{143}$$

We obtain evidence on ζ from the UK Quarterly LFS, looking at the previous quarter's employment status of newly-hired workers.[120] During 1993-2012, this fraction is on average 60.1%. Regressing ζ on the unemployment rate show that ζ is pro-cyclical, with a slope coefficient on unemployment of approximately 1 (see results reported in Table 38 below). Using equation 143, an average unemployment rate in the UK over 1993-2012 of 6.8% and an average ζ of 60.1% implies $\lambda^e/\lambda^u = 0.612$. Based on equation 138, the result $0 < \lambda^e/\lambda^u < 1$ would make reservation wages less cyclical than in the case with $\lambda^e = 0$, while the cyclicality of λ^e/λ^u would make reservation wages more cyclical.

To resolve this ambiguity, note that equation 143 implies an inverse relationship between ζ and unemployment even if λ^e/λ^u does not vary with the cycle. But the strength of the relationship between ζ and l shown in Table 38 is weaker than we would expect from equation 143 if λ^e/λ^u were acyclical. This implies that, as u rises, so does λ^e/λ^u. The estimates in Table 38 imply $\lambda^e/\lambda^u = 0.726$ for $u = 0.01$ and $\lambda^e/\lambda^u = 0.443$ for $u = 0.04$. According to equation 138, this mechanisms acts to make the reservation wage even more sensitive to the unemployment rate.

[120] We do not adjust this statistic for time aggregation, so it may be possible that a worker in employment this quarter and 3 quarters ago has had an intervening period of non-employment. Given the outflow rates from unemployment in the UK this makes little difference to the computations.

Table 38: The Cyclicality in the Fraction of New Hires from Previously Jobs

	Dependent variable: Fraction of new hires from previous jobs			
Unemployment rate	-1.51**	-1.91**	-1.02	-0.97
	(0.065)	(0.076)	(0.081)	(0.195)
Region effects	No	yes	No	yes
Year effects	No	no	Yes	yes
R-squared	0.57	0.69	0.67	0.75
No. observations	416	416	416	416

Notes: Each observation is a region-year cell, and all regressions are weighted by cell size. Cells based on less than 50 observations are omitted. Sample period: 1993-2012. Source: UK LFS.

The reservation wage with hyperbolic discounting

The models so far considered have assumed that individuals have rational expectations and time-consistent preferences, but a growing body of evidence casts doubt on both these assumptions. In the area of job search, Spinnewijn (2015) argues that the unemployed tend to be overoptimistic about their job prospects, and DellaVigna and Paserman (2005) and Paserman (2008) show that hyperbolic discounting has large effects on search intensity but very small effects on the reservation wage. They do not investigate the implication of hyperbolic discounting for the cyclicality of reservation wages, but the analysis below shows that hyperbolic discounting is not likely to have important consequences for the cyclicality of the reservation wage.

To stay close to our benchmark framework, we use the continuous time version of hyperbolic discounting developed by Haris and Laibson (2013). Consider the arrival rate of a shock – here denoted by δ – which turns one into a person (the future self) who cares less about the future than one's current self. The weight attached to the future self is denoted by Ψ.

The expectation is that the future self is a straightforward exponential discounter. The value function for being employed equation 36 is now modified to:

$$rW(w) = w - s[W(w) - U] + \delta[\Psi\tilde{W}(w) - W(w)], \tag{144}$$

where $\tilde{W}(w)$ is the value of being employed for the future non-hyperbolic self, given by equation 36:

$$r\tilde{W}(w) = w - s[\tilde{W}(w) - \tilde{U}] \tag{145}$$

The value function for the unemployed can similarly be written as:

$$rU = z + \lambda(W - U) + \delta(\Psi\tilde{U} - U), \tag{146}$$

where $\tilde{U}$ is given by equation 37. Thus:

$$\tilde{W} - \tilde{U} = \frac{w - z}{r + s + \lambda} \tag{147}$$

and:

$$r\tilde{U} = z + \frac{\lambda(w - z)}{r + s + \lambda} \tag{148}$$

From equation 144 and equation 146 one can then derive:

$$W - U = \frac{(w - z) + \delta\Psi(\tilde{W} - \tilde{U})}{r + s + \lambda + \delta} \tag{149}$$

Using equation 149, equation 146 and equation 147 one can, after some re-arrangement, derive:

$$rU = \frac{r + \delta\psi}{r + \delta} z + \frac{\lambda(w - z)[r(r + s + \lambda + \delta\Psi) + \delta\Psi(r + s + \lambda + s\delta)]}{(r + s + \lambda)(r + s + \lambda + \delta)(r + \delta)} \tag{150}$$

The reservation wage, ρ, must satisfy $W(\rho) = U$. Using equation 144, this implies:

$$rU = \rho + \delta\left[\Psi\tilde{W}(\rho) - U\right] \tag{151}$$

Using equation 145 we obtain:

$$\tilde{W}(\rho) = \frac{\rho + s\tilde{U}}{r + s} \tag{152}$$

Combining equation 151 and equation 152 leads to the following expression for the reservation wage:

$$\rho = \frac{(r + s)(r + \delta)U - \delta\Psi s\tilde{U}}{r + s + \lambda + \delta\Psi} \tag{153}$$

Substituting this into equation 150 and equation 148 and re-arranging leads to the following expression:

$$\rho = z + \frac{\lambda(w - z)}{r + s + \lambda}\left(\frac{r + s}{r + s + \lambda + \delta} + \frac{\delta\Psi}{r + s + \lambda + \delta\Psi}\right) \leq z + \frac{\lambda(w - z)}{r + s + \lambda} \tag{154}$$

The inequality shows that hyperbolic discounting (δ) lowers the reservation wage, and it reduces the weight on the wage in the determination of the reservation wage. Both results are intuitive as hyperbolic discounting makes an individual more present-oriented. The reduced weight on the wage makes the reservation wage less sensitive to the unemployment rate, but at the same time makes wages and reservation wages less strongly correlated.

In the calibration of Haris and Laibson (2013), $\delta = 2/3$ (at the annual level). In this case the reservation wage is very close to z, clearly making the reservation wage insensitive to unemployment, but at the cost of making it insensitive to the expected wage, while Table 33 shows that wages and reservation wages respond in very similar ways to most covariates considered.

References

Abraham, K. G. and Kearney, M. S. (2018). Explaining the Decline in the U.S. Employment-To-Population Ratio: a review of the evidence. *Nber*. 9

Acemoglu, D. (1998). Why Do New Technologies Complement Skills ? Directed Technical Change and Wage Inequality. *The Quarterly Journal of Economics*, 113(4):1055–1089. 5

Acemoglu, D. and Autor, D. H. (2011). Skills, Tasks and Technologies: Implications for Employment and Earnings. *Handbook of Labor Economics*, 4(PART B):1043–1171. 1, 8.D.1

Acemoglu, D., Laibson, D., and List, J. A. (2014). Equalizing superstars: The internet and the democratization of education. *American Economic Review*, 104(5):523–527. 3

Acemoglu, D. and Restrepo, P. (2017). Robots and Jobs: Evidence from US Labor Markets. *SSRN Electronic Journal*. 1, 9

Acemoglu, D. and Restrepo, P. (2018). The Race between Machine and Man: Implications of Technology for Growth, Factor Shares and Employment. *American Economic Review*, 108. 1

Aguiar, M., Bils, M., Charles, K. K., and Hurst, E. (2017). Leisure luxuries and the labor supply of young men. 9

Akerlof, G. A. and Yellen, J. L. (1990). The Fair Wage-Effort Hypothesis and Unemployment. *Quarterly Journal of Economics*, CV(May). 16

Akerman, A., Gaarder, I., and Mogstad, M. (2013). The Skill Complementarity of Broadband Internet. *Quarterly Journal of Economics*, (7762). 1, 9

Alvaredo, F., Atkinson, A., Piketty, T., and Saez, E. (2018). The World Inequality Database. 1

Ammermüller, A., Lucifora, C., Origo, F., and Zwick, T. (2009). Wage Flexibility in Regional Labour Markets : Evidence from Italy and Germany. *Regional Studies*, 44(4):401–421. 114

Armour, P., Burkhauser, R., and Larrimore, J. (2015). Using the Pareto Distribution to Improve Estimates of Topcoded Earnings. *Economic Inquiry*. 8.D.2

Atkinson, A. B. and Piketty, T. (2010). *Top Incomes - A Global Perspective*. 8.D.2

Atkinson, A. B., Piketty, T., and Saez, E. (2011). Top Incomes in the Long Run of History. *Journal of Economic Literature*, 49(1):3–71. 8.D.2

Autor, D. H. and Dorn, D. (2013). The Growth of Low Skill Service Jobs and the Polarization of the U . S . Labor Market. *American Economic Review*, 103(5):1553–97. 1, 3.3, 8.D.2

Autor, D. H., Dorn, D., and Hanson, G. H. (2015). Untangling Trade and Technology: Evidence from Local Labor Markets. *Economic Journal*, 125(7329):1–24. 9

Autor, D. H., Dorn, D., Katz, L. F., Patterson, C., Reenen, J. V., and Van Reenen, J. (2017). Concentrating on the Fall of the Labor Share. 1

Axtell, R. L. (2001). Zipf distribution of U.S. firm sizes. *Science*, 293(5536):1818–1820. 8.C

Baker, D. (2016). *Rigged: How Globalization and the Rules of the Modern Economy Were Structured to Make the Rich Richer*. 8

Bakija, J., Cole, A., and Heim, B. (2012). Jobs and Income Growth of Top Earners and the Cause of Changing Income Inequality: Evidence from US Tax Return Data. *unpublished manuscript, Williams College*. 1

Baltagi, B. H., Blien, U., and Wolf, K. (2009). New evidence on the dynamic wage curve for Western Germany : 1980 â 2004. *Labour Economics*, 16(1):47–51. 114

Bas, M., Paunov, C., and Rodriguez-Montemayor, E. (2018). Superstars in winner-take-all markets. (July). 3

Bell, B., Nickell, S., and Quintini, G. (2002). Wage equations, wage curves and all that. *Labour Economics*, 9:341–360. 114

Benabou, R. and Tirole, J. (2016). Bonus Culture : Competitive Pay , Screening , and Multitasking. *Journal of Political Economy*, 124(2). 7, 8

Benito, A. (2000). Inter-Industry Wage Differentials in Great Britain. *Oxford Bulletin of Economics and Statistics*, 62. 20

Bertrand, M. and Mullainathan, S. (2001). Are CEOS Rewarded for Luck? The Ones without Principals Are. *Quarterly Journal of Economics*, 116(3):901–932. 8

Blanchard, O. and Katz, L. F. (1992). Regional evolutions. *Brookings Papers on Economic Activity*, pages 1–75. 19.1

Blanchard, O. and Katz, L. F. (1999). Wage dynamics: Reconciling theory and evidence. *American Economic Review*, 89(2):69–74. 16

Blanchet, T. (2016). Wealth inequality in Europe and in the United States: estimations from surveys, national accounts and wealth rankings. (June). 8.D.2

Blanchet, T., Fournier, J., and Piketty, T. (2017). Generalized Pareto Curves: Theory and Applications. *WID.world Working Paper*, 2017/3(March). 36, 8.D.2

Blanchflower, D. G. and Oswald, A. (1994). *The Wage Curve*. MIT Press. 16, 18.1, 114

Blanchflower, D. G. and Oswald, A. J. (2005). The Wage Curve Reloaded. *NBER Working Paper*, (11338). 114

Blundell, R., Griffith, R., and Van Reenen, J. (1999). Market value, and innovation in a panel of British manufacturing firm. *Review of Economic Studies*, 66(3):529–554. 5, 2.1

Bok, D. (1993). *The Cost of Talent: How Executives and Professionals Are Paid and How It Affects America*. 8

Brynjolfsson, E. and Oh, J. H. (2012). The attention economy: Measuring the value of free digital services on the internet. *International Conference on Information Systems, ICIS 2012*, 4:3243–3261. 9

Burdett, K. and Mortensen, D. T. (1998). Wage Differentials, Employer Size, and Unemployment. *International Economic Review*, 39(2):257–273. 22.E, 22.E

Byrne, D. M., Fernald, J., and Marshall, R. (2016). Does the United States have a productivity slowdown or a measurement problem ? *Brookings Papers on Economic Activity*. 9

Calvo, G. A. (1983). Staggered Prices in a Utility Maximising Framework. *Journal of Monetary Economics*, 12:383–398. 17.1

Cameron, A. C. and Miller, D. L. (2015). A Practitioner's Guide to Cluster- Robust Inference. *The Journal of Human Resources*, (October). 25, 26, 27, 28, 29, 30, 31, 34, 35, **??**, **??**

Cantoni, D. and Bursztyn, L. (2016). A Tear in the Iron Curtain: The Impact of Western Television on Consumption Behavior. *The Review of Economics and Statistics*, 98(2):254 – 267. 1

Card, D. (1995). The Wage Curve : A Review. *Journal of Economic Literature*, 33(2):785–799. 16

Carruth, A., Collier, W., and Dickerson, A. (2004). Inter-industry Wage Differences and Individual Heterogeneity. *Oxford Bulletin of Economics and Statistics*, 5:811–847. 20

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *Quarterly Journal of Economics*, 129(4):1553–1623. 3.3

Chetverikov, D., Larsen, B., and Palmer, C. (2016). IV Quantile Regression for Group-Level Treatments, With an Application to the Distributional Effects of Trade. *Econometrica*, 84(2):809–833. 4.1, **??**, 8.D.1, 18, **??**

Chodorow-Reich, G. and Karabarbounis, L. (2016). The Cyclicality of the Opportunity Cost of Employment. *Journal of Political Economy*, 124(6):1563–1618. 105

Chong, A. and La Ferrara, E. (2009). Television and Divorce: Evidence From Brazilian Novelas. *Journal of the European Economic Association*, 7(2-3):458–468. 1

Cook, P. J. and Frank, R. (1995). *The Winner-Take-All Society*. 2

Costa, D. L. (1998). *The Evolution of Retirement: An American Economic History, 1880-1990*. University of Chicago Press. 9

Costain, J. S. and Reiter, M. (2008). Business cycles, unemployment insurance, and the calibration of matching models. *Journal of Economic Dynamics and Control*, 32(4):1120–1155. 22.C, 22.C

Costinot, A. and Vogel, J. (2010). Matching and Inequality in the World Economy. *Journal of Political Economy*, 118(4):747–786. 1

De Loecker, J. and Eeckhout, J. (2017). The Rise of Market Power and the Macroeconomic Implications. 8, 1

Della Vigna, S., Lindner, A., Reizer, B., and Schmieder, J. (2017). Reference-Dependent Job Search: Evidence from Hungary. *Quarterly Journal of Economics*, 132. 16, 17.4

DellaVigna, S. (2009). Psychology and Economics : Evidence from the Field. *Journal of Economic Literature*, 47(2):315–372. 16

DellaVigna, S. and Kaplan, E. (2007). The Fox News Effect: Media Bias and Voting. *The Quarterly Journal of Economics*, 3(122):1187–1234. 1

DellaVigna, S. and Paserman, M. D. (2005). Job Search and Impatience. *Journal of Labor Economics*, 23(3). 22.E

Devereux, P. J. and Hart, R. A. (2006). Real Wage Cyclicality of Job Stayers, Within-Company Job Movers, and Between-Company Job Movers. *Industrial and Labor Relations Review*, 60(1):105–119. 18.2, 114

Durante, R., Pinotti, P., and Tesei, A. (2015). The Political Legacy of Entertainment TV. *Working Paper*, (April). 1, 27, 10.1

Edmans, A. and Gabaix, X. (2016). Executive Compensation. *Journal of Economic Literature*, 54:150. 1

Eeckhout, J. and Kircher, P. (2018). Assortative Matching With Large Firms. *Econometrica*, 86(1):85–132. 1, 20, 65

Elsby, M. W. L. and Michaels, R. (2013). Marginal Jobs, Heterogeneous Firms, and Unemployment Flows. *American Economic Journal: Macroeconomics*, 5(1):1–48. 18.1, 18.1, 22.C

Enikolopov, R., Petrova, M., and Zhuravskaya, E. (2011). Media and political persuasion: Evidence from Russia. *American Economic Review*, 101(7):3253–3285. 27, 10.1

Fabiani, S., Galuscak, K., Kwapil, C., Lamo, A., and Room, T. (2010). Wage Rigidities and Labor Market Adjustment in Europe. *Journal of the European Economic Association*, 8(2-3):497–505. 18.3, 18.6, 24

Faggio, G. and Nickell, S. (2005). The Responsiveness of Wages to Labour Market Conditions in the UK. *Labour Economics*, 12:685–696. 114

Falk, A., Fehr, E., and Zehnder, C. (2006). Fairness Perceptions and Reservation Wages - The Behavioural Effects of Minimum Wage Laws. *The Quarterly Journal of Economics*, (November):1347–1381. 16, 17.4

Farber, H. S. (2008). Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers. *America Economic Review*, 98(3):1069–1082. 16

Feenberg, D. R. and Poterba, J. M. (1993). *Income Inequality and the Incomes of Very High-Income Taxpayers: Evidence from Tax Returns*, volume 7. 36, 8.D.2

Feldstein, M. and Poterba, J. (1984). Unemployment Insurance and Reservation Wages. *Journal of Public Economics*, 23:141–167. 19.2

Fenton, G. and Koenig, F. (2018). The Labor Supply Response to Innovation in Entertainment: Evidence from TV. *Working paper*. 1, 3.2, 27, **??**, **??**, 4

Finkelstein, M., Tucker, H. G., and Alan Veeh, J. (2006). Pareto Tail Index Estimation Revisited. *North American Actuarial Journal*, 10(1):1–10. 8.D.2

Fujita, S. and Ramey, G. (2009). The Cyclicality of Separation and Job Finding Rates. *International Economic Review*, 50(2). 18.1

Fujiwara, Y., Di Guilmi, C., Aoyama, H., Gallegati, M., and Souma, W. (2004). Do Pareto-Zipf and Gibrat laws hold true? An analysis with European firms. *Physica A*, 335(1-2):197–216. 8.C

Gabaix, X. and Landier, A. (2008). Why Has CEO Pay Increased So Much? *Quarterly Journal of Economics*, 123(1):49–100. 2, 1, 2.1, 2.1, 8.C

Gabaix, X., Moll, B., Lasry, J.-M., and Lions, P.-L. (2016). The Dynamics of Inequality. *Econometrica*, 84(6):1–45. 2, 2.1

Garicano, L. (2000). Hierarchies and the organization of knowledge in production. *Journal of political Economy*, 108(5):874–904. 9, 2.1

Garicano, L. and Hubbard, T. N. (2007). The return to knowledge hierarchies. *Working Paper 12815, National Bureau of Economic Research.* 2

Geerolf, F. (2014). A Static and Microfounded Theory of Zipf ' s Law for Firms and of the Top Labor Income Distribution. *Working Paper*, pages 1–76. 9

Geerolf, F. (2016). A Theory of Pareto Distributions. *Working Paper*, (May):1–49. 8.C

Genesove, D. and Mayer, C. (2001). Loss Aversion and Seller Behavior. 99, 17.4

Gentzkow, M. (2006). Television and voter turnout. *The Quarterly Journal of Economics*, 121(August):931–971. 1, 24, 28, 4.1.1, 9

Gentzkow, M. and Shapiro, J. (2008). Preschool Television Viewing and Adolescent Test Scores: Historical Evidence from the Coleman Study. *Quarterly Journal of Economics*, 123(February):279–323. 1, 4.1.1, 9

Gertler, M., Huckfeldt, C., and Trigari, A. (2016). Unemployment Fluctuations, Match Quality, and the Wage Cyclicality of New Hires. *NBER Working Paper Series*, 22341(2002):71. 19.1

Gertler, M., Sala, L., and Trigari, A. (2008). An Estimated Monetary DSGE Model with Unemployment and Staggered Nominal Wage Bargaining. *Journal of Money, Credit and Banking*, 40(8). 16, 17

Gertler, M. and Trigari, A. (2009). Unemployment Fluctuations with Staggered Nash Wage Bargaining. *Journal of Political Economy*, 117(1):38–86. 16, 17, 19.1

Gibbons, R. and Katz, L. (1992). Does Unmeasured Ability Explain Inter-Industry Wage Differentials? *The Review of Economic Studies*, 59(3):515–535. 20

Goldin, C. (1998). Labor Markets in the Twentieth Century. In *The Cambridge Economic History of the United States*, number 58, pages 549–624. 9

Goldin, C. and Olivetti, C. (2013). Shocking Labor Supply : A Reassessment of the Role of World War II on Women ' s Labor Supply. *American Economic Review P&P*, 103(3):257–262. 9

Goolsbee, A. and Klenow, P. (2018). Internet Rising, Prices Falling: Measuring Inflation in a World of E-Commerce. *American Economic Review P&P*, (January 2014):488–492. 9

Goos, M., Manning, A., and Salomons, A. (2010). Explaining Job Polarization: Routinization and Offshoring. 104(8):1–20. 6, 1

Gottschalk, P. (2005). Downward Nominal-Wage Flexibility: Real or Measurement Error? *The Review of Economics and Statistics*, 87(3):556–568. 18.3, 18.6

Gregg, P., Machin, S., and Fernandez-Salgado, M. (2014). Real Wages and Unemployment in the Big Squeeze. *Economic Journal*, 124:408–432. 114

Guellec, D. and Paunov, C. (2017). Digital Innovation and the Distribution of Income. 3

Haefke, C., Sonntag, M., and van Rens, T. (2013). Wage rigidity and job creation. *Journal of Monetary Economics*, 60(8):887–899. 16, 17, 18.2

Hagedorn, B. M. and Manovskii, I. (2008). The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited. *America Economic Review*, 98(4):1692–1706. 16, 18.1, 18.1, 22.C

Hall, R. (2005). Employment Fluctuations with Equilibrium Wage Stickiness. *American Economic Review*, 95:50–65. 16, 18.2

Hall, R. E. and Milgrom, P. R. (2008). The Limited Influence of Unemployment on the Wage Bargain. *America Economic Review*, 98(4):1653–1674. 16, 18.2

Hall, R. E. and Mueller, A. I. (2018). Wage Dispersion and Search Behavior: The Importance of Nonwage Job Values. *Journal of Political Economy*, 126(4):1594–1637. 19.2

Haris, C. and Laibson, D. (2013). Instantaneous Gratification. *The Quarterly Journal of Economics*, 128:205–248. 22.E, 22.E

Hines, J., Hoynes, H., and Krueger, A. B. (2001). Another Look at Whether a Rising Tide Lifts All Boats. In *The Roaring Nineties: Can Full Employment Be Sustained?* Russell Sage Foundation, New York. 19.1, 114

Holzer, H. (1986a). Black Youth Nonemployment Duration and Job Search. In Freeman, R. and Holzer, H., editors, *The Black Youth Employment Crisis*. 19.2

Holzer, H. J. (1986b). Reservation Wages and their Labor Market Effects for Black and White Male Youth. *The Journal of Human Resources*, 21(2):157–177. 19.2

Jones, S. R. G. (1988). The Relationship Between Unemployment Spells and Reservation Wages as a Test of Search Theory. *The Quarterly Journal of Economics*, (November). 19.2

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–292. 17.4

Kaplan, S. N. and Rauh, J. (2013). It's the Market: The Broad-Based Rise in the Return to Top Talent. *Journal of Economic Perspectives*, 27(3):35–56. 1

Katz, L. F. and Murphy, K. M. (1992). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *The Quarterly Journal of Economics*, 107(1):35–78. 1, 8.D.1

Kleven, H. J., Landais, C., Saez, E., and Schultz, E. A. (2013). Taxation and International Migration of Superstars: Evidence from European Football Market. *American Economic Review*, 103(5):1892–1924. 8.D.1

Kleven, H. J., Landais, C., Saez, E., and Schultz, E. A. (2014). Taxation and International Migration of Top Earners: Evidence from the Foreigner Tax Scheme in Denmark. *The Quarterly Journal of Economics*, 129(1):333–378. 8.D.1

Kline, P., Petkova, N., Williams, H., and Zidar, O. (2017). Who Profits from Patents? Rent-Sharing at Innovative Firms. (1151497). 8

Koszegi, B. and Rabin, M. (2006). A Model of Reference-Dependent Preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165. 17.4

Krueger, A. B. (2017). Where Have All the Workers Gone? An Inquiry into the Decline of the U.S. Labor Force Participation Rate. *Brookings Papers on Economic Activity*. 9

Krueger, A. B. and Mueller, A. (2011). Job Search, Emotional Well-Being, and Job Finding in a Period of Mass Unemployment: Evidence from High-Frequency Longitudinal Data. *Brookings Papers on Economic Activity*, (1):1–57. 16, 19.2

Krueger, A. B. and Mueller, A. I. (2012). Time Use, Emotional Well-Being, and Unemployment : Evidence from Longitudinal Data. *American Economic Review (paper & proceedings)*, 102(3):594–599. 16, 19.2

Krueger, A. B. and Mueller, A. I. (2016). A Contribution to the Empirics of Reservation Wages. *American Economic Journal: Economic Policy*, 8(1):142–179. 16, 19.2

Krueger, A. B. and Summers, L. H. (1988). Efficiency Wages and the Inter-Industry Wage Structure. *Econometrica*, 56(2):259–293. 18.1, 19.1, 20

Krusell, P., Mukoyama, T., and Sahin, A. (2010). Labour-Market Matching with Precautionary Savings and Aggregate Fluctuations. *The Review of Economic Studies*, 77(4):1477–1507. 106

Kudlyak, M. (2014). The cyclicality of the user cost of labor. *Journal of Monetary Economics*, 68:53–67. 16, 17

Kuznets, S. and Jenks, E. (1953). *Shares of Upper Income Groups in Savings*, volume I. 36, 8.D.2

Lancaster, T. and Chesher, A. (1983). An Econometric Analysis of Reservation Wages. *Econometrica*, 51(6):1661–1676. 19.2

Le Barbanchon, T., Rathelot, R., and Roulet, A. (2018). Unemployment insurance and reservation wages: Evidence from administrative data. *Journal of Public Economics*, pages 1–38. 16, 19.2

Loomes, G. and Sugden, R. (1986). Disappointment and Dynamic Consistency in Choice under Uncertainty. *The Review of Economic Studies*, 53(2):271–282. 17.4

Manning, A. (2003). *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton university press, Princeton. 7, 18.3, 20, 22.E

Manning, A. (2011). Imperfect Competition in the Labour Market. *Handbook of Labor Economics*. 24

Melitz, M. J. and Ottaviano, G. I. (2008). Market size, trade, and productivity. *Review of Economic Studies*, 75(1):295–316. 7

Michaels, G. and Graetz, G. (2018). Robots at Work. *Review of Economics and Statistics*. 1

Michaillat, P. (2012). Do Matching Frictions Explain Unemployment? Not in Bad Times. *America Economic Review*, 102(4):1721–50. 16

Moretti, E. and Wilson, D. (2017). The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists. *American Economic Review*, 107(7):1858–1903. 8.D.1

Mortensen, D. T. and Nagypál, É. (2007). More on unemployment and vacancy fluctuations. *Review of Economic Dynamics*, 10:327–347. 22.C

Murphy, K. M., Shleifer, A., and Vishny, R. W. (1993). Why Is Rent-Seeking So Costly to Growth? *American Economic Review P&P*, 83(2). 8

Nijkamp, P. and Poot, J. (2005). The Last Word on the Wage Curve? *Journal of Economic Surveys*, 19(3):421–450. 16

Nordhaus, W. D. (2006). Principles of national accounting for nonmarket accounts. In *A New Architecture for the U.S. National Accounts*, number May, pages 143–160. 9

OECD (2016). The Productivity-Inclusiveness Nexus. Technical Report June. 3

Olken, B. A. (2009). Do television and radio destroy social capital? Evidence from indonesian villages. *American Economic Journal: Applied Economics*, 1(4):1–33. 1, 27, 10.1

Paserman, M. D. (2008). Job Search and Hyperbolic Discounting : Structural Estimation and Policy Evaluation. *Economic Journal*, 118(531):1418–1452. 22.E

Penniman, C. and Heller, W. (1959). *State income tax administration*. Public Administration Service, University of Michican. 8.D.1

Petterson, S. M. (1998). Black-White Differences in Reservation Wages and Joblessness: A Replication. *The Journal of Human Resources*, 33(3):758–770. 19.2

Piketty, T. (2014). *Capital in the 21st Century*. Harvard University Press. 4.3.3

Piketty, T. and Saez, E. (2003). Income Inequality in the United States, 1913-1998. *The Quarterly Journal of Economics*, 68(February):1–39. 36, 4.3.3, 8.D.1, 8.D.2, 11

Piketty, T., Saez, E., and Stantcheva, S. (2014). Optimal Taxation of Top Labour Incomes: A Tale of Three Elasticities. *American Economic Journal: Economic Policy*, 6(1):230–271. 8, 8.D.1

Pissarides, C. A. (2000). *Equilibrium Unemployment Theory*. MIT Press, Cambridge, second edition. 16, 17, 17.1

Pissarides, C. A. (2009). The Unemployment Volatility Puzzle: Is Wage Stickiness the Answer? *Econometrica*, 77(5):1339–1369. 18.1, 18.2

Polivka, A. E. (2000). Using Earnings Data From the Monthly Current Population Survey. *SSRN Electronic Journal*, (October). 8.D.2

Powell, D. (2016). Quantile Regression with Nonadditive Fixed Effects. *RAND Working Paper*, (April):1–28. 8.D.1

Putnam, R. (1995). Bowling Alone: America's Declining Social Capital. *Journal of Democracy*, 6(1):65–78. 1

Robin, J.-M. (2011). On the Dynamics of Unemployment and Wage Distributions. *Econometrica*, 79(5):1327–1355. 18.1

Rogerson, R. and Shimer, R. (2011). Search in Macroeconomic Models of the Labor Market. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics, Vol. 4*, pages 619–700. 100, 22.C

Rosen, S. (1981). The Economics of Superstars. *American Economic Review*, 71(5):845–858. 3, 1, 9, 2.1, 2.1

Rudanko, L. (2009). Labor market dynamics under long-term wage contracting. *Journal of Monetary Economics*, 56:170–183. 16, 17

Ryscavage, P. (1988). Reservation Wages and Subsequent Acceptance Wages of Unemployed Persons. *SIPP Working Paper*, (66). 19.2

Sanderson, E. and Windmeijer, F. (2016). A weak instrument F -test in linear IV models with multiple. *Journal of Econometrics*, 190(2):212–221. 31

Sattinger, M. (1975). Comparative advantage and the distributions of earnings and abilities. *Econometrica*, 43(3):455–468. 3, 1, 2.1, 2.1, 14, 8.C

Sattinger, M. (1979). Differential Rents and the Distribution of Earnings. *Oxford Economic Papers*, 31(1):60–71. 1, 2.1

Sattinger, M. (1993). Assignment Models of the Distribution of Earnings. *Journal of Economic Literature*, 31(2):831–880. 65

Scheuer, F. and Werning, I. (2017). The Taxation of Superstars. *Quarterly Journal of Economics*, 132(1):211–270. 67, 8.D.1

Shimer, R. (2005). The Cyclical Behavior of Equilibrium Unemployment and Vacancies. *America Economic Review*, (March). 16, 100, 17

Shimer, R. and Werning, I. (2007). Reservation Wages and Unemployment Insurance. *The Quarterly Journal of Economics*, 122(3):1145–1185. 20

Silva, I. J. and Toledo, M. (2009). Labor Turnover Costs and the Cyclical Behavior of Vacancies and Unemployment. *Macroeconomic Dynamics*, 13(Supplement 1):76–96. 17.1

Spinnewijn, J. (2015). Unemployed but Optimistic: Optimal Insurance Design with Biased Beliefs. *Journal of the European Economic Association*, 13(1):130–167. 22.E

Sterne, J. (1999). Television under construction: American television and the problem of distribution, 1926-62. *Media, Culture & Society*, 21(199905):503–530. 4

Stigler, G. J. (1961). The Economics of Information. 1, 2.2.2

Syverson, C. (2017). Challenges To Mismeasurement Explanations for the U.S. Productivity Slowdown. *Journal of Economic Perspectives*, 31(2):165–186. 9

Taylor, J. B. (1999). Staggered Price and Wage Setting in Macroeconomics. In Taylor, J. B. and Woodford, M., editors, *Handbook of Macroeconomics*. Elsevier. 18.3, 18.6, 24

Terviö, M. (2008). The difference that CEOs make: An assignment model approach. *American Economic Review*, 98(3):642–668. 2, 1, 2.1, 2.1

Teulings, C. N. (1995). The Wage Distribution in a Model of the Assignment of Skills to Jobs. *Journal of Political Economy*, 103(2):280–315. 1

Tinbergen, J. (1956). On the Theory of Income Distribution. *Weltwirtschaftliches Archiv*, 77(2):155–175. 3, 1, 2.1

Wagner, J. (1994). German Wage Curves , 1979-1990. *Economic Letters*, 44:307–311. 114

Yanagizawa-Drott, D. (2014). Propaganda and Conflict: Theory and Evidence From the Rwandan Genocide. *Quarterly Journal of Economics*, 129(4):1947–1994. 10.1