Skill Costs and the Rise of Firm-level Productivity Dispersion

Jian Tang*

This version: November 5, 2022. Click here for the latest version.

Abstract

Firm-level productivity dispersion has increased dramatically in the US since the 1980s. I examine the contribution of rising skilled wages for this phenomenon, given that skilled labor is a key input for technological innovation and adoption. My hypothesis is that, to the extent that the skilled labor input required for technology adoption does not fully scale with current firm-level productivity, less productive firms will be particularly affected by higher skilled wages and will therefore adopt relatively less compared to high productivity firms. My contribution is twofold. First, using firm-level data from Compustat, I find that one third of the observed increase in productivity dispersion since the 1980s can indeed be attributed to differential productivity growth between low vs high productivity incumbents, leading to a decline in the rate of productivity convergence. I also find that the channel of rising skill costs is responsible for the near entirety of the decline in convergence. Second, I build a simple version of Hopenhayn's industry equilibrium model to illustrate the link between higher technological adoption costs and the rate of productivity convergence. In the calibrated model, I find that the observed increase in the skill premium generates over 75% of the increase in dispersion that is related to the observed decline in the rate of productivity convergence.

Keywords: productivity dispersion, convergence, skill premium, technology adoption

JEL Codes: D24, J31, L25, O30

*PhD candidate, Department of Economics, McGill University, jian.tang2@mail.mcgill.ca. I would like to thank Rui Castro, Markus Poschke for their advice and support. I am also grateful to all the McGill macro group members, and participants at McGill-CIREQ Lunch Seminar, the 56th Annual Conference of the Canadian Economics Association and the 17th CIREQ Ph.D. Students' Conference for their insightful comments.

1 Introduction

The firm-level productivity dispersion has been increasing in the recent decades in the United States and many other OECD countries. It is one of the pronounced stylized facts related to business dynamism in the US since the 1980s and has been cited as a contributing cause of the productivity slowdown, wage and income dispersion, and declining labor share. (Akcigit and Ates, 2021; Andrews et al., 2015; Berlingieri et al., 2017; Gouin-Bonenfant et al., 2018). This paper traces a large part of the rising dispersion to a lower rate of productivity convergence between high and low productivity firms, and attributes most of this effect to higher relative skilled wages.

I first document a similar trend of rising productivity dispersion among the US public firms from the 1980s to the 2000s. Based on the firm-level data in Compustat, the standard deviation of productivity increased by about 40% during this period. I also find that one third of this increase – a magnitude of 12% to 13% – can be attributed to a decline of productivity convergence among incumbent firms.

This paper then explores the role of higher skilled wages in slowing down the productivity convergence. I hypothesize that the rising skill premium from the 1980s to the 2000s made new technology adoption more costly for all firms, but more so for less productive ones, resulting in a relative slowdown of these firms compared to high productivity ones. Indeed, I find empirical evidences showing that, following the observed increase in aggregate skill premium, less productive firms exhibits less productivity growth compared to more productive ones. The magnitude of the effect is economically large. My estimates suggest that over 90% of the declining convergence and its dispersion effect can be accounted for by a higher skill premium.

In the second part of my paper, I develop a simple model of technology adoption to make the link between the skill premium and productivity dispersion explicit, and to provide a quantitative evaluation of my mechanism. The model is a simple variant of industry equilibrium model in Hopenhayn (1992). With some probability, firms have access to a new idea which may have a high impact on productivity growth. Adopting such an idea is costly, and skilled labor intensive. As relative skilled wages grow, all firms adopt less. However, to the extent that the skilled labor input required for adoption does not fully scale with current productivity, less productive firms are particularly affected.

In order to evaluate this mechanism quantitatively, I calibrate the model to be consistent with the level of productivity dispersion from Compustat firm in the 1980s. I then feed into the model the skilled wage that is consistent with the skill premium growth observed from the 1980s to the 2000s, and ask whether the model generates an increase in productivity dispersion induced by weaker convergence, as in the data. My model is able to generate 75% of the observed increase.

Overall, these results suggest that, (i) declining productivity convergence is an important contributor to the rising dispersion, and (ii) the rising skill cost is a major economic force behind the declining convergence.

There is a large literature also documenting and trying to understand the rise of productivity dispersion. As Syverson (2019) summarizes, the existing explanations for increasing dispersion and growing productivity gaps focus on various market imperfections, such as rising market power and concentration (De Loecker et al., 2020; Aghion et al., 2019; Covarrubias et al., 2020; Dorn et al., 2017), and increasing technology and knowledge diffusion costs (Andrews et al., 2015; Akcigit and Ates, 2021). These market frictions lead to growing disparities of technology gains and costs, and therefore generates different patterns of productivity growth among firms. In contrast, this paper does not assume any form of market imperfections and studies the heterogeneous effect of a higher skilled wage on productivity growth and the impact on productivity convergence.

This paper also contributes to the recent empirical literature which finds both evidences of technology advantage of leading firms (Tambe et al., 2020; Stiebale et al., 2020; Zolas et al., 2021) and delays in technology adoption by left-tail firms (Berlingieri et al., 2020) especially in information technology and skill-intensive industries. This study provides a particular explanation for these differential patterns of technology adoption between the frontier and laggard firms and highlights the role of skill costs for these phenomena.

Along the lines of linking skilled labor to technology adoption, researchers have found that, relative skilled wages, skill supply and skill efficiency, are all highly important for technology adoption and firm growth (Hjort et al., 2021; Signorelli, 2020; Lewis, 2011; Cavenaile et al., 2019). Following this strand of literature, in my model, I assume technology adoption only requires skilled labor input and examine the distributional effect of rising skill costs.

This study also speaks to the large body of work looking at how rising skill premium or the underlying skill-biased technology change fosters shifts of firm-level heterogeneity. A recent study by Criscuolo et al. (2021) provides cross-country evidences that skill composition and management competency explains about a third of the observed differences in productivity across firms, and frontier firms are more likely to hire a larger share of skilled workers. Poschke (2018) shows that the skill-biased technology change leads to the increase of dispersion in firm size distributions through entrepreneurial choices across countries. Earlier works, including Dunne et al. (2004), Caselli (1999), Katz et al. (1998), Doms et al. (1997), among many others, document how firm-level or plant-level wages, occupational mix, workforce education, and productivity vary with the adoption of new technology that complements skilled workers. This study particularly emphasizes the asymmetric effect of rising skill costs on technology adoption. In response to the increase in skill costs, high productivity firms are relatively more innovative than low productivity firms. Hence, the results are qualitatively in line with the empirical results found in the above literature.

The remaining of the paper is organized as following: Section 2 describes the basic facts, lays out the empirical strategy, and presents the main empirical findings, Section 3 describes the model, Section 4 gives the quantitative analysis, Section 5 concludes.

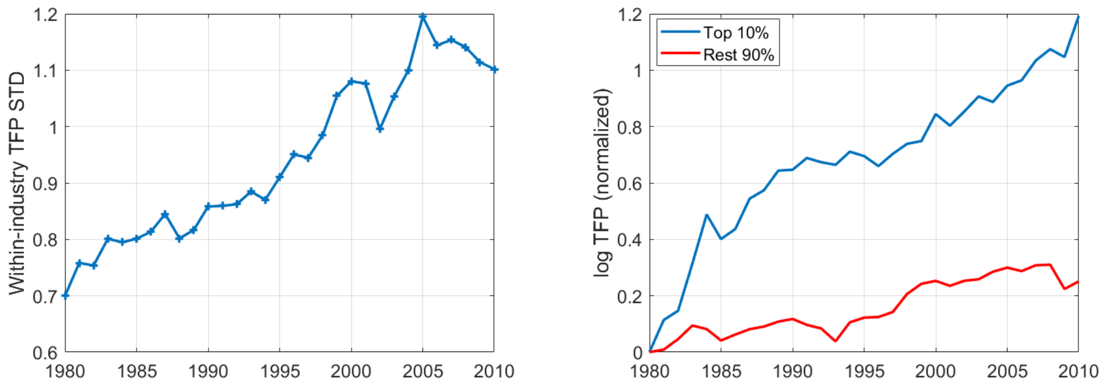
2 An empirical analysis of productivity dispersion

2.1 Productivity dispersion and convergence

In this section, I first document the evolution of firm-level productivity dispersion among the publicly listed US firms available in Compustat database, then I quantify how much of the increase can be attributed to declining convergence. In order to be comparable with the literature, I focus on the period between 1980 and 2010, during which the dispersion increased the most. Total factor productivity (TFP) is computed by taking into account of endogenous investment, following Olley and Pakes (1996). And to obtain the dispersion, a customary step, as described by Bartelsman et al. (2018), is to compute the within-industry log TFP by sweeping out industry and time effects. The dispersion based on the within-industry log TFP is also known as with-industry dispersion.[1]

Figure 1 provides two popular ways of describing the rising productivity dispersion in the literature. The left panel shows a drastically rising within-industry dispersion of log TFP among the US public firms. The right panel of Figure 1 describes depicts two very different growth paths of the top 10% vesus the rest 90% of the firms. With Compustat data, it replicates the phenomenon of the "the Best vs. the Rest" dynamics in Andrews et al. (2015) which shows a similar graph using the firm-level

[1] See Appendix A for details about sample selection and TFP computation.



(a) With-industry dispersion in Compustat

(b) Widening gaps of "the Best vs. the Rest"

Figure 1. Rising productivity dispersion among the US public firms

data from the OECD countries to highlight the widening productivity gaps between the frontier and the laggard firms.

It is noteworthy that the standard deviation of within-industry log TFP increases by about 40% from the 1980s to the 2000s. The magnitude is large compared to what has been reported in the literature using the US census data. The difference not only depends on the measurement of productivity but also on the particular samples. More specifically, it is well-known from Davis et al. (2006) that Compustat firms had a volatility surge since the 1980s when there was a large influx of newly listed firms which were more volatile in employment growth than older cohorts. The increasing disturbance from the entrants can in principle drive up the standard deviation of the distribution and has been used as an evidence against Compustat data for similar studies. This study acknowledges the selective nature of Compustat and investigates how much of rising dispersion can be attributed to the changes of rate of productivity convergence, which is presumably more associated with the fundamental business dynamism than the volatility shocks.

To this end, Figure 2 plots the dynamics of relative TFP growth rates of the firms across different initial TFP quintiles. Take the top left panel for instance. Each point represents the difference of five-year-average TFP growth rates between the lowest productive (first quintile) and the highest productive (fifth quintile) firms based on their initial average TFP levels five years earlier. It is evident that the TFP process features mean reversion, namely, less productive firms grow relatively fast. And it is also clear that such convergence force becomes weaker during this period since the relative growth rates declined. Figure 2 also indicates that the gaps of TFP growth rates are becoming smaller. This result coincides with another stylized fact of business dynamism found by Decker et al. (2016) that the dispersion of firm growth drops while the TFP dispersion increases. For the purpose of this study, the declining convergence rate could contribute to the rising dispersion, but it is not yet obvious to know how important this channel could be.

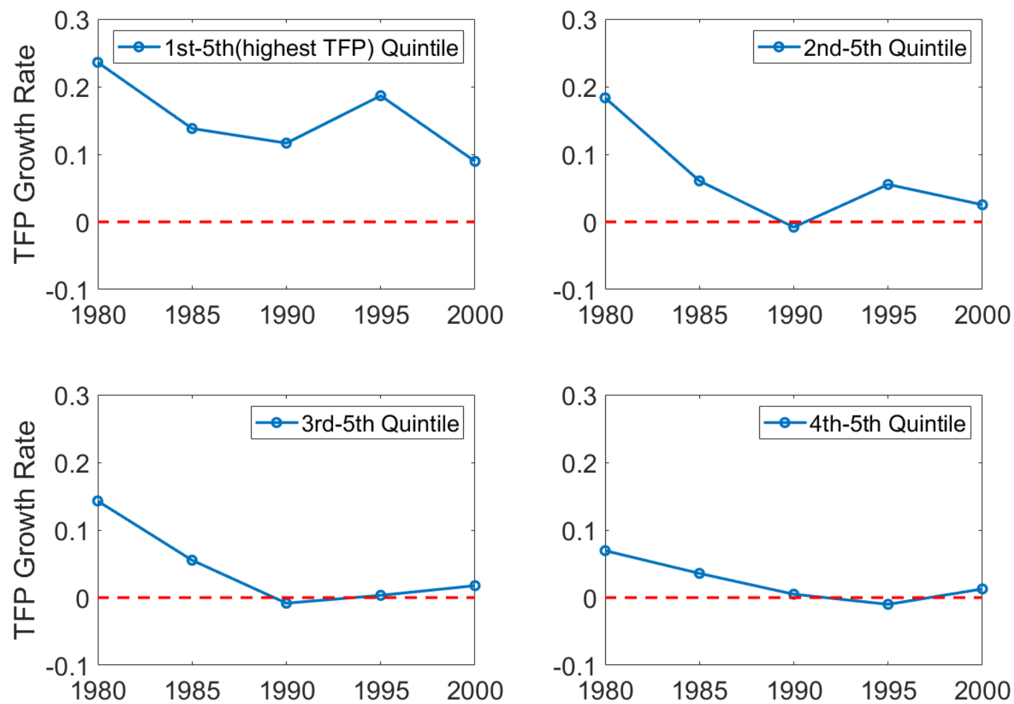


Figure 2. Evolution of relative TFP growth rates

To further quantify such contribution, I first estimate the convergence process to obtain the convergence rate for each decade. The simple regression is:

$$\Delta \log z_{i,t+1} = \alpha + \beta \log z_{i,t} + v_{i,t+1}$$

where $\Delta \log z_{i,t+1}$ is the TFP growth rate for firm i at period $t+1$ while $\log z_{i,t}$ is the initial TFP level at t. $v_{i,t+1}$ is the firm-specific shock. The β coefficient is known as the convergence coefficient in the economic growth literature. It is also convenient to control for the firm-specific permanent component and other variables related to the TFP process. The results are summarized in Table 1, where the first panel is pooled

OLS and the second panel is the fixed effect model with controls including firm size, capital stock and R&D intensity.

Table 1. TFP convergence rates in three decades

$\Delta \log z_{t+1}$	$1980s$	$1990s$	$2000s$
A: Pooled OLS			
$\log z_t$	-0.245	-0.209	-0.184
	(0.005)	(0.005)	(0.006)
B: Fixed Effects & Controls			
$\log z_t$	-0.706	-0.639	-0.626
	(0.025)	(0.023)	(0.025)
Observations	11,837	12,865	10,086

Note: all the estimated coefficients are significant at 0.01 level.

The results in Table 1 confirms that the TFP convergence has been slowing down. The absolute value of the estimated convergence coefficients decreases by 0.06 – 0.08. This implies that firms that are one standard deviation more productive than the mean – which is about 90% higher – relatively grow 5.4 – 7.2 percentage points more in the 2000s than in the 1980s in annual terms. This is a sizable effect even for the US public firms which on average grow about 5% annually in Compustat. With these coefficients, it is convenient to compute the implied change of dispersion using the simple time series formula. Note that the convergence regression from the pooled OLS can be written as

$$\log z' = \alpha + (1+\beta)\log z + v$$

The standard deviation of the stationary distribution of $\log z$ is simply

$$\sigma_z = \frac{\sigma_v}{\sqrt{1-(1+\beta)^2}} \tag{1}$$

Other things equal, the change of β from the 1980s to the 2000s implies the stationary standard deviation should rise by 13.4%. It is also possible to derive a similar formula for the fixed effect model (See Appendix B), which is

$$\sigma_z = \sqrt{\frac{\sigma_c^2}{\beta^2} + \frac{\sigma_v^2}{1-(1+\beta)^2}} \tag{2}$$

where σ_c is the standard deviation of the firm-specific permanent term. From the 1980s sample, estimated residual standard error σ_v is 0.32 and estimated σ_c is 1.82. Keeping these terms constant, the change of β based on the fixed effect regression gives an effect of 12.6% higher in stationary dispersion. The two models therefore provide a similar magnitude of the effect on the productivity dispersion, which accounts for about one third of the overall increase of productivity dispersion observed in Compustat.

To sum up, recognizing that the firms in the Compustat sample are selective and many specific elements could contribute to the rising productivity dispersion, I find one third of the increase in dispersion is associated with the declining productivity convergence, namely, relative slowdown of low productivity firms compared to the high productivity firms.

2.2 Convergence and skill costs

In this section, I explore the economic force behind the declining TFP convergence and hypothesize that rising skill costs have made technology progress more costly and particularly affected less productive firms more. This could be one of the forces that slow down the convergence process and increase the dispersion. I test this hypothesis empirically and then quantify the effect of this specific channel based on the regression outcomes.

Firstly, I measure skill costs by the series of skill premium or college premium which is equal to the ratio of the full-time labor incomes between college graduates and non-college individuals in the US. unskilled wage is essentially used to deflate the skilled wage based on the assumption that firms can always opt for unskilled production rather than acquiring costly new knowledge. I obtain these prices from the data of Current Population Survey (CPS) annual labor income by education. The measured skill premium increased by 18% from the 1980s to the 2000s.

To find out how firms of different TFP levels respond to the shocks of the skill premium in multiple time horizons, I run the following local projection regressions

$$\log z_{i,t+h} - \log z_{i,t} = \alpha_i^h + \alpha_{s,t}^h + \beta^h \log z_{i,t} + \gamma^h(\log z_{i,t} \times w_{t+1}) + \theta^h X_{i,t} + v_{i,t+h} \tag{3}$$

where $\log z_{i,t+h}$ is the log TFP for firm i at period $t+h$ and $h = 1, 2, ..., 5$, α_i is the firm-specific fixed effect, $\alpha_{s,t}$ is sector-year-fixed effect (sector defined on Compustat 2-digit SIC code), w_{t+1} is growth rate of skill premium at $t+1$ which is common for all firms and $X_{i,t}$ controls for firm characteristics including firm size, capital stock and R&D intensity. The superscript h on the RHS denotes the parameters associated with the regression at the horizon h. The coefficient of interest is γ^h which measures the differential responses across the levels of initial TFP to the growth rate of the aggregate skill premium. A positive sign of γ^h indicates that more productive firms respond to the increase of skill premium more positively.

Table 2. Heterogeneous responses to skill premium growth

$\Delta_h \log z$	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$
$\log z \times w$	0.304	0.752**	0.783**	0.895*	0.478
	(0.264)	(0.364)	(0.396)	(0.484)	(0.344)
Obs.	37,467	30,146	24,526	20,120	16,644
Controls & Fixed effect	✓	✓	✓	✓	✓
R^2	0.411	0.562	0.659	0.700	0.734

Note: *p<0.1; **p<0.05; ***p<0.01. All standard errors are clustered at firm level.

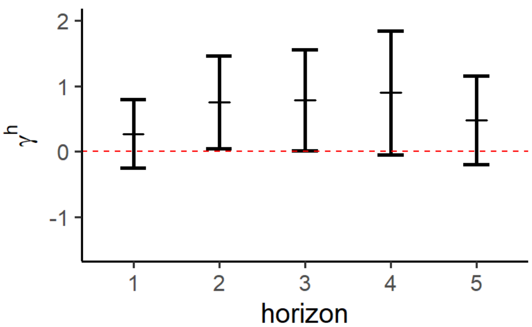
The result of the regression is reported in Table 2. Although the contemporaneous effect at $h = 1$ is not significant with the clustered standard errors, the signs of γ^h tends to be positive and significantly positive from $h = 2$ to $h = 4$, implying a lagging effect of the skill premium growth. And such effect indicates that indeed less productive firms tend to grow relatively slower than more productive ones following a rise in the skill premium. This specific asymmetric growth channel may be important to explain the relative change of growth rates or the declining convergence seen in the last section. The result is robust to including the post-2010 sample (see more details in Appendix E).

To quantify the change of stationary dispersion due to this effect, the coefficients of the local projection in Table 2 are firstly annualized. This is because the local projection generates only short-term responses. The annualized effect from the first four periods is about 0.3. This in turn decreases the absolute value of the convergence rate – which is the β coefficient – by 0.054 following an 18% skill premium growth from the 1980s to the 2000s. In words, firms with TFP one standard deviation higher than the mean respond 4.86 percentage points more following an 18% skill premium growth. This is a magnitude quite close to the decline in the convergence rate obtained in the last section. Using the estimated parameters and the same dispersion formula for the fixed effect model in the last section, the resulting increase in stationary dispersion is 11.2%, which is more than 90% of the rising dispersion induced by the declining convergence.[2] It indicates that the rising skill premium could be a crucial trend that is associated with the declining productivity convergence.

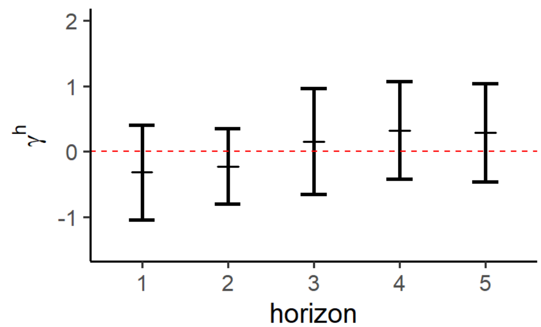
A natural question related to this local projection exercise is whether or not the effect is causal. Generally, the local projection regression is not causal *per se*. But it's worth noting that all the independent variables on the RHS of the regression are either predetermined or aggregate, meaning that the likelihood of a reverse casual effect is low. One concern is that the series of skill premium growth might have picked up other persistent aggregate trends. One way to mitigate this concern is by advancing the shock series by a few periods. Since future skill premium growth should not affect the current technology adoption decision or at least in a lesser magnitude, the effect is expected to be smaller. Figure 3 compares the regression results implementing this method. The right panel of Figure 3 provides a placebo exercise by feeding in the skill premium shocks that will happen only five years later. Compared with the left panel which comes from the regression (3), the effects in the right panel are not only smaller but cannot be distinguished from zero given the 95% confidence interval with the clustered standard errors.

One remaining concern is about how to interpret the result, in particular, the economic implication of rising skill premium. It is intertwined with the identification problem in the following way. The rising skill premium could indicate two phenomena, namely, the simple increasing costs of skilled workers and the underlying technology change or skilled-biased technology change (SBTC). Since the rising skill premium is commonly treated as a proxy in the literature for the SBTC, it is unknown if the asymmetric responses are associated with the rising skill costs or the technology advancement or both. It is plausible that the asymmetric effects imply that all firms

[2] Again, this is one third of the overall increase in dispersion. It should be noted that a more robust comparison would require a benchmark fixed effect model that entirely mirrors (3) but differs only by replacing the skill premium shocks with the decade dummies. This helps make the fixed effects constant across the decades just as in (3). See Appendix C for such exercise.



(a) Coefficients from the local projection (3)



(b) Coefficients from the placebo projection

Figure 3. Robustness check for other trends

benefit from the technology advancement while the better firms benefit more,[3] or alternatively, all firms suffer from more expensive skill costs while worse firms suffer more. A mixed effect from these two separate forces is also possible, which could lead to a scenario where an exogenous technology shock potentially benefits all firms but, as a result of the rising costs, only benefits a few who can afford the skill costs. This is essentially in line with the simple cost shock interpretation. As the regression (3) cannot estimate the average effect of the skill premium directly because of the included time fixed effect, this problem cannot be addressed with the above regression setup. Informally, I replace the time fixed effects with the polynomials of time periods. The average effects tend to be negative but are not statistically significant (See Appendix D). It should be noted that, although it raises the question about the underlying force at play, this problem does not compromise the asymmetric effect hence the implied increase in dispersion at all.

Despite the limitation of (3), it is still possible to address this concern by exploring the other dimension of firm heterogeneity as a "sign identification." Specifically, I run the following regression.

$$\log z_{i,t+h} - \log z_{i,t} = \alpha_i^h + \alpha_{s,t}^h + \beta^h \log z_{i,t} + \delta^h (r_{i,t} \times w_{t+1}) + \theta^h X_{i,t} + v_{i,t+h}$$

where $r_{i,t} = \frac{R\&D_{i,t}}{Sales_{i,t}}$, is the R&D intensity and I use it as a proxy for skill or knowledge intensity, following the implication of Machin and Van Reenen (1998). The rest of the variables are the same as in (3). The idea behind this regression is that, if the

[3]Nevertheless, the mechanism of such asymmetric effect due to the SBTC is not straightforward at least in the conventional literature on the directed technology change (e.g., Acemoglu 2002).

skill premium growth represents purely a positive technology change that benefits some of the productive firms more, and since the SBTC by its nature complements the skilled workers, it should also benefit the firms which are initially more skill or knowledge intensive. However, because the skill costs increase simultaneously, the cost effect also particularly affects these firms but in a negative way since they have a larger share of skill cost exposure. If the result shows that skill premium growth makes these firms worse off, it is an evidence that the cost channel associated with the skill premium is more important than the pure technology shocks. Therefore, the key coefficient of interest is δ^h. The regression result is given in Table 3.

Table 3. Negative responses among R&D intensive firms

$\Delta_h \log z$	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$
$r \times w$	-0.015^{**}	-0.035^{***}	-0.024^{**}	-0.031^{**}	-0.006
	(0.007)	(0.010)	(0.011)	(0.013)	(0.012)
Obs.	37,467	30,146	24,526	20,120	16,644
Controls & Fixed effect	✓	✓	✓	✓	✓
R^2	0.443	0.584	0.659	0.701	0.734

Note: *p<0.1; **p<0.05; ***p<0.01. All standard errors are clustered at firm level.

Table 3 shows that the coefficients of the interaction term are indeed persistently negative and significant. Evidently, the negative TFP responses are particularly dominant among the firms with higher initial R&D intensity. Although this is not a direct test on the average effect of skill premium, it further characterizes the effect of skill premium as a negative impact on the growth for the firms that are more R&D intensive, which are presumably more skill intensive and more exposed to the skill cost shocks. Such negative impact therefore supports the interpretation that the rising skill premium could be the burden of cost on the productivity progress to some type of firms.

To summarize the empirical results, the rising skill premium generates asymmetric productivity growth among the firms and leads to a decline of convergence. It accounts for more than 90% of the increasing dispersion associated with the declining convergence. Furthermore, the rising skill premium is characterized as a cost shock since it also impacts more negatively on the firms with higher initial skill cost exposure, measured by the R&D intensity. However, it still remains unclear how the skill cost induces the asymmetric productivity growth. This motivates a more structural characterization of the skill cost effect on the rate of productivity convergence.

3 Model

In this section, I start with a static technology adoption model and show the basic mechanism why the rising skill costs have an asymmetric effect on the productivity growth. Then I embed the simple framework in a standard industry equilibrium model in the spirit of Hopenhayn (1992). Lastly, I calibrate the dynamic model to quantify the dispersion effect of rising skill premium.

3.1 A static model

The basic model is static and stylized. There are a number of firms in the market. Each firm is operating on a firm-specific productivity z, and draws a new technology idea g from some distribution, then decides on the adoption of the new idea. That is, firms solve

$$\max\{zg - \kappa g, z\}$$

where κ is the skill cost. One can think κ represents the skilled wage, or the skill cost that accounts for certain amount of skilled labor required for the adoption. For simplicity, κ is referred as skill cost in this section. The total adoption cost increases in g linearly. If the firm adopts the new idea g, the payoff is $zg - \kappa g$, and if not, the payoff remains z. Here, the new idea g is immediately realized in payoff after the firm pays the adoption cost κg. It is important to notice that the adoption of new idea generates an growth effect, which is a common assumption in the endogenous growth literature, and the adoption benefit zg is always larger for firms who start to be more productive.

The analysis of the problem is fairly simple. The cutoff rule or the adoption threshold for each adopting firm is

$$\bar{g} = \frac{z}{z - \kappa}$$

The firm will adopt if $g > \bar{g}$. If the adoption threshold is higher, the overall adoption probability is lower and hence the firm is less innovative. The value of threshold crucially depends on the initial productivity z. It is easy to show that $\frac{\partial \bar{g}}{\partial z} < 0$, implying that the adoption threshold is lower for larger z. This is because despite a

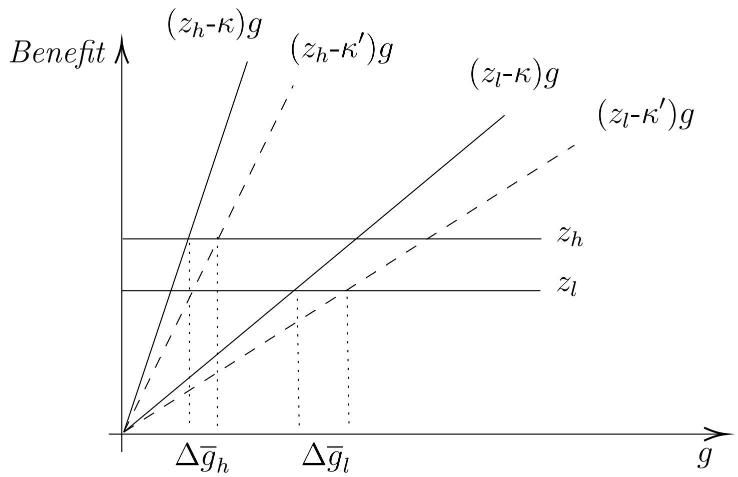


Figure 4. Graph of the asymmetric effect on the adoption thresholds

larger reservation value equal to z, an increase of z raises the benefit of technology adoption more since $\bar{g}$ is always greater than 1. And simple algebra further yields

$$\frac{\partial \bar{g}}{\partial \kappa} = \frac{\bar{g}}{z-\kappa} > 0, \quad \frac{\partial}{\partial z}\frac{\partial \bar{g}}{\partial \kappa} = -\frac{\kappa}{(z-\kappa)^3} < 0 \tag{4}$$

The first inequality shows that the adoption threshold must increase in κ for each z. The skill cost increase should make all adopting firms less innovative. This is straightforward since the rising cost makes firms more picky in adopting new ideas. The second inequality shows that the magnitude of the increase in adoption threshold is decreasing in z, which implies that more productive firms should be less affected by the rising skill cost.

Figure 4 provides a graphic illustration of the decision problems of two firms with productivity $z_h > z_l$. The upward sloping lines represent the net adoption benefits while the horizontal lines of z_h and z_l are the reservation values. First it is clear that the adoption threshold of firm z_l is higher than that of firm z_h because the less productive firm needs better ideas to cover the same skill cost. And because the adoption benefit of firm z_h is steeper, the same shift from κ to κ' causes less increase in $\bar{g}_h$ than in $\bar{g}_l$.

This asymmetric effect arises precisely because the technology adoption induces a growth effect on the productivity and the payoff of adoption therefore depends on the initial levels of productivity. The skill cost is however fixed relative to the levels of productivity. As a result, a rise of skill cost will particularly affect those firms whose adoption benefits are lower. In general, to the extent that the skill input required for adoption does not fully scale with current firm-level productivity, less productive firms will be particularly affected by higher skilled costs and therefore will adopt relatively less compared to high productivity firms.

How does the effect on the thresholds map to the relative change of growth rates? To obtain an analytic answer to this question, one has to impose an assumption for the g distribution. As a simple example, suppose it is a uniform distribution $U(0, G)$. The impact of rising skill cost on productivity growth is similar to that on the adoption probability mentioned above. The expected gross growth rate for each adopting firm has two components, namely, growing with adoption or staying at the current level. That is,

$$\eta(z,\kappa) = \int_{\bar{g}(z,\kappa)}^{G} g\ U(g) + \int_{0}^{\bar{g}(z,\kappa)} 1\ U(g)$$

or

$$\eta(z,\kappa) = \frac{G-\bar{g}(z,\kappa)}{G} \cdot \frac{G+\bar{g}(z,\kappa)}{2} + \frac{\bar{g}(z,\kappa)}{G}$$

where η denotes the expected growth rate of firm z facing the skill cost κ. And with (z,κ) subdued

$$\frac{\partial \eta}{\partial \kappa} = -\frac{1}{G}\frac{\partial \bar{g}}{\partial \kappa}(\bar{g}-1) < 0$$

since $\bar{g} > 1$ and $\frac{\partial \bar{g}}{\partial \kappa} > 0$. This is saying that the expected growth will decline in the skill cost, which is intuitive. Similarly, with the inequalities (4), one can show

$$\frac{\partial}{\partial z}\frac{\partial \eta}{\partial \kappa} = -\frac{1}{G}\Big[\frac{\partial \bar{g}}{\partial z}\frac{\partial \bar{g}}{\partial \kappa} + (\bar{g}-1)\frac{\partial}{\partial z}\frac{\partial \bar{g}}{\partial \kappa}\Big] > 0.$$

Therefore, facing rising skill cost κ, more productive firms on average grow relatively faster than less productive ones because of less affected adoption. Lower productive firms grow even less since they wait longer and become less active in adoption, although conditional on adopting, the adopted growth rate is higher for these firms since the increase in $\bar{g}$ raises the expected value of technology draws.

Aside from the main purpose of this model, there are other important observations that can be drawn from the static model. The first is declining aggregate productivity growth rate. This can be trivially shown as an increase in κ decreases technology adoption hence productivity growth for all firms. Thus the model can speak to the well-documented trend of slowdown in productivity growth. Moreover, it explains why the increase in productivity gaps or rising dispersion is associated with decreasing aggregate or industrial productivity growth, which is investigated by Andrews et al. (2015) and Kehrig (2015). It also implies that the dispersion of firm growth rates will drop while the productivity dispersion increases, as noted in Decker et al. (2016). The decline of dispersion in firm growth rates is simply because all firms are less innovative and stay inactive more often after the skill cost rises.

3.2 Dynamics

In order to examine the effect of increasing skill cost with the firm-level data in Compustat through the above simple mechanism, I embed the static setup in an industry dynamic model in the spirit of Hopenhayn (1992) with firm entry and exit, which is a workhorse model in industry dynamics and has been applied widely to study firm-level heterogeneity. In the conventional Hopenhayn type of models, the firm-level productivity process is assumed to be exogenous, usually featuring a mean reversion or AR(1) process. The key difference between the model in this paper and the conventional type is that the model in this paper allows firms to purposefully grow their own productivity by adopting new technology, in addition to receiving idiosyncratic productivity shocks from an AR(1) process. In this way, I can employ this model to assess the effect of increasing skill cost and generate a productivity convergence and distributions that can be comparable with data.

In this model, time is discrete and the horizon is infinite. The economy consists of a mass of heterogeneous firms.

Incumbent firms. At time t, a positive mass of price-taking firms produces a homogeneous good according to the following production function:

$$y = (e^z)^{1-\alpha} n_u^\alpha$$

where $\alpha \in (0, 1)$, z is the log productivity, and n_u denotes the employment of unskilled workers which is the only factor input in production. Firms' profit maximization problem is

$$\Pi(z) = \max_{n_u} (e^z)^{1-\alpha} n_u^\alpha - w_u n_u$$

where w_u is the unskilled worker's wage. A firm's productivity is driven by exogenous shocks and the outcomes of the firm's technology adoption. The law of motion of the log productivity is therefore a mixture of two processes. In each period, the firm receives a productivity shock from an exogenous AR(1) process. In addition, with probability $a \in (0, 1)$, it has an opportunity of drawing a new idea g from a normal distribution $N(0, \sigma_g)$, and decides on the adoption of the idea g. If the firm adopts, its productivity grows by g, and if not, it stays with whatever it received from the AR(1) innovations. The idea distribution is symmetric with mean 0, implying that there are equal chances of drawing good and bad ideas. For simplicity, I assume firms exit the market permanently with constant probability $(1 - s)$. The decision of adoption and the values of g will depend on the firms' value function, which is given by

$$W(z) = \Pi(z) + s\beta\mathbb{E}_{z'|z}\Big(a\mathbb{E}_g \max\big\{W(z' + g) - \kappa\exp(g), W(z')\big\} + (1 - a)W(z')\Big)$$

where the first conditional expectation represents the exogenous AR(1) component, $z' = \rho z + \epsilon$, with ρ being the persistence coefficient, and $\epsilon \sim N(m_\epsilon, \sigma_\epsilon)$. It consistently shifts the current levels of productivity for all firms. The second expectation takes across the possible realizations of new ideas. a denotes the arrival probability of such adoption opportunity. If $a = 0$, the process is identical to an AR(1), which is the standard Hopenhayn setup. If the firm adopts a new idea g, it pays the adoption or implementation cost, which is equal to $\kappa\exp(g)$, and its log productivity improves immediately by g. Again, κ denotes the skill cost. Since g affects log productivity, the adoption cost is assumed to increase in g exponentially so that the cost can be measured by firm profit under the same scale. This is in line with the setup of the static model as z and g here can be both substituted by their exponential forms. The continuation value is discounted by β and subject to the exogenous survival rate s.

The decision of adoption, the arrival probability of new ideas, the survival rate and the exogenous shocks determine the productivity transition. Denote the productivity transition operator by Q, which summarizes the effect of these transitions on the firm productivity distribution.

Entrants. There is a large mass of potential entrants, of which a constant mass $m > 0$ enters the market each period. And entry requires a sunk investment c_e in units of unskilled labor. New entrants draw their initial level of productivity from a distribution $H(z)$. An entrant starts business as long as its value exceeds its cost, that is

$$\int W(z)dH(z) \geq w_u c_e.$$

This formulation implies that upon paying entry cost, entrants' productivity is revealed and they only hire labor and operate in the subsequent production cycle. Under free entry condition it must be that, in equilibrium, the expected value of entry equals its cost whenever there is positive entry.

Together with the incumbents and exits, they drive the law of motion of the productivity distribution of active firms μ, that is,

$$\mu' = Q\mu + mH.$$

The stationary distribution exists as long as the mean reversion process is dominant or the exogenous exit is significant.

In addition, I assume labor is exogenous and inelastically supplied. As the setup stays close to Hopenhayn (1992), it also features a partial equilibrium.

Definition of stationary equilibrium. A stationary equilibrium consists of value function $W(z)$, policy function $n_u(z), g(z)$, price level w_u, a measure of incumbent firms μ, and a mass of entrants m such that: given a skill cost κ,

1. $W(z), n_u(z), g(z)$ solve the incumbent firm's problem
2. The free entry condition holds: $\int W(z)dH(z) = w_u c_e$
3. The labor market clears at w_u
4. The distribution of incumbent firms is stationary: $\mu' = \mu$.

Effect of rising skill costs on thresholds. Figure 5 illustrates the effect of rising skill costs. On the y-axis, any values above the thresholds are the adoptable ideas. $\kappa' > \kappa$, representing a higher skill cost corresponding to a higher skilled wage. The adoption thresholds shift up for all firms under κ', but the thresholds shift up more for low productivity firms, precisely for the reasons aforementioned. As a result of the higher skill cost, low productivity firms will relatively adopt less technology and therefore grow less.

4 Quantitative analysis

In this section, I first calibrate the model to match some of the key moments of the 1980s firm sample in Compustat. Then I change the skill cost to measure the effect of rising skill cost on the productivity dispersion, by matching the skill premium growth according to the empirical data while fixing all the other parameters at the 1980s levels.

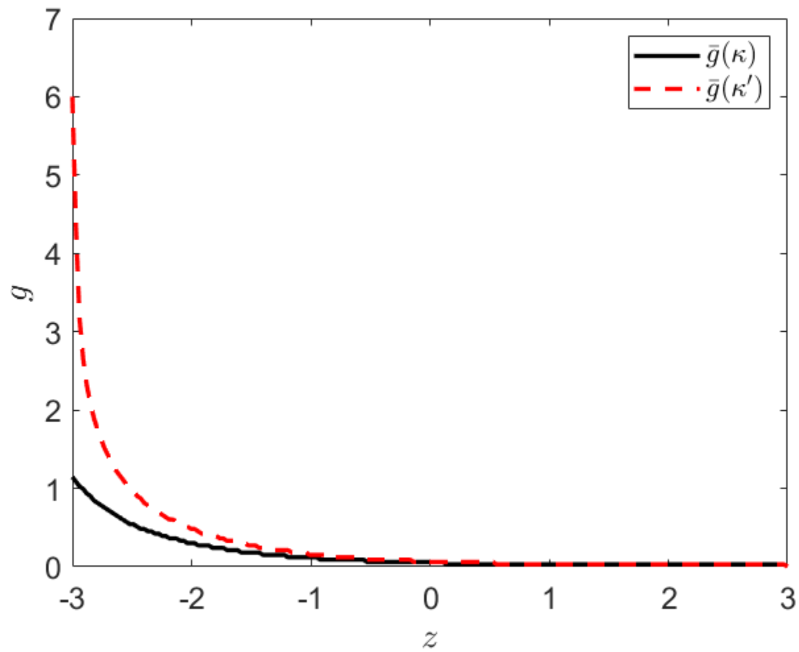


Figure 5. Change of adoption thresholds when κ rises

4.1 Model calibration

I preset a few baseline parameters based on prior information, and jointly calibrate the remaining parameters to match the characteristics of the firm size distribution and TFP process in Compustat.

The length of a time period is set to one year. The discount rate is set to be 0.96. The labor share is 2/3 or 0.67. The share of profit is therefore 1/3. Ideally, the labor share could be set to the Compustat labor share that is computed in the empirical analysis, but it would also require a production function involving capital, which inevitably complicates the model and computation without too much gain. The average one-year survival rate in the 1980s is 0.957 in Compustat. Since Compustat does not officially report exiting firms, I compute the proportion of firms who no longer report financial data as the exit rate. The idea distribution is assumed to be normal with mean zero. Lastly, I preset the mean of entry distribution to 11.21, which is the first 3-year average of log TFP of the new firms entering the 1980s sample. The mean of entrants' productivity is externally set provided that I will need to capture the average productivity across all firms to inform the parameters of the exogenous AR(1) process in the model, which will be discussed shortly. The preset parameters are summarized in Table 4.

Table 4. Calibration summary: preset parameters

Name	Parameter	Value
Labor share	α	2/3
Discount	β	0.96
Survival rate	s	0.957
Mean of entry distribution	m_H	11.21
Mean of idea distribution	m_g	0

The remaining parameters are jointly calibrated so that the model generates realistic distributions of productivity and its process.

One important block of parameters are the standard deviation σ_ϵ, mean m_ϵ, and persistence coefficient ρ that characterize the primitive AR(1) process in the model. Note that they cannot be preset to their empirical counterparts directly because in the model the productivity process is a mixture of two separate schemes of processes. However, the empirical AR(1) can indeed provide a set of auxiliary estimates that inform the corresponding primitive parameters. In order to get these statistics from the model, I simulate the model to obtain the sample paths of productivity, based on the transition matrix Q without including the survival rate, and then run the AR(1) regressions.

The rest of the parameters are the entry cost, c_e, standard deviation of idea distribution, σ_g, standard deviation of entry distribution, σ_H, arrival probability of new adoption opportunity, a, and the level of skill cost, κ. The entry cost affects the average value of operating firms and hence the average firm size. I target the average firm size in Compustat in the 1980s which is about 4,040. The standard deviation of idea distribution is crucial in this model as it determines the quality of a new idea and how far the firm can move upward after each adoption. A larger σ_g implies more firm growth and larger firm size at the top. The share of employment in the 10% largest firms – which is about 78% in the 1980s – is therefore informative about σ_g. For σ_H, I obtain information from the share of employment in small firms, which often are young. In Compustat, firms below average size account for 19% of employment. I calibrate the above three parameters following Gabler and Poschke (2013) closely since both models share some similar features about productivity process despite different datasets. Another important parameter to pin down is the arrival probability of new

Table 5. Calibration summary: model statistics and targets

Name	*Data*	*Model*
Auxiliary AR(1) *of log TFP*:		
Estimated standard error of residuals	0.49	0.52
Estimated persistence	0.75	0.73
Estimated mean	11.27	11.30
Other targeted moments:		
Average firm size (thousands)	4.04	4.06
Share of employment in largest 10% firms	0.78	0.78
Share of employment in firms below average size	0.19	0.20
Median/mean size	0.10	0.09
Standard deviation of log TFP in the 1980s	0.92	0.92
Untargeted moments (not used in calibration):		
Share of employment in largest 5% firms	0.65	0.68
Share of employment in largest 25% firms	0.93	0.94
fraction of firms below average size	0.84	0.87

Table 6. Calibration summary: jointly calibrated parameters

Name	*Parameter*	*Value*
Primitive AR(1) parameters:		
Standard deviation	σ_ϵ	0.37
Coefficient	ρ	0.64
Mean	m_ϵ	3.98
Other parameters:		
Entry cost (relative to avg profit)	c_e	2.21
Standard deviation of idea distribution	σ_g	1.44
Standard deviation of entry distribution	σ_H	0.81
Arrival probability of new adoption	a	0.63
Skill cost level (relative to avg profit)	κ	0.17

adoption opportunity. a is calibrated to target the ratio between median and mean of the firm size distribution because the higher a is, the more often firms can adopt, resulting in a more skewed firm size distribution. In other words, with a smaller a, the median should be closer to the mean. The median/mean of the firm size distribution is 0.10 in the Compustat data. Lastly, I calibrate κ to target the standard deviation of productivity in the 1980s directly – which is about 0.92 – provided the theoretic implication of rising κ.[4]

Table 5 reports the values of the targeted moments for the data and the model, and Table 6 summarizes all the internally calibrated parameters. The calibration fits reasonably well for the firm size distribution. Even in dimensions that were not targeted, the model statistics, which are fairly comparable to Gabler and Poschke (2013), are close to the moments in the data. In particular, it fits the productivity dispersion in the 1980s perfectly. The target that fits less well is the auxiliary AR(1) parameters which require simulation and regression. The overall fit is acceptable.

Regarding the calibrated parameters, note that the primitive AR(1) coefficient is

[4] This is the standard deviation of average productivity among the 1980s firm sample. One can also use the within-industry dispersion as described in the empirical analysis section. The result does not change much but it's less sensible statistically to take averages of the within-industry productivity measures in each decade which already take the industry-year effects out.

significantly smaller than the empirical counterpart and all the similar estimates in the literature, implying a stronger underlying convergence. But it should be emphasized that endogenous adoption itself is a divergent force since more productive firms adopt more often. Together they generate a relatively reasonable level of persistence. Entry cost is high relative to the average profit since the Compustat firms are typically very large. The standard deviation of idea distribution is considerably higher than that of the random shocks thanks to a concentrated right tail. Together with a large arrival probability of new adoption opportunities and relatively low skill cost level, it implies that the Compustat firms are quite innovative with fairly fast growth. There exists large dispersion among entrants but still substantially falls short of the dispersion among incumbents, which is not entirely consistent with Davis et al. (2006) who find there exists a surge of volatility among new listed firms in the 1980s.[5] The reason that the model predicts a slightly smaller entrant dispersion is because the much skewed firm size distribution requires an important role of endogenous adoption. By contrast, a larger entrant dispersion would imply a much more dominant role of AR(1) process or fewer opportunities of adoption in order to shrink the entrant distribution to a less dispersed distribution of incumbents, which cannot align with other moments. This could be a limitation of the simple assumptions the model relies on.

Another technical limitation of the computation is about the discretization of the idea distribution. The model assumes a common idea distribution for all firms. However, to implement the idea distribution in computation and following Gabler and Poschke (2013), I truncate the support of the idea distribution according to the level of productivity. For example, the firm with highest productivity gets $g = 0$ as the maximum draw and the firm with lowest productivity gets $g = 0$ as the minimum draw from the distribution. For the rest of the productivity levels, the distance from itself to the max (min) grid determines the largest (smallest) g the firm can possibly draw. This setup is clearly for computation simplicity and it helps achieve stationarity in an easier way by limiting the growth of right-tail firms. Hence, it may compromise the divergent mechanism implied by the static model.

4.2 Effect of rising skill cost

In this section, I use the calibrated model to evaluate the effect of rising skill cost on the productivity dispersion. Specifically, I compute the stationary equilibrium asso-

[5] For comparison, the standard deviation of first three-year average of new firms' TFP is about 0.89 in Compustat.

ciated with the new skill cost that matches the actual growth of skill premium from the 1980s to the 2000s, and compare it with the calibrated benchmark equilibrium with parameters in the 1980s.

Before the experiment, it is useful to establish the link between the skill cost κ and the skill premium that is used in the empirical analysis. This will facilitate a direct comparison between the results from the empirical analysis and the quantitative experiment in this section. To do so, recall that in the previous stylized model, κ denotes skilled wage or skill cost. To be more specific, κ can take the form of

$$\kappa = w_s \bar{n_s}$$

where w_s is skilled wage and $\bar{n_s}$ is the constant amount of skilled labor required for a given level of technology adoption.[6] $\bar{n_s}$ is clearly associated with skill efficiency in technology adoption which will be assumed constant throughout the time.

Note that in the dynamic model, the unskilled worker's wage is determined endogenously by the free entry condition. To connect the model with the empirical result, I vary skill cost κ indirectly by changing the skill premium. To do this, note that κ can be written as

$$\kappa = \frac{w_s}{w_u} w_u \bar{n_s}.$$

This step is to factor out the skill premium $\frac{w_s}{w_u}$ from κ. To understand the effect of the rising skill premium, first notice that when the skill premium rises, κ will increase, which leads to a lower firm's value for all incumbent firms. In order to have free entry condition hold, the unskilled wage must decrease to raise the operating profits and hence the firm's value. The drop of unskilled wage is mitigated – but not entirely offset – by the effect of decreasing entry cost since entry cost is in the unit of unskilled labor. As a result, κ rises less than the skill premium because equilibrium w_u drops.

Now let the skill cost κ increase such that the skill premium $\frac{w_s}{w_u}$ grows by 18% as observed in the data, while keeping all the other parameters constant. Table 7 summarizes the results.

First note that indeed the unskilled wage drops by about 3%, an equilibrium effect that reduces the impact of rising skill premium. And average firm size increases be-

[6] Alternatively, in the spirit of Bloom et al. (2020) as well as similar endogenous growth models, one can consider the technology upgrading process as a linear production function $g = \theta n_s$ where n_s is the skilled labor input and θ is the skill efficiency. Denote the skilled wage as w_s. Then, in the context of the static model, the net adoption benefit becomes $zg - \frac{w_s}{\theta} g$. Let $\kappa = \frac{w_s}{\theta}$. The decision problem in the previous model remains the same.

Table 7. Effect of an 18% higher skill premium

Variables	κ	κ'
unskilled wage	2.41	2.34
Average firm size (thousands)	4.06	4.18
Share of employment in largest 10% firms	0.78	0.74
Share of employment in firms below average size	0.20	0.28
Median/mean size	0.09	0.11
Productivity dispersion	0.92	1.01

cause the labor cost is lowered. Since fewer firms can effectively move to the top, the share of employment in the largest 10% firms drops whereas the share of employment in the small firms increases, and the skewness of the distribution – measured by median/mean ratio – improves slightly. The effect on the small firms is noticeably larger since the rising skill cost particularly affects more on low productivity firms. The log productivity dispersion increases by about 9.8%, which is not large considering the overall 40% increase in dispersion but quite significant if compared with the 12-13% increase induced by the declining convergence. Since it is a model of asymmetric growth, the rising skill premium explains about 75% of dispersion induced by the relative change of growth rates.

To illustrate the comparison of rising productivity dispersion based on empirical and structural models, Figure 6 shows three growth trends from different computation sources. The bold line represents the increase in dispersion induced by the declining convergence using the fixed effect estimates in (2).[7] At the risk of mislabelling, it is named as "Data" to emphasize that this trend imposes the minimum of assumptions and is indeed one part of the dispersion decomposition of the overall trend in data. This trend grew the most from the 1980s to the 1990s by more than 10% while slowed down during the 2000s. The overall increase is about 12.6%. The dot line presents the estimated results from the regression (3) where the skill premium growth is shown to be associated with the asymmetric productivity growth. In total it finds an 11% increase in dispersion overall, out of which 8% increase during the second decade. The

[7] The reason to use the fixed effect model (2) rather than the pooling OLS (1) is obviously for a better comparison with the regression (3). But the growth differences across decades are indeed less obvious with the OLS results.

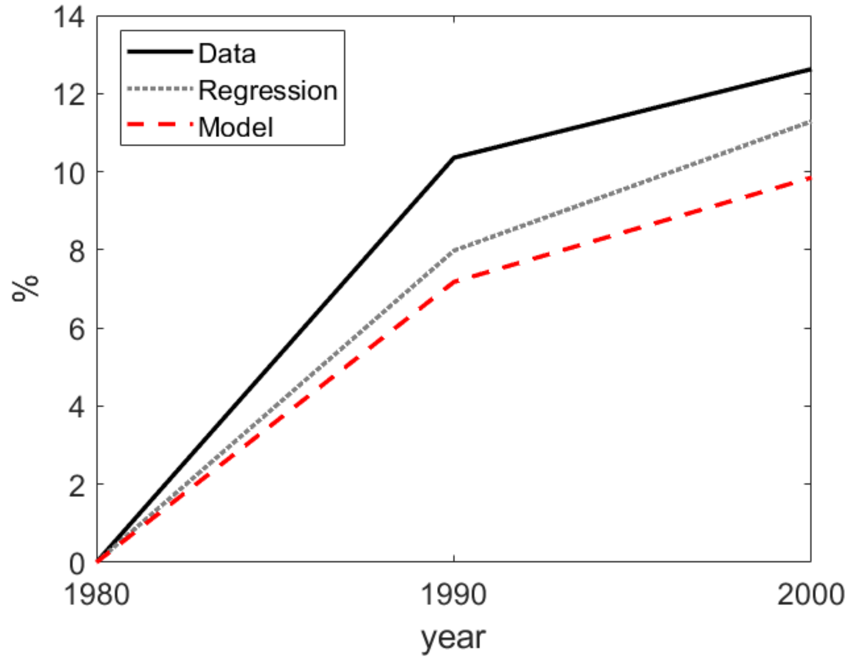


Figure 6. Comparison of changes of productivity dispersion

dashed line shows the outcome from the calibrated model. It predicts an increase of 9.8% in dispersion, which is about 75% of the Data trend. It also shows that 7% of the increase happened in the 1990s. The model gives a lower magnitude of the effect compared to the regression-based outcome, but still provides an evidence that the rising skill cost is an important source of the declining productivity convergence and increasing dispersion.

The dispersion effect generated by the model is not as large as the one estimated by the local projection regression in Section 2.2, although the model implies a larger drop of the absolute value of convergence rate, which is shown in the left panel of Figure 7. There are many things that could lead to this gap. One noticeable reason is associated with the estimated standard error of residuals in the AR(1) regressions. Recall that in the empirical analysis, I examine the effect declining convergence rate by changing the β coefficient only while holding all the other variables constant, that is, in (1), (2) and (3), the estimated standard error is held to its the 1980s level. This is intuitive since without the structural model, no clear relationship between the coefficient and the standard error could be drawn and it is better to treat the term as an exogenous shock. This is not the case with the model. In the stationary equilibrium, when the skill cost rises and affects less productive firms more negatively,

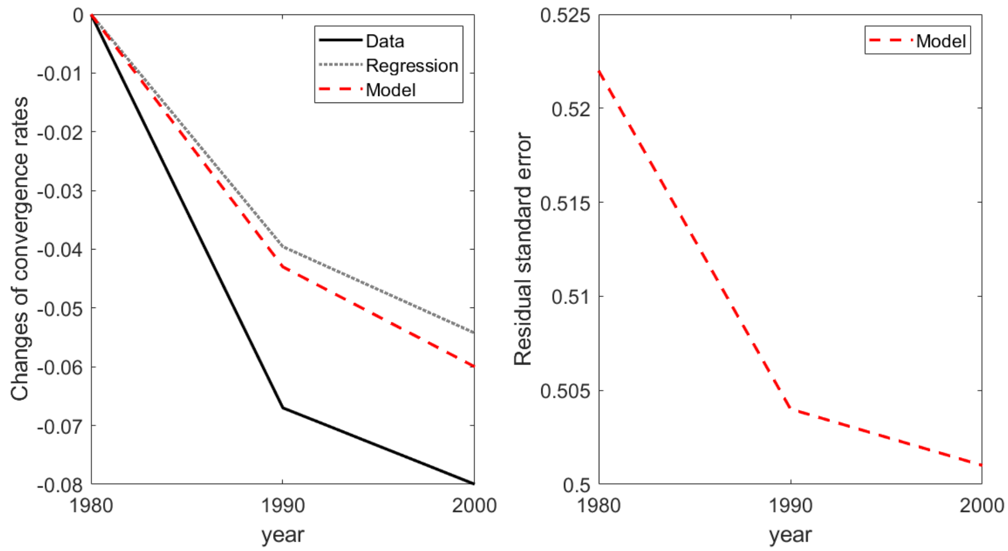


Figure 7. Residual standard errors and comparison of changes of convergence rates (in absolute values)

it not only generates a decline of convergence rate which can relate to the reduce-formed estimates, but it also makes the firms less innovative overall, which is shown by the decreasing standard error estimated by the auxiliary AR(1) regression in the right panel of Figure 7. In this panel, the residual standard errors from Data and Regression is irrelevant or treated as constant. Indeed, the overall decrease of convergence rate from the model is 0.06, slightly larger than the change of β coefficient estimated in the local projection (3) in Section 2.2. But here the standard error also drops by about 0.02 and effectively reduces the impact of rising skill cost. This channel is numerically significant. Keeping standard error the same and using only the change of convergence coefficients, the increase in dispersion implied by the auxiliary AR(1) is about 11%. But including the decline of standard errors, the increase in dispersion implied by the auxiliary AR(1) alone is only 8%. Through the lens of the model, the evolution of convergence rates is associated with the changes of growth volatility. Although such relationship cannot be found in Compustat given the large increase in volatility during this period, it can speak to the general business dynamism noted by Decker et al. (2016) and Akcigit and Ates (2021).

5 Conclusion

Since the 1980s, the differences in productivity performance across firms in the US and other OECD countries have been increasing. In this paper, using the firm-level data from Compustat, I have shown that one third of the increasing productivity dispersion can be attributed to a decline of productivity convergence among incumbent firms. I hypothesize that the rising skill premium from the 1980s to the 2000s made new technology adoption more costly for all firms, but more so for less productive ones, resulting in a relative slowdown of these firms compared to high productivity ones. Indeed, I find empirical evidences showing that, following the observed increase in aggregate skill premium, less productive firms exhibits less productivity growth compared to more productive ones. The magnitude of the effect is economically large. My estimates suggest that over 90% of the declining convergence and its dispersion effect can be accounted for by a higher skill premium.

To understand the basic mechanism behind the empirical facts, I develop a simple model of technology adoption to make the link between the skill premium and productivity dispersion explicit. I show that, to the extent that the skilled labor input required for technology adoption does not fully scale with current firm-level productivity, less productive firms will be particularly affected by higher skilled wages and will therefore adopt relatively less compared to high productivity firms. The rising skill cost therefore makes all firms worse off but matters more for firms with lower productivity. As a result, despite the same increase in skill cost, lower productive firms adopt and grow less.

In order to evaluate this mechanism quantitatively, I embed the static model in a standard dynamic framework following Hopenhayn (1992). I calibrate the model to be consistent with the level of productivity dispersion from Compustat firm in the 1980s, and then feed into the model the skilled wage that is consistent with the skill premium growth observed from the 1980s to the 2000s. The calibrated model predicts that 75% of increasing dispersion induced by the declining convergence can be explained by the rising skill premium.

This paper has many policy implications related to the skill costs and the heterogeneity of firm growth. Rather than attribute the rising dispersion to market imperfections, this paper argues that the increasing dispersion or widening productivity gaps between the frontier and laggard firms may be a natural outcome of rising skilled wages, which could be a result of the skill-biased technology change. To the extent that the increase in skilled wage or the technology change is decelerating, the rising productivity gaps might slow down as well, which is indeed supported by the evidence that productivity dispersion slightly decreases after 2010 (Cunningham et al., 2021) together with a slowdown of skill premium growth (more in Appendix E). It also provides a different perspective related to the inequality of technology adoption and R&D activities. When there exists financial market frictions, low productivity firms are more likely to be constrained by higher research expenditure and growing skilled wages. Aside from R&D tax credits or other means of financial support, an education reform or immigration policy in favor of skilled workers might help promote technology upgrading of these laggard firms.

There are other aspects of business dynamism that the model can be employed to investigate. From the static model, the rising skill costs also imply a decline of aggregate productivity growth. It is possible to incorporate a balanced growth in the dynamic model in the spirit of Luttmer (2007) and Poschke (2009). It also speaks to other stylized business dynamism in the US since the 1980s, notably the relationship between increasing productivity dispersion and decreasing volatility of firm growth. The model can qualitatively deliver these features but the quantitative implication remains to be examined.

References

D. Acemoglu. Directed technical change. *The review of economic studies*, 69(4): 781–809, 2002.

P. Aghion, A. Bergeaud, T. Boppart, P. J. Klenow, and H. Li. A theory of falling growth and rising rents. Technical report, National Bureau of Economic Research, 2019.

U. Akcigit and S. T. Ates. Ten facts on declining business dynamism and lessons from endogenous growth theory. *American Economic Journal: Macroeconomics*, 13(1): 257–98, 2021.

D. Andrews, C. Criscuolo, and P. N. Gal. Frontier firms, technology diffusion and public policy: Micro evidence from oecd countries. 2015.

E. J. Bartelsman, Z. Wolf, E. Grifell-Tatjé, and C. K. Lovell. Measuring productivity dispersion. In *The Oxford Handbook of Productivity Analysis*. 2018.

G. Berlingieri, P. Blanchenay, and C. Criscuolo. The great divergence (s). 2017.

G. Berlingieri, S. Calligaris, C. Criscuolo, and R. Verlhac. Laggard firms, technology diffusion and its structural and policy determinants. 2020.

N. Bloom, C. I. Jones, J. Van Reenen, and M. Webb. Are ideas getting harder to find? *American Economic Review*, 110(4):1104–44, 2020.

F. Caselli. Technological revolutions. *American economic review*, 89(1):78–102, 1999.

L. Cavenaile, M. A. Celik, and X. Tian. Are markups too high? competition, strategic innovation, and industry dynamics. *Competition, Strategic Innovation, and Industry Dynamics (August 1, 2019)*, 2019.

M. Covarrubias, G. Gutiérrez, and T. Philippon. From good to bad concentration? us industries over the past 30 years. *NBER Macroeconomics Annual*, 34(1):1–46, 2020.

C. Criscuolo, P. Gal, T. Leidecker, and G. Nicoletti. The human side of productivity: Uncovering the role of skills and diversity for firm productivity. 2021.

C. Cunningham, L. Foster, C. Grim, J. Haltiwanger, S. W. Pabilonia, J. Stewart, and Z. Wolf. Dispersion in dispersion: Measuring establishment-level differences in productivity. 2021.

S. J. Davis, J. Haltiwanger, R. Jarmin, J. Miranda, C. Foote, and E. Nagypal. Volatility and dispersion in business growth rates: Publicly traded versus privately held firms [with comments and discussion]. *NBER macroeconomics annual*, 21:107–179, 2006.

J. De Loecker, J. Eeckhout, and G. Unger. The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, 135(2):561–644, 2020.

R. A. Decker, J. Haltiwanger, R. S. Jarmin, and J. Miranda. Where has all the skewness gone? the decline in high-growth (young) firms in the us. *European Economic Review*, 86:4–23, 2016.

M. Doms, T. Dunne, and K. R. Troske. Workers, wages, and technology. *The Quarterly Journal of Economics*, 112(1):253–290, 1997.

D. Dorn, L. F. Katz, C. Patterson, J. Van Reenen, et al. Concentrating on the fall of the labor share. *American Economic Review*, 107(5):180–85, 2017.

T. Dunne, L. Foster, J. Haltiwanger, and K. R. Troske. Wage and productivity dispersion in united states manufacturing: The role of computer investment. *Journal of Labor Economics*, 22(2):397–429, 2004.

A. Gabler and M. Poschke. Experimentation by firms, distortions, and aggregate productivity. *Review of Economic Dynamics*, 16(1):26–38, 2013.

E. Gouin-Bonenfant et al. Productivity dispersion, between-firm competition, and the labor share. 2018.

J. Hjort, H. Malmberg, and T. Schoellman. The missing middle managers: Labor costs, firm structure, and development. 2021.

H. A. Hopenhayn. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica: Journal of the Econometric Society*, pages 1127–1150, 1992.

L. F. Katz, A. B. Krueger, et al. Computing inequality: have computers changed the labor market? *Quarterly Journal of Economics*, 113(4):1169–1173, 1998.

B. Kaymak and I. Schott. Loss-offset provisions in the corporate tax code and misallocation of capital. *Journal of Monetary Economics*, 105:1–20, 2019.

M. Kehrig. The cyclical nature of the productivity distribution. *Earlier version: US Census Bureau Center for Economic Studies Paper No. CES-WP-11-15*, 2015.

E. Lewis. Immigration, skill mix, and capital skill complementarity. *The Quarterly Journal of Economics*, 126(2):1029–1069, 2011.

E. G. Luttmer. Selection, growth, and the size distribution of firms. *The Quarterly Journal of Economics*, 122(3):1103–1144, 2007.

S. Machin and J. Van Reenen. Technology and changes in skill structure: evidence from seven oecd countries. *The quarterly journal of economics*, 113(4):1215–1244, 1998.

G. Olley and A. Pakes. The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297, 1996.

M. Poschke. Employment protection, firm selection, and growth. *Journal of Monetary Economics*, 56(8):1074–1085, 2009.

M. Poschke. The firm size distribution across countries and skill-biased change in entrepreneurial technology. *American Economic Journal: Macroeconomics*, 10(3): 1–41, 2018.

S. Signorelli. Too constrained to grow analysis of firms' response to the alleviation of skill shortages. 2020.

J. Stiebale, J. Suedekum, and N. Woessner. Robots and the rise of european superstar firms. 2020.

C. Syverson. Creating economic opportunity for more americans through productivity growth. Technical report, Aspen Economic Strategy Group, 2019.

P. Tambe, L. Hitt, D. Rock, and E. Brynjolfsson. Digital capital and superstar firms. Technical report, National Bureau of Economic Research, 2020.

N. Zolas, Z. Kroff, E. Brynjolfsson, K. McElheran, D. N. Beede, C. Buffington, N. Goldschlag, L. Foster, and E. Dinlersoz. Advanced technologies adoption and use by us firms: Evidence from the annual business survey. Technical report, National Bureau of Economic Research, 2021.

Appendices

A Data

The data used for the empirical results is the Compustat database, which provides balance sheet information on the US public firms. The sample selection criterion and definition of variables follow Kaymak and Schott (2019) closely. I drop firms that ever had an over 50% of its sales contributed by mergers and acquisitions and drop any year during which the firm had an M&A activity. I exclude firms in financial and utilities sectors (SIC's 4900–4999 and 6900–6999). My sample consists of all the remaining firms that have valid data on sales, employment and capital. For the analyses with R&D, I further select the sample with positive R&D expenditure. Capital stock is constructed using inventory methods based on gross and net property, plant and equipment (PPEGT and PPENT). I estimate firm-level TFP following the Olley-Pakes method using capital as a control variable. In the first step, log output is regressed on third-order polynomials in the logs of capital stock and investment including interactions, as well as log employment. A full set of year and industry dummies are also included. Then I construct the sample including the residuals and lags. In the third step, I re-estimate the coefficient of capital by minimizing the sum of residuals from the regression using the residuals and lags and their third-order polynomials from the second step. Lastly, I compute log TFP assuming a Cobb-Douglas production function, and normalize it to have a mean of zero in each year and industry whenever with-industry dispersion is used.

B Dispersion formula for the fixed effect model

First note that the fixed effect model can be written as

$$z_{i,t+1} = \rho z_{i,t} + c_i + v_{i,t+1}.$$

The stationary variance is

$$\begin{aligned} V(z_{i,t+1}) =& V(\rho z_{i,t} + c_i + v_{i,t+1}) \\ =& V(\rho z_{i,t}) + \sigma_c^2 + 2 \times \text{cov}(\rho z_{i,t}, c_i) + \sigma_v^2 \end{aligned}$$

where

$$\begin{aligned}\mathrm{cov}(z_{i,t}, c_i) =& \mathrm{cov}(\rho z_{i,t-1} + c_i + v_{i,t}, c_i) \\ =& \mathrm{cov}(\rho z_{i,t-1}, c_i) + \sigma_c^2\end{aligned}$$

The variance formula becomes

$$\begin{aligned}V(z_{i,t+1}) =& \rho^2 V(z_{i,t}) + \sigma_c^2(1 + 2\rho + 2\rho^2 + ...) + \sigma_v^2 \\ =& \rho^2 V(z_{i,t}) + \frac{1+\rho}{1-\rho}\sigma_c^2 + \sigma_v^2\end{aligned}$$

which then implies the stationary variance is

$$\sigma_{z,ss}^2 = \frac{\sigma_c^2}{(1-\rho)^2} + \frac{\sigma_v^2}{1-\rho^2}$$

C Fixed effect regression with decade dummies

The regression is

$$\Delta \log z_{i,t+1} = \gamma_0 \log z_{i,t} + \gamma_1(\log z_{i,t} \times D_{1990}) + \gamma_2(\log z_{i,t} \times D_{2000}) + \alpha_i + \alpha_{s,t} + \theta X_{i,t} + v_{it}$$

Table 8. Fixed effect regression with decade dummies

	(1)	(2)
$\log z$	-0.573^{***}	-0.576^{***}
	(0.019)	(0.018)
$\log z \times D_{1990}$	0.065^{***}	0.065^{***}
	(0.020)	(0.020)
$\log z \times D_{2000}$	0.045^{*}	0.045^{*}
	(0.024)	(0.024)
Controls	No	Yes
R^2	0.402	0.412

The setup of this regression is similar to regression (3) except that here the decade dummies are included to impose constant fixed effects across decades. Therefore this is more comparable to the results from regression (3). The results are shown in Table

8. The stationary standard deviation of the log TFP is given by

$$\sigma_z = \sqrt{\frac{\sigma_c^2}{\beta^2} + \frac{\sigma_v^2}{1-(1+\beta)^2}}$$

where $\sigma_v = 0.50$, $\sigma_c = 2.46$ in the 1980s. The implied effects on the dispersion of γ_1 and γ_2 are 12.3% and 8.4% respectively but with a statistically non-significant difference between them. The effect of skill premium growth is even larger with this formulation, despite a decline of contribution of the convergence channel to the overall dispersion growth.

D Average effect of skill premium growth

The following regression estimates the average effect (η_1) by replacing the year dummies with polynomials of years. η_1 can be estimated with or without the interaction term. The effects have a negative sign but are mostly non-significant, indicating that the average effect under this specification is inconclusive.

$$\log z_{i,t+h} - \log z_{i,t} = \alpha_i^h + \alpha_s^h + \beta^h poly(t,5) + \eta_1^h w_{t+1} + \eta_2^h(\log z_{i,t} \times w_{t+1}) + \theta^h X_{i,t} + v_{it}^h$$

Table 9. Average effects with year polynomials

$A : \Delta_h \log z$	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$
w	-0.158	-0.372^{***}	-0.051	-0.002	-0.154
	(0.101)	(0.133)	(0.149)	(0.173)	(0.184)
R^2	0.416	0.560	0.635	0.678	0.712
$B : \Delta_h \log z$	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$
w	-2.777	-8.869^{**}	-9.666^{**}	-10.074^{*}	-5.220
	(2.968)	(4.147)	(4.593)	(5.316)	(3.988)
$\log z \times w$	0.232	0.753^{**}	0.853^{**}	0.896^{*}	0.451
	(0.258)	(0.361)	(0.401)	(0.466)	(0.348)
R^2	0.416	0.560	0.636	0.678	0.712

E Including post-2010

Including the sample after 2010 does not change the main results very much. In fact, from Table 10, one can conclude that rate of convergence reverts in the 2010s. This indeed aligns with the empirical observation in Cunningham et al. (2021) based on the US census data, that the productivity dispersion declines since 2010.

Table 10. TFP convergence rates in four decades

$\Delta \log z_{t+1}$	1980s	1990s	2000s	2010s
A: Pooled OLS				
$\log z_t$	-0.245	-0.209	-0.184	-0.186
	(0.005)	(0.005)	(0.006)	(0.007)
B: Fixed Effects & Controls				
$\log z_t$	-0.706	-0.639	-0.626	-0.666
	(0.025)	(0.023)	(0.025)	(0.038)
Observations	11,837	12,865	10,045	6,405

Note: all the estimated coefficients are significant at 0.01 level. For the regression in panel B, error is clustered at firm level.

Moreover, because skill premium has a much weaker growth after 2010, the relationship between convergence rate and skill premium described in Section 2.2 still holds, as shown in Table 11, which includes the observations after 2010.

Table 11. Heterogeneous responses to skill premium growth

$\Delta_h \log z$	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$
$\log z \times w$	0.289	0.708**	0.646*	0.854*	0.478
	(0.266)	(0.348)	(0.390)	(0.478)	(0.344)
Obs.	40,427	31,999	25,429	20,454	16,644
Controls & Fixed effect	✓	✓	✓	✓	✓
R^2	0.436	0.574	0.652	0.698	0.734

Note: *p<0.1; **p<0.05; ***p<0.01. All standard errors are clustered at firm level.