Productivity Convergence and Firm's Training Strategy

Maastricht University

Mantej Pardesi

Productivity Convergence and Firm's Training Strategy

RM/24/003

ISSN: 2666-8807

GSBE

Maastricht University School of Business and Economics
Graduate School of Business and Economics

P.O Box 616
NL- 6200 MD Maastricht
The Netherlands

Productivity Convergence and Firm's Training Strategy

Mantej Pardesi*

Abstract

In this paper, I study how converging to the productivity frontier influences a firm's training investments. Although productivity growth induces a high-skill bias in firm's workforce structure, little is known about its training incentives for vocational and technical skills. I address endogeneity in productivity growth using a two-stage control function approach where I use productivity shocks as exogenous changes to a firm's position in intra-industry distribution. I find that closing the gap to the frontier leads to a negative effect on firm's investment in training in vocational skills. The negative effect is stronger for large, multi-plant, innovative and technical advanced firms. Using a model for firm sponsored training, I explain the results via a technology effect, cost of training effect and labour substitution effect. First, productivity convergence induces technology upgradation that is skill biased against vocational skills. Second, high expected costs of training augments this skill-biasedness. Third, compositional shift in workforce induces firms to demand fewer vocational and technical skills.

Keywords: Firm Productivity, Frontier, Skill formation, Training, Technological Change, Labour Polarisation

JEL Codes: J240, J420, D24, D23

* Research Center for Education and Labor Market (ROA), Maastricht University, Netherlands

Federal Institute for Vocational Education and Training (BIBB), Bonn, Germany

Corresponding author – mantej.pardesi@maastrichtuniversity.nl

† I am thankful for receiving guidance from Frank Cörvers and Harald Pfeifer for this paper. I am also grateful for comments from Andries de Grip, Thomas Zwick, Ken Mayhew, Caroline Neuber-Pohl, Pelin Özgül, Nicholas Rounding, Nico Pestel, Salman Patel and Eduard Storm. Lastly, I express my gratitude to the feedback and comments from seminar and conference participants at the GINI Conference (University of Agder), Skills for the Future Conference (LISER – Luxembourg), BIBB - Bonn, IAB – Nuremberg, Morse Conference (Maastricht University) and the Swiss Leading House VPET-ECON.

1. Introduction

Studies on heterogeneous firms document persistent and widening gaps in firm productivity, even within narrowly defined industries (Syverson, 2011). These "productivity gaps" account for inequalities in the labour market as highly productive firms offer higher wages, attract high ability workers and invest in labour saving technologies[2]. In the productivity gap literature, firms closer to the "frontier" use highly educated workers much more efficiently than firms further way from the frontier due to capital-skill complementarity (Vandenbussche et al., 2006; Aghion et al., 2009). While the complementarity between advanced-education and proximity to the frontier has been empirically assessed (Bartelsman et al., 2015; Vandenberghe, 2018), little is known about firm's training investments across the productivity distribution. Training is important as it increases workforce productivity and firm's competitiveness (Mohrenweiser & Zwick, 2009; Konings & Vanormelingen, 2015). Moreover, training can be provided to workers across the skill spectrum thereby expanding the analysis to workers without tertiary education. As far as I know, the current literature on firm sponsored training focuses solely on the effect of training on firm performance/productivity. In contrast, this paper studies the dynamic effect of change in a firm's position in the productivity distribution on its training investment.

A firm moves up the productivity distribution if its productivity growth is higher than the productivity growth of the frontier. Theoretically, this process of convergence is due to knowledge spill-overs from the frontier to the extent that knowledge is non-rival and not fully appropriable (Acemoglu et al., 2006). However, due to constraints on firms' absorptive capacity, the learning process is restricted to a certain threshold. To achieve growth beyond this threshold, firms need to attract workers to perform complex and innovative tasks that have an efficiency enhancing effect. Whereas literature in personnel economics documents the heterogeneity in hiring practices across the productivity distribution (see Oyer and Schaefer (2010, 2011) for a systematic review), heterogeneity in training strategies are limited to studies linking training to firm size, monopsony power and labour market institutions (Bassanini et al., 2005; Dustmann & Schönberg, 2009; Manning, 2011). In their seminal contribution to the training literature, Acemoglu & Pischke (1999a; 1999b) argue that firms invest in training even if the knowledge endowed is non-rivalrous and transferable. This is due to market imperfections which allow firms to extract higher rent from trained workers. Concurrently, the "superstar" firm hypothesis argues that firms closer to the frontier have a lower labour share of value added (compressed wage structure) than firms away from the frontier (Autor et al., 2020). Therefore, if training is complementary to proximity to the frontier and allows the firm to extract rents from trained workers, converging to the productivity frontier should influence firm's training investments.

In this paper, I assess the effect of converging to the productivity frontier on a firm's training investment. The paper contributes to the literature in various dimensions. Firstly, I use a large establishment-level longitudinal data set that provides rich information to simultaneously estimate establishment specific productivity and training strategies. The establishment level data ensures consistency in analysis and reduces aggregation bias to capture our main effects on

[2] In their cross-country study on wage inequality, Criscuolo et al. (2020; 2021) argue that a third of overall wage inequality is due to differences in pay between firms. Highly productive firms pay a productivity premia to all of their workers, possess greater wage setting power and are able to attract high ability workers. Further empirical studies can be found with Dunne et al., (2004), Faggio et al. (2010), Barth et al., (2016) and Card et al., (2018) for a theoretical study. For technology investments and variations in managerial practices across firms, see Andrews et al. (2015) and Bloom & van Reenen (2010).

training more precisely. Secondly, I focus on apprenticeship training that constitutes skill formation in transferable skills for young trainees at the establishment level. Since apprenticeship training is firm financed, costly, time intensive and reflects the establishment's skill demand expectations, apprentice vacancies reflects an establishment's intensity to invest in training (Wolter & Ryan, 2011; BiBB, 2018, 2022). Thirdly, I control for simultaneity between productivity and training using dynamic panel instruments to extract time variant establishment specific productivity shocks similar to Guiso et al. (2005) and Guertzgen (2014). These productivity shocks identify exogenous and unanticipated changes to a firm's position in the productivity distribution. Fourth, due to the longitudinal structure of the data, I control for time-invariant unobserved confounders at the establishment level in a within-between model (Neuhaus & Kalbfleisch, 1998) and account for heterogeneity in establishment characteristics using split-sample regressions.

I find that a 1 % increase in an establishment's proximity to the frontier leads to a 25% to 40% decrease in the demand for apprentices, depending on the specification. This effect is greater in magnitude for large establishments, positive for medium sized establishments and non-linear across the productivity distribution. I explain these effects using three mechanisms: First, the *convergence effect* is negative as productivity growth is Hicks neutral for small and medium-sized firms and skill-biased for large firms. Second, proximity to the frontier induces establishments to be technologically advanced. In line with the skill-biased technical change literature, new technologies polarize labour demand against workers with vocational qualifications in favour of occupations with manual and/or cognitive tasks (Acemoglu & Autor, 2011). I find that firms reduce their demand for workers with vocational qualifications and increase the demand for low-skilled workers as they converge to the productivity frontier. The *labour substitution effect* depicts that when firms do not expect to hire workers for vocational qualifications, they will also reduce training in such occupations. Third, high net costs of training[3] buttress the perverse incentive for firms to recruit apprentices resulting in a negative *cost of training effect* (Mühlemann, 2016). Furthermore, I show how the results are uninfluenced by firm's ability to fill apprentice vacancies and retain apprentices. Conditional on training being general and firm sponsored, the study holds value for other types of firm sponsored training in general/transferable skills. Therefore, productivity growth might induce firms to either hire or train in firm specific skills or replace workers with capital investment as a strategy to meet its skill demand.

The remainder of the paper is structured as follows: Section 2 provides the institutional and empirical overview of the German apprenticeship system. Section 3 uses the extant literature to build a model of firm sponsored apprenticeship training that incorporates productivity gaps, technology endowments and costs of training. Sections 4 and 5 detail the data, methodology and identification strategies used in this paper. Section 6 and 7 contain the main results and mechanisms driving our results along with several robustness checks and heterogeneity analysis. The study concludes with contributions to the literature and recommendations for future work in Section 8.

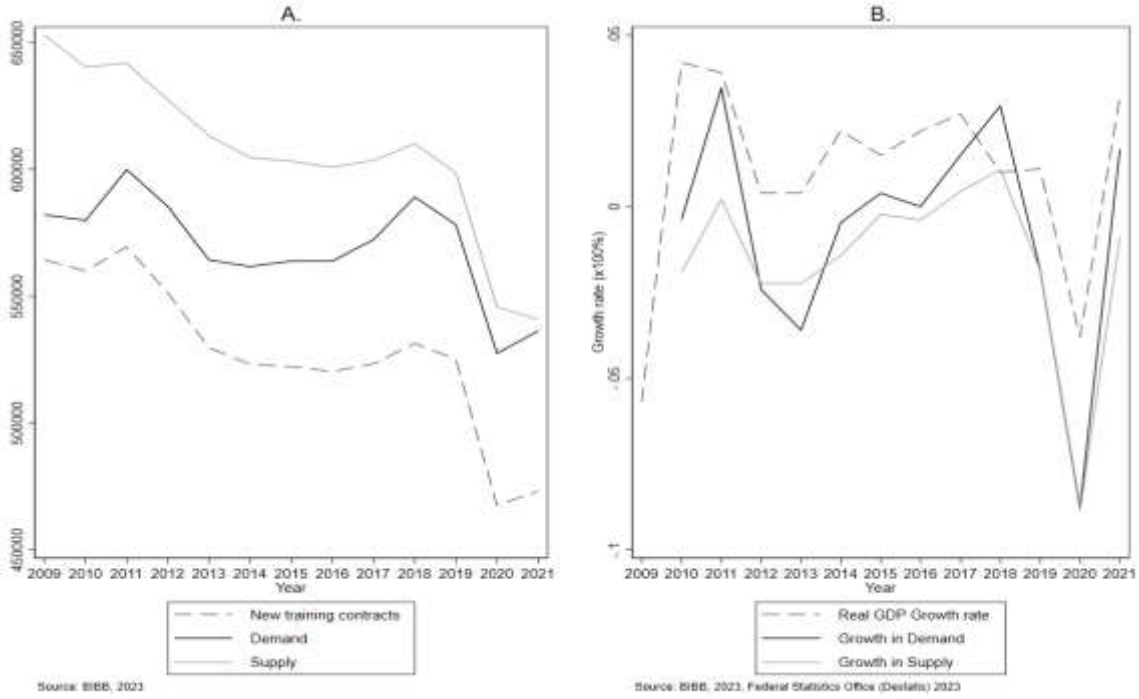
[3] Net costs of training are gross costs of training less the benefits from apprentice's productive work.

2. Apprenticeship Training in Germany

2.1. Institutional Context

In Germany, initial vocational education training (IVET) is institutionalized under i) the dual-system of apprenticeship training, ii) the school-based system; and the, iii) pre-vocational training measures[4]. Every year more than 50% of each secondary school graduating cohort starts an apprenticeship program (Uhly, 2020). Apprentices are trained through a combination of practical job-oriented learning at companies and vocational schools. Training in the dual-system is based on a standardized training curricula for each occupation. Due to the standardized training curricula and a streamlined certification process, skills learned by apprentices in their training company are transferable to other companies (Franz & Soskice, 1995; Clark, 2000; Franz & Zimmermann, 2002). These training curricula are formed via a consensus-based structure[5] providing the incentives for firms to train apprentices in transferable skills within nationally defined occupations (Solga et al., 2014).

Figure 1. Evolution of demand and supply of apprentices in Germany



Source: BIBB, 2023 & Federal Statistics Office (*Destatis*) 2023

In the past decade, the number of new training contracts has gone down from more than 600,000 in 2008 to about 470,000 in 2021 (see panel A in Figure 1). The decline in newly concluded training contracts is due to: 1) demographic decline,

[4] This is referred to as the 'transitional' system that is mostly school-based and focused on preparing individuals for pursuing occupation credentials in either the dual system or the fully qualifying school-based system.

[5] Consensus-based structure implies decisions regarding the changes to regulation of apprenticeship occupations are made with the consent of the relevant stakeholders, i.e., employers (employer associations and chambers), trade unions, the 16 German Länder states and the federal government.

2) unattractiveness of vocational training, and 3) declining demand by firms. Due to demographic decline, fewer students graduate from school decreasing absolute number of eligible apprentices. Moreover, fewer school leavers show interest in apprenticeship training, further decreasing the pool of potential of apprentices[6] (Haasler, 2020). In panel B of Figure 1, change in apprenticeship demand and real GDP growth are congruent with each other. This provides evidence for the pro-cyclicality of demand for apprentices. The sharp drop in demand, supply and GDP growth witnessed in 2020 is due to the COVID-19 pandemic. The scarring effect of COVID-19 is seen as the number of new training contracts in 2021 is considerably less than in 2019 (panel A). Nevertheless, the cyclicality in demand and persistent excess supply is consistent with the literature on youth labour markets (Clark & Summers, 1980; Card, Kluve, et al., 2018). Therefore, this study has external validity in the broader literature of youth labour markets.

2.2. Why do firms train apprentices?

Firms train apprentices for a multitude of reasons: 1) as a screening device for future recruitment (Ryan, 2010); 2) as low cost substitutes for skilled workers (Mohrenweiser & Backes-Gellner, 2010); and 3) as a strategy to meet their optimal demand for present and future skills (Pfeifer & Backes-Gellner, 2018). The literature on firm's demand for apprentices broadly falls into two groups: 1) the feasibility of training and 2) the attractiveness of training. In Germany, employers bear the most prominent costs in terms of training, such as apprentice wages, costs of training staff and equipment [7] [8]. The feasibility of firms to train depends upon their ability to bear the burden of costs, i.e., smaller firms train less as it is relatively more expensive for them to train than for larger firms (Jansen et al., 2015). Simultaneously, a convex shaped hiring cost function makes training more appealing for firms as hiring skilled workers becomes increasingly difficult (Blatter et al., 2016; Muehlemann & Pfeifer, 2016). Existing literature suggests that the ability and willingness to bear the net-costs of training are influenced by factors such as firm size, employee representation, training motive, business cycle (and business expectations), labour market rigidities and the degree of labour market competition [9].

Several studies have documented the rising costs of training due to changing skill needs, technological progress and tightness in the labour market (Muehlemann & Pfeifer, 2016; Mohrenweiser et al., 2019). As training responds to wider

[6] In the BIBB (2022), the institutionally registered supply of apprentices is defined as the sum of newly concluded training contracts, registered applicants for vocational training positions minus applicants who entered vocational training.

[7] The need to be competitive induces higher investment expenditure in new and advanced equipment and machinery. However, more training reduces the time spent on using the equipment for productive purposes.

[8] According to Schönfeld et al. (2010), wages paid to apprentices comprise the largest component of training expenditure, reaching up to 50% of the total expenditure.

[9] Evidence on how larger firms train and demand more apprentices is in the VET reports published by the BiBB (2022). Koch et al. (2019) provide evidence of how the presence of works councils influences demand and quality of apprenticeship training in Germany. Jansen et al., (2015) show how apprenticeship training in Germany is motivated by investment into future skills. Muehlemann et al., (2009) and Dietrich & Gerner (2007) analyse the role of business cycle and business expectations towards the demand for apprentices. Theoretically, the work by Dietrich & Gerner (2007) is very close to our work. Lastly, Baethge et al., (2007) and Stockinger & Zwick (2017) provide evidence of how competitive pressure increases the risk of poaching and thus reduces the demand for apprentices. In addition, proximity to the frontier might trigger reputation effects of apprenticeship training, where firms might demand more apprentices as their improved reputation deems them to post apprenticeship vacancies (Wolter & Ryan, 2011).

technological changes, implementing new technologies increase the costs of training. On one hand, observed and opportunity costs due to increasing technologies create a perverse incentives for firms to participate in the training market. This is because time spent on using such technologies for production competes with time spent on training (Thelen, 2007). On the other hand, participation in the training market allows firms to catch up with the latest technologies and production processes in the industry furthering technology diffusion (Rupietta & Backes-Gellner, 2019; Schultheiss & Backes-Gellner, 2020). As productivity is a crude measure of technological intensity and firms vary in their technology endowment, one can expect highly productive firms to spend less on training if their occupations are at a high risk of automation and hiring costs to replace skilled workers are low. In their recent work, Heß et al. (2023) show how not only the firm but the workers in occupations that are more exposed to automation technologies are less likely to participate and receive firm sponsored training.

Beyond these firm-specific reasons, firm's demand for apprentices is pro-cyclical to business cycles and business expectations similar to other types of firm investments (Bellmann, Gerner, et al., 2014; Lüthi & Wolter, 2020) . The intuition is that a firm would be more likely to bear the costs of training and have higher returns from training in a boom rather than a slowdown (Dietrich & Gerner, 2007; Muehlemann et al., 2009). Therefore, closing the gap to the productivity frontier should induce positive business expectations and increase firm's demand for apprenticeship training. This expectation is conditional on the role of apprenticeship training in a firm's productive activity. Existing literature only investigates the impact of apprenticeship training on firm performance. These studies show that the effect of apprenticeship training on firm productivity presents a positive effect in studies on Swiss firms, a negative effect for Hungarian and Austrian firms and mixed effects for German firms (Böheim et al., 2009; Dionisius et al., 2009; Mohrenweiser & Zwick, 2009; Cabus & Nagy, 2021).

The returns of training apprentices further influence path-dependency in a firm's training motive, i.e., whether firms train apprentices to "invest" for future demand for skills or whether firms train apprentices to "substitute" present skilled labour requirements (Lindley, 1975; Mohrenweiser & Zwick, 2009). Mohrenweiser and Backes-Gellner (2010) show how firms with substitution motive have lower retention rates. The increase in the pool of trained apprentices as skilled workers in the market contributes to labour market tightness. In another analysis of the training motive, Mohrenweiser and Backes-Gellner (2008) found no effect of the share of apprentices in a firm on its profits hinting at a possible investment rationale for training apprentices. This is because if apprentices do not contribute to the contemporaneous profits of a firm, then a firm trains them for their contribution to profits in the future. To what extent productivity convergence influences a firm's rationale to train apprentices is, yet, an unexplored topic largely due to little variation in a firm's training motive[10].

At the macro-level, falling labour costs in the global south might induce firms to further shift away production processes to other countries ('off-shoring'), reducing labour demand for occupations and tasks that the countries in global south might have a comparative advantage in (Goos et al., 2014). The resultant decline in demand for apprentices is consequential for firms as it increases the shortage of skilled workers, increasing the need for expensive on-the-job training and decreases its innovative capacity (Forster & Bol, 2018). Overall, positive expectations, pro cyclicality, and

[10] Using year-to-year retention rates as a metric for training rationale limits the sample to large firms with graduating apprentices every year. Therefore, a longer horizon such as a three-year average retention rate is a better measure for training rationale.

monopsony power positively affect the demand for apprentices whereas rising costs of training, technological change and competition negatively influence the demand for apprentices. In this context, this study links the prior literature by means of potential confounders and adds another explanation for determinants of apprenticeship demand.

2.3. Equilibrium in the Apprenticeship Training Market

The decision to demand apprentices depends upon the role of apprentices in the production processes of a firm that is converging to the productivity frontier. Let us consider an economy comprising of $i = 1, \dots, N$ firms each using a mix of skilled, unskilled workers and apprentices, $S_{i,t}, U_{i,t}$ and $T_{i,t}$ respectively, to produce a given unit of output at time t. For simplicity, let us assume that the production function for each firm i in industry j is $y_{i,j,t} = A_{i,t} f(S_{i,t}, U_{i,t}, T_{i,t}, K_{i,t})$ where $A_{i,t}$ is a Hicks-neutral technical efficiency parameter and $K_{i,t}$ is firm's capital stock. For wages $w_{i,t}^S, w_{i,t}^U$ and $w_{i,t}^T$ of workers and apprentices respectively, $r_{i,t}$ as rental rate of capital, profit maximisation is given as $\pi_{i,t} = p y_{i,j,t} - w_{i,t}^S S_{i,t} - w_{i,t}^U U_{i,t} - c_{i,t} T_{i,t} - r_{i,t} K_{i,t}$ where $c_{i,t}$ is the cost of training per apprentice per year. In a competitive market, equilibrium for apprenticeship demand is attained when marginal product equals the marginal cost of training. I can write the value of marginal product as $VMP_{i,t} = p * MP_T = p * \frac{dy_{i,j,t}}{dT_{i,t}}$ and the equilibrium condition in equation 1 as:

$$p A_{i,t} \frac{df(S_{i,t}, U_{i,t}, T_{i,t}, K_{i,t})}{dT_{i,t}} = c(w_T, h^T(\theta), c^o) \qquad \forall\, i = 1, \dots, F \tag{1}$$

In equation 1, the marginal cost of training an additional apprentice is a function of the wage of an apprentice (w_T), hours spent on training (h^T), quality of training (θ), and other costs (c^o) such as training material, machinery and training personnel. Mühlemann (2016) denotes the marginal product (benefit) of training an apprentice in year t as a function of time spent by the apprentice doing skilled work ($h^W \alpha_T$) and the relative productivity of the apprentice to skilled ($\gamma_\alpha w_{i,t}^S$) and unskilled worker ($\beta_T w_{i,t}^U$). I can denote this in equation 2 as:

$$MP_{i,t} = A_{i,t} \frac{df(S_{i,t}, U_{i,t}, T_{i,t}, K_{i,t})}{dT_{i,t}} = B_t^T = h^W(\alpha_T * \gamma_\alpha w_{i,t}^S + \beta_T * w_{i,t}^U) \tag{2}$$

Where, $0 < \gamma_\alpha < 1$ as apprentices have lower productivity than a skilled worker. α_T and β_T represent the share of skilled and unskilled tasks done by the apprentice during the time spent in training. Thus, the marginal benefit for a firm is positively related on how productively the firm uses apprentices in its production process. If a firm trains T apprentices and every apprentice trains for a period of s years, the total benefit from apprenticeship training can be derived in equation 3:

$$TB_T = T^\rho \sum_{t=1}^{s} h^W(\alpha_T * \gamma_\alpha w_{i,t}^S + \beta_T * w_{i,t}^U) \tag{3}$$

Where $\rho < 1$ implies a concave benefit structure (decreasing marginal training benefits). The law of diminishing returns to scale implies that a large firm with a high number of apprentices will get a lower benefit from the additional apprentice it trains than a small firm. Analogously, total training costs for T apprentices for s years can be expressed in equation 4 as:

$$TC_T = T^\sigma \sum_{t=1}^{s} c(w_T, h^T(\theta), c^o) = T^\sigma \sum_{t=1}^{s} w_T + h^T(\theta) w_{i,t}^S + c^o \tag{4}$$

Where, $\sigma > 1$ implies a convex cost structure. The assumption regarding convex cost structure is backed by literature on firm size where larger firms have a higher cost of recruitment and training (Muehlemann & Pfeifer, 2016). The equilibrium condition can be achieved when marginal training benefits equal marginal training costs, i.e.,

$$\frac{dTB}{dT} = \frac{dTC}{dT}$$
$$\rho T^{\rho-1} B_t^T = \sigma T^{\sigma-1} C_t^T \quad \text{-(5)-}$$

Where, $C_t^T = c(w_T, h^T(\theta), c^o)$ is the marginal cost of training an additional apprentice per year. Simplifying equation 5 to get the optimal demand for apprentices of firm i:

$$T_{it} = \left(\frac{\sigma B_t^T}{\rho C_t^T}\right)^{\frac{1}{\sigma-\rho}} \quad \text{-(6)-}$$

Therefore, a firm's steady state demand for apprentices depends positively on the marginal benefit-cost ratio of providing apprenticeship training. In addition the demand is negatively related to the degree of concavity of the benefit function (ρ) and positively on the degree of convexity of the cost function (σ).

Although $\rho < 1$ is theoretically motivated, $\sigma > 1$ cannot be unambiguously theorised. Cost savings can occur as firms can train many apprentices simultaneously with a single trainer. Conversely, capacity constraints might push costs further if firms want to train more apprentices as they would have to invest in training facilities and trainers. Therefore, one relies on empirical evidence where cost of recruiting apprentices are seen to be higher for larger firms than for small firms (Muehlemann & Pfeifer, 2016). The assumption of $\sigma > \rho$ concerns the shapes of the marginal benefit and marginal cost curves. In equation 6, a high ρ indicates that a large training firms will have a much greater decline in marginal benefits from an additional apprentice than small and medium firms. Since the benefits and costs incurred from additional apprentice are unique for each firm, equation 6 provides a training equilibrium for the i^{th} firm.

Using factor price equalization, market equilibrium is achieved in the apprenticeship market when the ratio of marginal benefit and marginal cost in equation 5 is equal for all firms in the industry. Let us assume two types of firms in the industry, a frontier (F) and a non-frontier firm (i). In this industry, frontier firms are the productivity leaders and they are characterised as having superior technology than non-frontier firms, i.e., $A_i < A_F$. I assume that the degree of concavity in benefits and degree of convexity in costs structure to be the same throughout the industry. This is because the apprenticeship system in Germany is institutionalised where costs are floored using minimum training allowances and apprentice contribution is assessed using standardized certification process. This means that $\sigma_i = \sigma_F = \sigma$ and $\rho_i = \rho_F = \rho$. Factor price equalization in the apprenticeship market allows us to obtain the following relation between a non-frontier firm and a frontier firm:

$$\frac{\rho T_i^{\rho-1} B_{i,t}^T}{\sigma T_i^{\sigma-1} C_{i,t}^T} = \frac{\rho T_F^{\rho-1} B_{F,t}^T}{\sigma T_F^{\sigma-1} C_{F,t}^T}$$
$$T_i^{\rho-\sigma} A_{i,t} \frac{MB_{i,t}^T}{C_{i,t}^T} = T_F^{\rho-\sigma} A_{F,t} \frac{MB_{F,t}^T}{C_{F,t}^T}$$
$$\frac{T_{i,t}}{T_{F,t}} = \left(\frac{A_{F,t}}{A_{i,t}} \frac{MB_{F,t}^T}{MB_{i,t}^T} \frac{C_{i,t}^T}{C_{F,t}^T}\right)^{\frac{1}{\rho-\sigma}} \quad \text{-(7)-}$$

Where the relative ratio of apprentices between the non-frontier and frontier firm depends upon the relative intensities of technical efficiency, relative ratio of marginal benefits and marginal costs of apprenticeship training. Taking natural logarithm of equation 7 and removing the time scripts, I get the following estimable relation:

$$\log T_i = \frac{1}{\sigma - \rho}\log\frac{A_i}{A_F} - \frac{1}{\sigma - \rho}\log\frac{C_i}{B_i} - \frac{1}{\sigma - \rho}\log\frac{C_F}{B_F} + \log T_F \qquad \text{-(8)-}$$

Equation 8 provides us with a set of propositions that can be empirically tested:

Proposition 1 (Convergence Effect): *For a Hicks neutral technical efficiency parameter, closing the gap to the frontier (increase in* $\frac{A_i}{A_F}$*) is positively related to demand for apprentices.* As total factor productivity is a proxy for technology intensity of a firm, more efficient firms should use more apprentices as long as $\sigma > \rho$ and A_i is Hicks neutral. Productivity convergence that induces an increase in demand for all the factors of production would also increase the demand for apprentices. Heterogeneity in the effect of productivity convergence would highlight how different firms face different marginal cost and marginal benefit curves.

Proposition 2 (Cost of training effect): *Net Costs of training induces a negative effect if firms expect costs of training larger than expected benefits of training* (C_i/B_i).

On one hand, the allocative function of a firm that distributes apprentices to more skilled tasks (α_T) rather than unskilled tasks (β_T) increases the demand for apprentices. This is regarded as the reason why apprentices contribute to firm productivity in Switzerland relative to Germany (Dionisius et al., 2009). On the other hand, demand for apprentices would decrease as firms expect to incur a higher marginal cost of training for the additional apprentice. Conditional on firm characteristics, the net cost (costs-benefits) of training might be positively or negatively related to demand for apprentices depending upon the how firms are able to use apprentices in their production processes. Firms with negative net costs, i.e., benefits from training exceed costs of training, would increase the demand for apprentices as they are able to extract greater productivity from every additional apprentice.

Proposition 3 (Labour Substitution effect): *Skill biased productivity growth induces substitution of apprentices for high and low skilled workers.*

Factor augmenting productivity growth that is biased towards high skilled workers might reduce the role of apprentices in the production process. Moreover, productivity growth that increases the demand for low-skilled workers at the cost of medium skilled workers contributes to polarization in the labour market. This is backed by the literature since the assumption of Hicks neutrality confounds with the evidence obtained in the skill-biased technological change literature (Abowd et al., 2007). Since apprentices are typically trained in technical and semi-skilled occupations that are substituted for high and low skilled workers, a decrease in the demand for apprentices with a decrease in the demand for semi-skilled workers would create a labour substitution effect.

3. Data

For this study, I use the IAB Establishment Panel (*Betriebspanel*) from the Federal Employment Agency in Germany. The establishment panel is an annual survey covering around 16,000 German establishments with at least one employee

subject to social security contributions, spanning all industries and sizes (Bellmann et al., 2022)[11]. The establishment data comes from the universe of establishments that comply with mandated social security notification of their employees. The establishment panel sample is stratified by size, industry, and federal state (Länder)[12]. The survey provides information on key factors of production, turnover, firm characteristics, and personnel changes, particularly apprenticeships. I limit my analysis to the period from 2009 to 2019 to ensure consistent industry classification in the establishment panel and mitigate the demand shock to the stock of apprentices seen during the 2008-09 financial crisis and the COVID-19 pandemic in 2020. I exclude non-profit and public administration establishments, as these firms might not be considered profit oriented establishments. I further exclude establishments that do not report turnover in sales, such as banks/financial institutions, insurance companies, non-industrial organizations, regional and local authorities.

After these exclusions, I have a sample size of more than 150,000 establishment-year observations. After accounting for missing information on turnover, intermediate costs, investments, wages, etc. I achieve an unbalanced estimation panel of approximately 19,000 establishments. I derive capital stock per establishment using the modified perpetual inventory approach tailored specifically to the establishment panel (Müller, 2008, 2017). I use the methodology by Stiel and Schiersch (2022) to deflate turnover, investments, capital stock with industry-level deflators with 2015 as the base year (OECD, 2017). Wages and other cost related data are deflated with consumer price indices with 2015 as the base year.

I complement the establishment panel with information on school leavers from the Federal Statistics Office to serve as a proxy for the supply of apprentices at the state level (Muehlemann et al., 2022). I use the total number of school leavers (graduates and dropouts) per year at the state level by their school-leaving certificate[13]. The literature on apprenticeship training documents how firm-sponsored training crucially depends upon the costs of training apprentices and the alternate costs of using skilled workers[14]. To account for the costs of training apprentices, I use the BIBB Cost-Benefit Survey waves of 2012-13 and 2017-18 (Jansen et al., 2017; Pfeifer et al., 2021). Specifically, I aggregate training costs per apprentice (both net and gross costs) into 112 cells of sector (14 groups), size (4 groups), west and east German establishments (2 groups). The intuition is that the aggregated costs and benefits reflect the *expected costs/benefits per additional apprentice* for a firm with aforementioned characteristics. I merge these aggregates to the IAB Establishment Panel ensuring group similarities and deflate the costs using the consumer price index. To fill the gaps between the two survey waves and project backwards to 2009, I use linear imputation techniques at the establishment level.

In Table 1, I show the descriptive statistics of key variables in the estimation sample by firm's status of productivity convergence. Specifically, I compare firms that improve their position in the productivity distribution 'convergers' to firms that do not improve their position, viz. 'non-convergers'. In terms of turnover, capital stock and workforce size, I

[11] The establishment survey does not include self-employed and establishments that employ only people not covered by social security (mineworkers, farmers, artists, journalists, etc.) as well as public employers with solely civil servants.

[12] See Fischer et al. (2009) and Ellguth et al. (2014) for a detailed overview of the IAB establishment panel.

[13] A substantial literature on the German apprenticeship system argues that the decline in new training contracts is due to demographic trends and tertiarisation of education (Solga et al., 2014; Thelen, 2007). I use the motivation in Dummert et al. (2019) and Muehlemann et al. (2022) to use school leaver aggregates as proxies for these supply side trends.

[14] See Dionisius et al. (2009), Jansen et al. (2015), Pfeifer et al. (2011) and Muehlemann et al. (2005) for literature on firm-level costs of training apprentices. See Walden & Troltsch (2011) and Bellmann et al. (2014) for literature on wage costs of skilled workers as a determinant for demand for apprentices.

do not observe significant difference between a converger and a non-converger. However, non-convergers have a higher share of intermediate inputs and investment in their turnover. Large expenditure on raw material as a share of turnover indicates inefficient use of resources and supply chains and thus lower productivity. The difference in investment intensity might reflect the delayed effect of investments on an establishment's productivity.

In terms of training characteristics, I find limited structural difference between convergers and non-convergers. Approximately, 58% of our estimation sample are establishments that train at least one apprentice in the corresponding year, defined as a "training" establishment. This proportion is equal between convergers and non-convergers indicating that the extensive margin decision to train or not is not influenced by proximity to the frontier. This is important since apprenticeship training and productivity growth are positively related to firm size. One might expect convergers to be dominated by large firms who are more likely to train apprentices (Van Biesebroeck, 2005; Jansen et al., 2015). However, we find limited evidence for this productivity-size nexus in our sample. Similarly, I observe 5% of the workforce to be comprised of apprentices among both convergers and non-convergers. Although convergers appear to post more vacancies for apprentices than non-convergers, this difference is not statistically significant.

Table 1. Descriptive statistics, by firm's status of productivity convergence

Variable	Estimation Sample (n=67951)		Convergers (n=26401)		Non-convergers (n=27215)		Difference	
	Mean	SD	Mean	SD	Mean	SD	T-Statistic	
Turnover (in million €)	18.3	224	21	295	17.2	181	-1.81	
Capital Stock (in million €)	21.2	319	24.9	426	19.2	238	-1.93	
Intermediate Inputs Intensity	0.48	0.22	0.45	0.21	0.51	0.22	31.17	***
Investment Intensity	0.06	0.20	0.05	0.20	0.06	0.16	4.13	***
Workforce Size	137	1066	148	1329	132	954	-1.68	
Composition: Low skilled	0.15	0.21	0.15	0.22	0.14	0.21	-7.26	***
Composition: Medium skilled	0.63	0.23	0.63	0.23	0.64	0.23	4.82	***
Composition: High skilled	0.07	0.14	0.07	0.14	0.08	0.15	4.07	***
Workforce Turnover Rate	0.01	0.09	0.01	0.09	0.00	0.09	-7.71	***
Works Council (%)	0.25	0.43	0.25	0.43	0.25	0.43	0.27	
Collective agreements (%)	0.40	0.49	0.40	0.49	0.40	0.49	-0.59	
Training firms (%)	0.58	0.49	0.57	0.50	0.58	0.49	1.64	
Apprenticeship Rate	0.05	0.08	0.05	0.08	0.05	0.08	-0.08	
Number of Vacancies Posted	2.84	16.47	2.97	19.46	2.69	14.35	-1.89	
Unfilled Vacancy Rate†	0.19	0.34	0.20	0.35	0.22	0.36	3.52	***

† Variable with fewer observations than the estimation sample
Note: Convergers are establishments that improved their position in the productivity distribution. Training establishments are those that train at least one apprentice in year t. Intermediate Inputs Intensity = Intermediate Inputs/Turnover. Investment Intensity = Investments/Turnover. Workforce Turnover Rate = (New Hires - Separations)/Workforce Size. Apprenticeship rate is defined as the share of apprentices in the total establishment workforce. Unfilled vacancy rate is the share of apprentice vacancies that were not filled by apprentices in year t. * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Furthermore, the difference in workforce composition is important for the study results. The estimation sample is consistent with the widely documented properties of the German labour market: a large share of medium-skilled workers and lower shares of low skilled and high skilled workers (Baethge & Wolter, 2015; Rohrbach-Schmidt & Tiemann, 2016)[15]. Whereas non-convergers have a greater composition of medium and high skilled workers, convergers have a greater share of low-skilled workers. Moreover, establishments that converge to the productivity frontier have a greater turnover rate than establishment that do not converge to the frontier. Differences in workforce composition are a threat to internal validity and thus need to be included as control variables in our analysis. It has been emphasised in earlier literature that labour market institutions influence not only personnel fluctuations but also a firm's training behaviour (Thelen, 2004; Fregin et al., 2020). I observe that around 25% of establishments have employee representation in the form of works councils and 40% of establishments have wages determined by sectoral collective labour agreements. I do not observe any statistical difference in labour market institutions between convergers and non-convergers.

4. Methodology

The main estimation strategy is based on equation 8 where demand for apprentices[16], T_{it}, is a function of closeness to the productivity frontier $CTF_{it} = \frac{A_{it}}{A_{Ft}}$. Demand for apprentices is defined as the number of apprentice vacancies posted by the establishment in year t. The estimable version of equation 8 is written as:

$$T_{it} = \beta_0 + \beta_1 CTF_{it} + X_{it} + H_i + \delta_t + \epsilon_{it} \qquad \text{-(9)-}$$

In equation 9, $X_{i,t}$ is a vector of time-varying establishment controls such as establishment size, number of skilled workers (in log), turnover rate (new hires minus separations as a share of the workforce), share of part-time workers, share of workers with university degrees, presence of works council and sectoral collective labour agreement. Time invariant controls are a set of sector and region dummies in H_i and δ_t contains time-varying variables such as year dummies and number of high school graduates at the Länder level. The latter variable reflect the supply side of the apprenticeship training market allowing us to effectively control for demographic effects (Muehlemann et al., 2005). Since vacancies for apprentices is a count variable with over-dispersion, I use negative binomial regression proposed by Cameron and Trivedi (1986) (hereafter referred to as the NB2 model)[17]. I restrict the estimation sample to firms that are below the frontier ($CTF_{it} < 1$) to avoid bias emanating from firms that are at or above the frontier. For instance, if the frontier firms

[15] The workforce composition totals do not add to 1 because I can only identify employees subject to social security with their educational levels. This excludes working proprietors, civils servants, interns, freelancers and marginal part-time workers. On average, training firms have a lower share of employees implying that they have a higher share of the latter group of workforce. For more information on managerial and part-time work in Germany, please refer to Brenke (2011).

[16] We define demand for apprenticeship training as the number of apprentice vacancies posted by the firm in year t. Since this variable is a discrete variable we use a negative binomial regression. This is important since many firms display volatility in their apprenticeship demand, i.e., posting 0 vacancies in year t, 2 in year $t+1$, and 0 in year $t+2$. Taking a logarithm of the demand variable (typical in the literature) omits the years where there has been no demand creating sample bias.

[17] Wooldridge (1999) show how a Poisson model works equally well due to its generalizable assumptions. In our robustness checks, we show how our main results our equal regardless of the method of estimation used.

are less likely to train apprentices, our estimates will be downward biased due to structurally different training characteristics in the upper-tail of the productivity distribution.

In order to have reliable estimates for equation 9, I need to ensure, 1) a robust method for estimating firm productivity, 2) derive exogenous changes to the productivity measure, and 3) consistently identify the causal effect in the empirical strategy. As stated earlier, literature on firm productivity and training has focused on the potential benefits of workforce training on firm productivity. In this paper, I control for this reverse causality by identifying productivity shocks to the establishment. The intuition is that unexpected and unanticipated changes in productivity growth influence firm's demand for apprentices through the pass-through of the shock on firm's position in the productivity distribution. In addition to the productivity shock, I control for time-invariant unobserved characteristics in equation 9 using the within-between model first proposed by Mundlak (1978) and formalised by Neuhaus and Kalbfleisch (1998). This allows me to consistently estimate β_1 in equation 9 whilst accounting for potential endogeneity.

4.1. Estimating establishment-level productivity

An establishment's closeness to the productivity frontier is a ratio of its productivity to the productivity of the frontier establishment. Without assuming constant returns to scale, I use a skill-augmented value-added production function approach to estimate establishment-level productivity[18] as:

$$va_{it} = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \beta_s s'_{it} + v_{it} + u_{it} \quad \text{-(10)-}$$
$$such\ that, TFP = \omega_{it} = \beta_0 + v_{it}$$

Where β_l and β_k are output elasticities of labour (l) and capital (k), in log scale respectively. va_{it} is logarithm of gross value added, i.e., turnover minus intermediate costs[19] and s' includes the share of low-skilled workers, medium skilled workers, high skilled workers and apprentices out of the total workforce[20]. I run equation 10 at the sectoral level to derive output elasticities for each sector separately. Therefore, total factor productivity (TFP) is denoted as ω_{it}, β_0 represents the average productivity level at the sector level and v_{it} represents the time-varying establishment-specific part of productivity. Intuitively, ω_{it} represents the part of a firm's value added that is unexplained by observable/embodied factors of production. Thus, TFP is a proxy indicator for disembodied factor augmenting technological change (Van Beveren, 2012; Autor & Salomons, 2017). The error term, u_{it}, is an i.i.d. component representing white noise.

Cross-sectional and ordinary least squares (OLS) estimation of equation 10 renders biased estimates of output elasticities due to unobserved heterogeneity and simultaneity between value added and explanatory variables. Moreover, the quality of factors of production and input choices might be correlated with firm productivity (Olley & Pakes, 1996; Levinsohn & Petrin, 2003) . To alleviate these concerns, I use the semi-parametric estimation proposed by Levinsohn and Petrin (LP) (2003). The basic assumptions are a) input and output prices are uniform across firms in a sector, b) capital stock is

[18] See Black & Lynch (2001), Zwick (2006) and Konings & Vanormelingen (2015) for earlier work using heterogeneous worker based production functions.

[19] In the establishment panel, intermediate costs include costs for raw materials, merchandise, wage work, rents and other internal and external costs. I use value added rather than turnover based on Lochner & Schulz, (2021)

[20] I categorize low skilled workers as employees in menial jobs that do not require specific vocational education. Medium skilled workers are employees in skilled jobs that require a vocational qualification or comparable training on the job or relevant professional experience. High skilled workers are employees in skilled jobs that require a university degree.

predetermined in period $t-1$ as a state variable, and, c) labour and intermediate costs respond to productivity in period $t-1$ (see Appendix 1 for a technical explanation of the LP technique). Productivity is proxied by a non-parametric function of intermediate input demand (the control function) that is strictly increasing in (scalar and unobserved) firm productivity (Konings, 2008). Inverting this function allows us to control for unobserved firm productivity by substituting it out of the production function[21]. Conditional on the probability of survival of the establishment in the next period (χ_{it})[22], productivity is estimated using a first-order Markov process.

$$\omega_{it} = E(\omega_{it}|\omega_{i,t-1}, \chi_{it}) + \xi_{it} = \lambda\omega_{i,t-1} + \xi_{it} \quad \text{-(11)-}$$

Where ξ_{it} is a change to productivity that is uncorrelated with ω_{it} or k_{it} where I assume that ω_{it} follows an AR(1) process with parameter λ. Deriving output elasticities using LP estimation allows us to control for unobserved heterogeneity, however, literature suggests that this method suffers from serial correlation of the estimated coefficients (Ackerberg et al., 2006; Ackerberg et al., 2007; Ackerberg et al., 2015). Therefore, I deploy the Ackerberg-Caves-Frazer (ACF) correction to avoid issues related to collinearity in estimating labour and capital coefficients (Ackerberg et al., 2006). In the ACF correction, the control function is augmented by including labour input as an additional proxy for productivity. This conditioning eliminates the collinearity between intermediate input demand and labour input that depends upon wage costs and potentially training costs (Konings & Vanormelingen, 2015; Lochner & Schulz, 2021).

Productivity convergence the dynamic process of catching-up to the frontier, I use a relative measure of productivity based on the "proximity to the frontier" literature (Acemoglu et al., 2006; Bartelsman et al., 2015; Vandenberghe, 2018). The "proximity to the frontier" approach suggests that heterogeneous firms should select strategies that depend on their relative performance within their industry (Bartelsman et al., 2008). Changes in CTF can either be due to $\Delta\omega_{it}$ or $\Delta\omega_{Ft}$, i.e., a firm converges to the productivity frontier if $\Delta\omega_{it} > \Delta\omega_{Ft}$. Firm-level studies that use productivity as an independent variable typically rely on industry level shocks as a source of exogenous variation in firm level productivity. These can be export shocks, exchange rate shocks, technology shocks among others. This strategy is insufficient since I need exogenous variation that is time variant and firm specific. Moreover, literature discussed above shows how firm level training influences firm's productivity hinting at potential reverse causality (Dearden et al., 2006; Zwick, 2006, 2007; Cabus & Nagy, 2021). Establishments that demand apprentices might be structurally more productive due to factors such as management quality, exposition to long-term technical change, management-employee relations. Therefore, I isolate exogenous changes in CTF through a firm specific component and a sector specific component.

4.2. Instrumental Strategy : Identifying exogenous variation in CTF

Any potential firm-specific instrument should be uncorrelated with both observable and unobservable determinants of estimated productivity, ω_{it}. In the LP estimation, ω_{it} evolves in a Markov process and, thus, can be decomposed into

[21] Levinsohn and Petrin (2003) argue that proxying for ω_{it} allows a more flexible interpretation of productivity. This is different to reducing ω_{it} to a "fixed" (time invariant) firm effect and hence is a less "costly" solution to the omitted variable and/or simultaneity problem.

[22] Low productive firms might drop out of their sample due to their low productivity creating a potential selection bias. If I assume that a firm operates if its productivity is above a minimum sectoral level, Olley and Pakes (1996) propose to include a survival indicator variable $\chi_{it} = 1\ s.t.\ \omega_{it} > \underline{\omega}_{it}$ where $\underline{\omega}_{it}$ is the sectoral minimum.

two factors: a trend element and an idiosyncratic shock ξ_{it}. Writing equation 11 in first differences, adding a further lagged period and time dummies, I get the following:

$$\Delta\omega_{it} = \lambda_1\Delta\omega_{i,t-1} + \lambda_2\Delta\omega_{i,t-2} + \tau_t + \Delta\xi_{it} \quad \text{-(12)-}$$

Where λ_1 and λ_2 estimate the effect of growth in past periods on today's growth, and the year dummies control for business cycle effects. First differencing allows us to achieve within-panel stationarity in estimated productivity (ω_{it}). Equation 12 is characteristically similar to the work by Guiso et al. (2005), Carlsson et al. (2016) and in the German context, by Guertzgen (2014)[23]. Specifically, if I remove the trend component (τ_t) and growth persistence ($\Delta\omega_{i,t-1}$ and $\Delta\omega_{i,t-2}$) from productivity growth in period t, the remaining variation in $\Delta\omega_{i,t}$ is the deviation from trend and persistence that can be considered as a shock to the firm in period t ($\Delta\xi_{it}$). In equation 12, growth in productivity in period t is influenced by growth in productivity in period $t-1$, $t-2$ and an idiosyncratic shock to productivity growth. On the assumption that $\Delta\xi$ is serially uncorrelated, I can derive consistent estimates for λ_1 and λ_2 using a two-step difference generalized method of moments (GMM) approach proposed by Arellano and Bond (1991). I use lagged values of ω_{it} dated $t-3$ and earlier as instruments for $\Delta\omega_{t-1}$ and $\Delta\omega_{t-2}$ as proposed by Guiso et al. (2005).

Table 2. Two-step difference GMM estimation for TFP shock estimation

Dependent Variable = TFP Growth	(1)	(2)
Productivity Growth (t-1)	-0.192***	0.055
	(0.007)	(0.174)
Productivity Growth (t-2)		-0.052
		(0.034)
Year Dummies	Yes	Yes
Number of Instruments	17	15
Arellano-Bond test for autocorrelation		
AR(1) (Z-statistic)	-39.56***	-4.07***
AR(2) (Z-Statistic)	-8.05***	2.04*
Sargan test for over identifying restrictions (χ^2 Test Statistic)	23.84***	2.26
Hansen test for over identifying restrictions (χ^2 Test Statistic)	21.48**	2.35
N (Establishments)	11908	9737
N	50882	38772

Note: Dynamic panel data estimation, two-step difference GMM. GMM type instruments include omega t-2 (t-3) and earlier as uncorrelated with the error term in year t. AR(p) tests for pth order serial correlation in errors. Sargan and Hansen tests for overidentification of instruments test the null hypothesis that the instruments are valid and correctly specified. Standard Errors in the parentheses. * if p<0.05, ** if p<0.01, *** if p<0.001.

[23] Guiso et al. (2005) provides a similar equation to us but use value added instead of productivity. I believe that both equations reflect firm-level shock estimation but use different indicators for firm performance.

Table 3. Auto-covariance structure of estimated TFP shock with its lags

Lag Levels	TFP Shock
t	0.227
t-1	-0.074
t-2	-0.002
t-3	-0.007
t-4	-0.004
t-5	-0.002
t-6	-0.001
t-7	-0.010

Note: Autocovariance at lag level t is the variance of the variable

In Table 2, I show the estimation of equation 12 by changing the lag structure of the explanatory variables. The test for serial correlation of the residuals derived by Arellano and Bond (AR-2 test statistic) tests the assumption on the correlation of $\Delta\xi_{it}$. The test statistic is standard normally distributed under the null hypothesis of no serial correlation. In column 2 where I have two lags for productivity growth, I see that the test statistic for the second lag is significant at 5% level of significance. However, the covariance structure of $\Delta\xi_{it}$ shows no evidence of serial correlation beyond the first lag (see Table 3). To test for specification bias, I use the standard over-identifying restrictions test (Sargan and Hansen tests). Under the null hypothesis of correct specification, I observe that the χ^2 test statistic is insignificant at 5% level of significance (see Column 2). This provides further evidence against misspecification in TFP shock estimation. The insignificant coefficients in Column 2 can be interpreted as evidence for lack of trend effects in productivity growth. Based on the above tests in Table 2 and Table 3, I conclude that the estimated $\Delta\xi_{it}$ can be interpreted as a "true" exogenous shock to a firm's productivity in period t. The estimated $\Delta\xi_{it}$ represent shocks of small magnitude compared to more structural shocks such as COVID-19, migration boom, trade shock, unanticipated technology shock, etc. Nevertheless, it is more appropriate to isolate time varying establishment specific changes to productivity.

To supplement $\Delta\xi_{it}$, I also use the productivity of the frontier establishment ($\omega^F_{i,t}$) as a sector specific time variant instrument for CTF. This is based on the assumption that changes to the productivity of the frontier establishment does not have a direct influence on the demand for apprentices for the i^{th} establishment. The influence of frontier would only materialize if it leads to convergence or divergence of the ith establishment, i.e., changes to the productivity of the i^{th} establishment through technology diffusion (Comin & Hobiijn, 2006; Andrews et al., 2015). Hence, I use $\Delta\xi_{it}$ (proxy for firm specific productivity shock) and $\omega^F_{i,t}$ (proxy for sector-specific changes to CTF) as plausible instruments to derive exogenous variation in "closeness to the frontier" variable.

4.3. Control function approach

The two instruments are combined into a first-stage regression as follows:

$$CTF_{i,t} = {\omega_{i,t}}\Big/{\omega^F_{i,t}} = f(\omega^F_{i,t}, \Delta\xi_{i,t}, \omega_{i,t-1}) \qquad \text{-(13)-}$$

Such that $f(.)$ is a linear function of parameters and I add ω_{t-1} to incorporate level effects of productivity. Since $Cov(\Delta\xi_{it}, \Delta\xi_{i,t-1}) \neq 0$ (see Table 3), including ω_{it-1} ensures including a potential omitted variable in the first stage

regression. However, typical two-stage least squares does not give precise estimates when the dependent variable in second stage is a count variable with variance greater than the mean, i.e., over-dispersion (Hausman et al., 1984)[24]. Furthermore, Wooldridge (2010) notes that combining a linear first-stage regression with non-linear second stage creates the "forbidden regression"[25]. This functional misspecification occurs as the fitted values from the linear first-stage do not satisfy the exclusion restriction required in a two-stage least squares system.

Wooldridge (2010) proposes a control function approach where the error term from the first-stage linear regression is included in the second stage linear regression. The intuition is that first-stage error term refers to the unexplained variation in CTF and if one controls for this 'back-door', consistent estimates in the second stage regression can be achieved by bootstrapping the first and second stage regressions (Guo & Small, 2016; Huntington-Klein, 2021). In this study, the first stage excludes the variation in CTF due to idiosyncratic changes in productivity and frontier productivity (the instruments). The unexplained variation of CTF is then controlled in the second stage to produce causal effect on the demand for apprentices. The necessary assumption for the control function approach is that the instruments be uncorrelated with the first-stage error term. The error term acts similar to any other control variable in the second stage regression, i.e., its significance can be examined using conventional hypothesis testing in the second stage results. Using the control function approach offers a flexible generalization of the direction of causality problem in instrumental variable regressions.

I incorporate the control function approach in equation 9 by assuming a linear form for the first stage equation and incorporating the error term from this equation into the second stage model:

$$\text{First Stage:} \quad CTF_{it} = \alpha_0 + \alpha_1 \omega_{i,t}^{F} + \alpha_3 \Delta\xi_{i,t} + \alpha_4 \omega_{i,t-1} + \alpha_5 X_{it} + \eta_{it} \qquad \text{-(14)-}$$

$$\text{Second Stage:} \quad T_{it} = \beta_0 + \beta_1 CTF_{it} + \beta_\eta \hat{\eta}_{it} + X_{it} + H_i + \delta_t + \epsilon_{it} \qquad \text{-(15)-}$$

Where, $\hat{\eta}_{it}$ is the predicted error term from the first stage regression substituted in the second stage negative binomial regression with β_η as the parameter of interest. I estimate equation 14 using a simple random effects model for the entire sample with heteroskedastic robust standard errors clustered at the establishment level. I test the instrument strength in the first-stage regression with an F-test and the validity of the instruments using a Hansen *J*-statistic test for over-identifying restriction test for equation 14.

The last problem with equation 15 is that it potentially suffers from omitted variable bias as time-invariant unobserved factors might confound our analysis. In linear estimation, such factors are controlled by using fixed effects regression. However, in non-linear estimation models like negative binomial regression, conditional fixed effects have been shown

[24] I test for over-dispersion using a *goodness-of-fit* test of the model with a poisson model. I compute both deviance statistic and Pearson statistic and show that they are significantly different to zero. I also conduct the over-dispersion test proposed by Cameron & Trivedi (2010) and find over-dispersion in the variable for demand for apprentices.

[25] First coined by Jerry Hausman, "forbidden regression" describes a system of equations where I replace a non-linear function of endogenous explanatory variables with the same non-linear function of fitted values from a first-stage estimation (see Wooldridge (2010) for more on consistency of instrumental variables regression).

to be unreliable and not a 'true-fixed effects model' (Allison & Waterman, 2002; Guimaraes, 2008)[26]. To overcome this issue, Allison (2009) proposes a model where time-varying regressors are decomposed into a within-establishment component and a between-establishment component (Mundlak, 1978; Neuhaus & Kalbfleisch, 1998). Under this approach, the coefficients for the within-establishment component will be identical to a fixed effects regression. Therefore, I decompose each time-varying variable equation 8 into a between component, such that, $\overline{CTF_i} = n^{-1}\sum_{t=1}^{n_i} CTF_{it}$ and a cluster component $(CTF_{it} - \overline{CTF_i})$. By making some simplifications, I can then change equation 15 to equation 16:

$$y_{it} = \beta_0 + \beta_1 CTF_{it} + \beta_\eta \hat{\eta}_{it} + X_{it} + H_i + \delta_t + \gamma\overline{CTF_i} + \phi\overline{X_i} + \delta\overline{\eta_i} + \epsilon_{it} \qquad \text{-(16)-}$$

β_1 captures the within effect of change in CTF on the establishment's demand for apprentices. Equation 16 is also referred to as the correlated random effects model (CRE), first proposed by (Mundlak, 1978) and allows us to simultaneously estimate the establishment specific effect whilst controlling for between effects[27]. In equation 17, γ, ϕ and δ are the difference of the within and between effects and allow us to implicitly perform an augmented regression test. For instance, if γ or ϕ or $\delta = 0$, equation 16 collapses into equation 15. This implies that there would not be any difference in the fixed effects and the between effects model (Schunck, 2013). The combination of the CRE model and the control function approach allows us to provide a novel methodological contribution to the literature as I not only account for omitted variable bias but also introduce a exogenous variation in CTF.

5. What does the frontier look like?

The "proximity to the frontier" literature defines the frontier based on sectoral leaders, national leaders, regional or global leaders (Bartelsman et al., 2008; Andrews et al., 2015). In Figure 2, I show some key stylized facts about firm productivity in Germany. The frontier is defined as the 95th percentile of the intra-industry productivity distribution. Turnover (defined as log of sales, panel (a)) appears to have a strict positive relation with total factor productivity. The most productive firms observably differ based on turnover than even the near frontier firms at the 80th percentile of the productivity distribution. This is different to the analysis by Lochner and Schulz (2021) who argue that firms at the 90th percentile are not different to the firms at the 80th percentile in Germany. Similar to Lochner and Schulz (2021), I observe an inverse U-shaped relation between productivity and firm size (defined as log of employment, panel (b)). The positive relation between productivity and size is only seen up to the median firm from where decreasing returns to scale induces high productive firms to have a smaller workforce.

Although frontier firms are smaller in size, panel c) shows that they pay higher average wages which is congruent with cross-country literature on wage premiums and within-country literature on wage inequality (Card, Cardoso, et al., 2018;

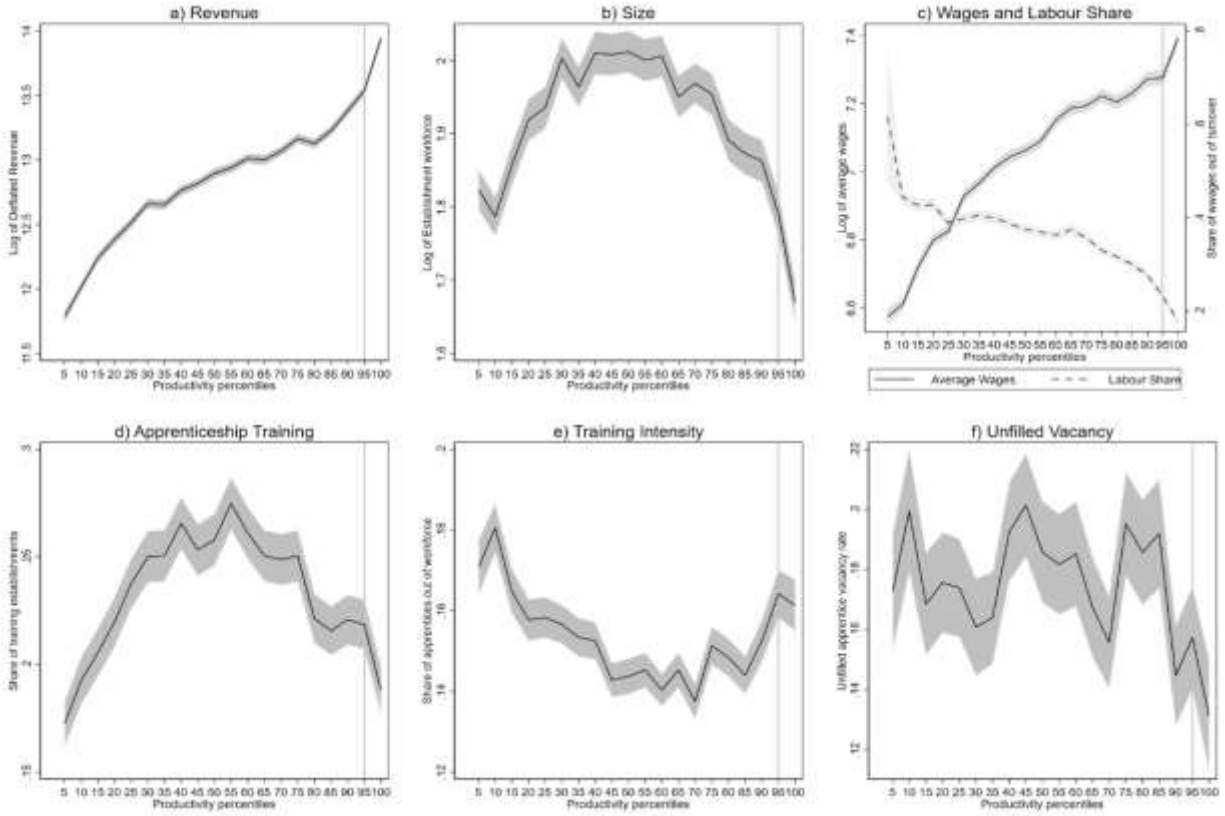
[26] Econometrically, using conditional likelihood to incorporate establishment fixed effects in a negative binomial structure only demeans the over-dispersion parameter and not the "true fixed effects" in terms of time-invariant unobserved establishment specific variables

[27] Simulation studies assessing the consistency of the hybrid model report that this method *could* yield small bias in coefficient estimates for non-linear models such as in a negative binomial regression (Brumback et al., 2013, Goetgeluk & Vansteelandt, 2008). I test the presence of such bias by adding polynomial functions of the means of the explanatory variable. In our tests, I see that these polynomial terms are not statistically significant and our estimates do not change much and therefore I can be confident that specification bias is not going to be an issue for our study.

Criscuolo et al., 2020). Furthermore, I see the similar trend of falling labour share of value added among the frontier firms in panel c) as seen by Autor et al. (2020). Therefore, firms at the frontier, or "superstar firms" have greater ability to extract rents from its skilled workers through higher monopsony power in wage determination. So on one hand, frontier firms pay higher average wages, but on the other hand, these wages go to very few individuals highlighting increasing sorting of workers towards the frontier firms.

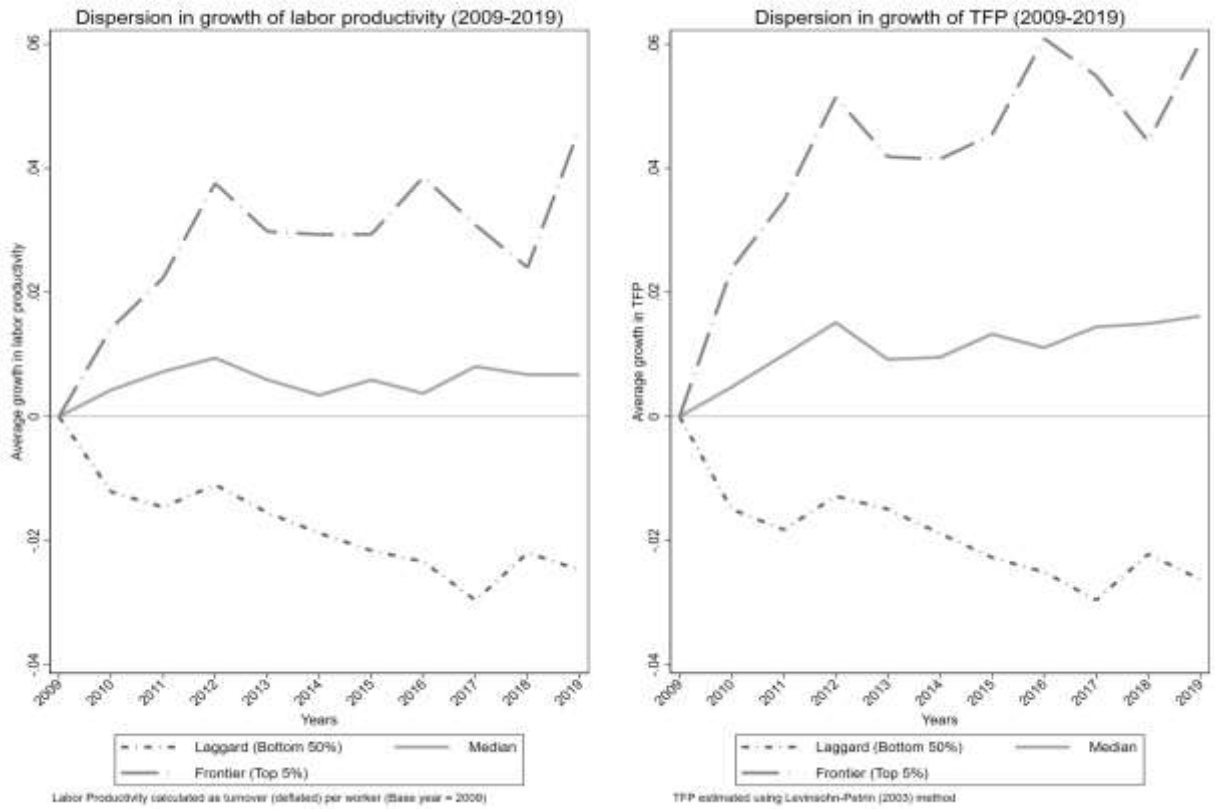
With respect to the incidence of training firms, panel d) shows the share of training firms by productivity percentile. Following a similar inverse U-trend to firm size, firms that are closer to the productivity frontier have a lower incidence of apprenticeship training. Here apprenticeship training is defined as a firm with at least one apprentice. Out of those who do train apprentices, frontier firms have a higher training intensity than median firms but not more than low productivity firms who appear to be the most apprenticeship intensive (panel e). A similar sorting story can be seen in the apprenticeship market as frontier firms are more likely to fill their apprentice vacancies (panel f) than non frontier firms. As far as I know, this is the first evidence regarding the relation between frontier firms and apprenticeship training in Germany. Frontier firms are less likely to train apprentices but if they are a training firm, they train more apprentices than non-frontier firms.

Figure 2. Characteristics of frontier firms by key firm observables



I further contextualize the productivity differences between frontier and non-frontier firms in Figure 3 by dividing firms into frontier, median and laggard firms[28]. In Figure 3, the vertical axis denotes the average growth in year t within a firm type from the base year, i.e., 2009. The dispersion in productivity matches the conclusions for a wide range of countries (Andrews et al., 2015, 2016; Bouche et al., 2021). For both labour productivity (defined as log of value added per worker) and total factor productivity, I see average annualized productivity growth to be negative and decreasing for laggards but positive and increasing for the frontier establishments with near-zero growth for the median group of establishments. I normalize 2009 as the start of the figure and I see that the three years after the financial crisis in 2008 saw the greatest divergence between the three types of firms. In this period, average growth of the frontier establishments increased to more than three times to that of the median establishments. Literature on the role of the financial crisis in influencing post-crisis firm performance attributes this dispersion to financial fragility, exposure to tight credit conditions and exposure to international markets (Duval et al., 2020). The macro trends reveal the significance of widening productivity growth between the frontier and the non-frontier firms.

Figure 3. Evolution of firm productivity growth by frontier, median and laggard firms in Germany



[28] Frontier establishments are the ones with productivity greater than equal to the 95th percentile of the within-industry productivity distribution, median establishments have productivity between 50th and 95th percentile and laggards are those that have productivity less than 50th percentile.

6. Results

6.1. Proposition 1: *Convergence effect*

The first proposition derived from equation 8 suggests that relative position in the productivity distribution will have a positive effect on demand for apprentices as long as $\sigma > \rho$. In Table 4[29], I test this proposition by estimating equations 14-16. The first column reports the effect of CTF on demand for apprentices only controlling for establishment size: an increase in number of vacancies by 48.7% for 1% increase in CTF ($\beta = 0.397, p < 0.001$). The positive effect goes in line with our theoretical model in equation 8. The effect reduces in magnitude as I control for sector, region, time fixed effects in column 2. With a full set of canonical controls and controlling for demeaned values of time-varying confounders (Column 4), I conclude that a 1% increase in the closeness to the productivity frontier is associated with a 25% decline in the demand for apprentices ($\beta = -0.288, p < 0.001$). The change from a positive to a negative effect is due to workforce characteristics and time-invariant confounders. I comply with the literature on apprenticeship training that emphasises the importance of workforce variables as proxies for skill endowment, hiring and training costs, and business outlook (Mohrenweiser & Zwick, 2009; Mühlemann, 2016). Using the coefficient on the between effect as an augmented regression test, I find no discernible difference between a traditional fixed effects model and the random effects model.

The control function approach allows us to further control potential reverse causality emanating between a firm's position in the productivity distribution and its demand for apprentices. In columns 5 and 6, I only use productivity of the frontier establishment as an instrument. Although the instrument is marginally strong, it fails to provide statistical significance in the second stage. Moving to the fully specified model in Column 8, I observe a strong, negative and statistically significant relation between closeness to the frontier and demand for apprentices. Specifically, a 1% increase in CTF leads to a 39.8% decrease ($\beta = -0.504, p < 0.001$) in the firm's demand for apprentices, *ceteris paribus*. In the Column 8, the explanatory power of 'between' effect ($\overline{\overline{CTF}}$) and the first-stage residual is weak and insignificant. This means that once we control for the 'backdoor" explanations, first stage residual, time constant factors do not influence our model. However, since the augmented regression test reveals statistical significance of the demeaned variables, Column 8 is still preferred over Column 7.

As long as instruments satisfy the assumptions of 1) instrument strength, and 2) exclusion restriction, I can credibly argue that the main effects in Column 8 are strong and representative. First stage regression results as well as joint F-test provide conclusive evidence of instrument strength. The assumption of exclusion restriction implies that the instruments are uncorrelated with the demand for apprentices and only influence demand through a firm's position on the productivity distribution. I compute the Hansen J-Statistic test for over-identifying restrictions to provide support for exogeneity assumption. A low value of the J-Statistic (J-statistic = 5.31, Column 8) fails to reject the null hypothesis of instrument exogeneity. Based on these tests and the identification rationale, I argue that I have valid instruments for our study. If one views total factor productivity as a proxy for technological endowment of a firm, convergence in a firm's technological ability to an industry frontier leads to negative effects on its demand for apprentices.

[29] The coefficients in Table 4 show the incidence-rate ratios (IRR) that refer to the change in log of expected counts of the dependent variable. To simplify interpretation, one can interpret the coefficients as percentage change ($\% \, change = [\exp(IRR) - 1] * 100$) for a 1 p.p. change in CTF.

Table 4. Effect of Productivity Convergence on demand for apprentices

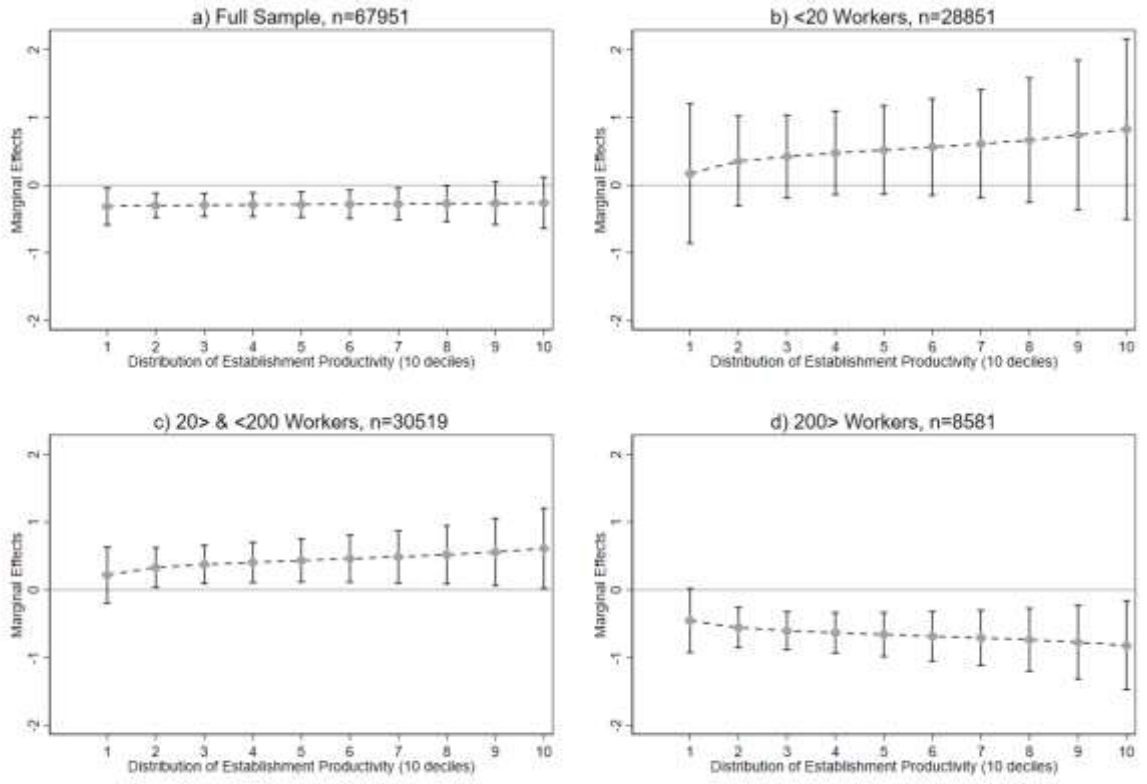
	Demand for Apprentices							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
					First Stage			
Productivity of the frontier (ω_{Ft})					-0.003*	-0.021***	-0.088***	-0.086***
					(0.001)	(0.002)	(0.000)	(0.000)
Productivity Shock ($\Delta\xi_{it}$)							0.097***	0.098***
							(0.000)	(0.000)
Lagged productivity (ω_{t-1})							0.099***	0.101***
							(0.000)	(0.000)
					Second Stage			
CTF	0.397***	0.196*	-0.153*	-0.288***	0.434	-1.734	-0.442***	-0.504***
	(0.091)	(0.091)	(0.078)	(0.086)	(6.521)	(1.147)	(0.128)	(0.139)
$\overline{CTF}$				0.133		1.147		-0.014
				(0.175)		(0.973)		(0.295)
First Stage Residual (η)					-0.588	1.455	0.305	1.883
					(6.522)	(1.149)	(1.039)	(1.084)
Controls								
Establishment Controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Year, Region, Sector	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Augmented Regression Test				163.47***		121.05***		75.24***
F-Stat (First Stage)					6.08*	70.81***	3.2e+05***	1.6e+05***
Hansen J-Statistic							10.35	5.31
N (Establishments)	19034	19034	18267	18267	18267	18267	8438	8438
N	71562	71562	67951	67951	67951	67951	31184	31184

Note: Establishment controls include number of skilled workers (log), share of hihgly skilled workers, workforce turnover rate and share of part time workers. Other control variables include establishment size dummies, works council and (log) number of school graduates. Standard Errors in the parentheses. * if p<0.05, ** if p<0.01, *** if p<0.001. Augmented Regression test is a test of joint significance of all the between effects in the regression where the null hypothesis represents absence of between effects. F-statistic for instrument strength estimated for 3 instruments with 33 degrees of freedom. Null Hypothesis of Hansen J test is of exogeneity in instruments.

In Figure 2, apprenticeship training intensity and training firm characteristics are non-uniform across the productivity distribution. Therefore, I expect the main effect to vary based on firm's position in the productivity distribution. To test this, I include a quadratic term in equation 9 and control for establishment means of time-varying variables. In panel (a) of Figure 4, I do not find non-linear effect for the full sample of establishments. The effect size is a decrease in demand for apprentices by 25% on average which mirrors the main effect in Column 4 in Table 4. For establishments with less than 20 workers (panel b), I find positive and increasing effect of productivity catch-up on apprentice demand. The lack of statistical insignificance indicates that small establishments are concerned more by whether to train or not (extensive margin) rather than how many apprentices to train (intensive margin) (Muehlemann & Wolter, 2014).

The results are different when comparing medium sized to large establishments (panel c and d). On one hand, I observe a positive and increasing catch-up effect for medium sized firms. On the other hand, for large establishments, I find negative and decreasing catch-up effect similar to the main results. Since large firms have a high ρ, i.e., decrease in marginal benefits from an additional apprentice due to decreasing returns to scale, the negative effect for large firms is intuitive from equation 6. The existence of non-linear effects shows that there might be a functional relationship between ρ and CTF for large firms. For medium sized firms, the effect is positive indicating the ability of these firms to extract higher benefits from apprenticeship training (high B_t^T). Therefore, the link between CTF and demand for apprentices depends upon the size of the firm.

Figure 4. Non-linear effects of productivity catch-up by establishment size



Note: Marginal effects are Incidence rate ratios. % change = exp(IRR)-1. Confidence intervals displayed at 5% level of significance

6.2. Proposition 2: *Cost of training effect*

For this analysis, I proxy costs of training using an industry-region-size aggregated measure from the BIBB Cost-Benefit Surveys. Since these measures are aggregates, I view these costs as proxies for expected costs for apprenticeship training per apprentice rather than realized costs. Expected costs are appropriate for studying demand for apprentices as it indicates the revealed preferences of the firm rather than actual realised preferences. In Table 5, I provide alternate specifications conditional on net costs of training apprentices (gross costs – gross benefits)[30]. Controlling for net costs of training yields insignificant change to our main results (Columns 1 and 2 in Table 5). Therefore, I argue that productivity convergence has maintains its negative relation with apprentice demand even after controlling for net costs (gross costs and gross benefits in Appendix).

However, the cost of training effect stipulates how the heterogeneity in expectation of training costs influences demand for apprentices. I test this effect by interacting CTF with expectation regarding net costs of training in Columns 3 and 4 of Table 5. I find a negative interaction effect which shows that firm's training cost expectations play a role across the productivity distribution. Specifically, for establishments that are closer to the frontier, increase in expected costs decreases the demand for apprentices by more than for establishments further away from the frontier. These 'near-frontier' establishments face a larger decrease in marginal productivity of apprentices than firms further away from the frontier (high ρ). Near the frontier, firms that catch-up need to be efficient in their human resource costs and thus if they see costs that do not have proportionate benefits, they cut down on those costs and hence reduce their demand for apprentices.

Table 5. Effect of catch-up on demand for apprentices conditional on expected net costs of training

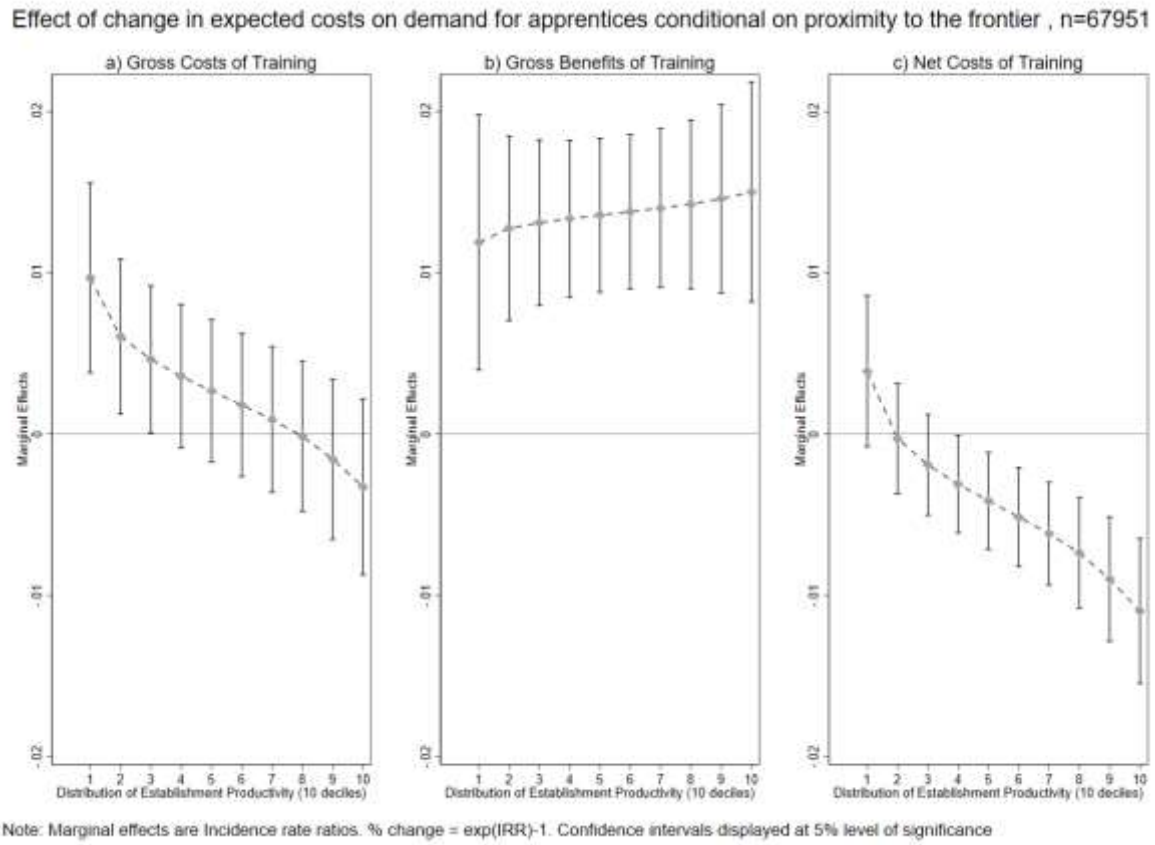
	Demand for Apprentices			
	(1)	(2)	(3)	(4)
CTF	-0.276**	-0.497***	0.251	0.737**
	(0.087)	(0.138)	(0.153)	(0.233)
Net Costs	-0.004*	-0.000	0.052***	0.136***
	(0.002)	(0.003)	(0.013)	(0.021)
CTF# Net Costs			-0.064***	-0.155***
			(0.015)	(0.023)
First Stage Residual (η)		1.956		3.827***
		(1.082)		(1.116)
N (Establishments)	18267	8438	18267	8438
N	67951	31184	67951	31184

Note: Costs measured as costs of training per apprentice per year in '000 Euros. Establishment controls include number of skilled workers (log), share of highly skilled workers, workforce turnover rate and share of part time workers. Other control variables include establishment size dummies, works council and (log) number of school graduates. All models control for centered values of time varying independent variables. Standard Errors in the parentheses. * if p<0.05, ** if p<0.01, *** if p<0.001.

[30] See Appendix Table 7 and Table 8 for tables on gross benefits and gross costs of training.

To further explain the cost of training effect, I show the marginal effects of expected gross costs, gross benefits and net costs on demand for apprentices across the productivity distribution (Figure 5). Increase in net costs of training appear to dissuade firms closer to the productivity frontier to demand apprentices. For firms further away from the frontier, I get weak positive and insignificant effects. The non-linearity in effects for net costs of training is similar to those for gross costs of training (Panel A). For expected gross benefits of apprenticeship training, I find a statistically significant positive effects on demand for apprentices. Since expected benefits are derived from an apprentice's productive work in the firm, demand for apprentices would depend upon the complementarity of apprentice's productivity on firm productivity (Muehlemann & Wolter, 2014). I argue that firms closer to the productivity frontier demand fewer apprentices because apprentice's marginal increase in productivity is less than its marginal increase in costs of training. This is in line with our model in equation 8 where the net cost of training have a negative effect on establishment's demand for apprentices. These results have policy consequences since near-frontier establishments appear to be 1) more sensitive to changes in expected costs, and 2) decrease their demand for apprentices further due to cost concerns. Since these establishments are the highly productive, appropriate training incentives need to be channelled to them to create more productive demand for apprentices.

Figure 5. Non-linear effects of expected costs of training on apprenticeship training

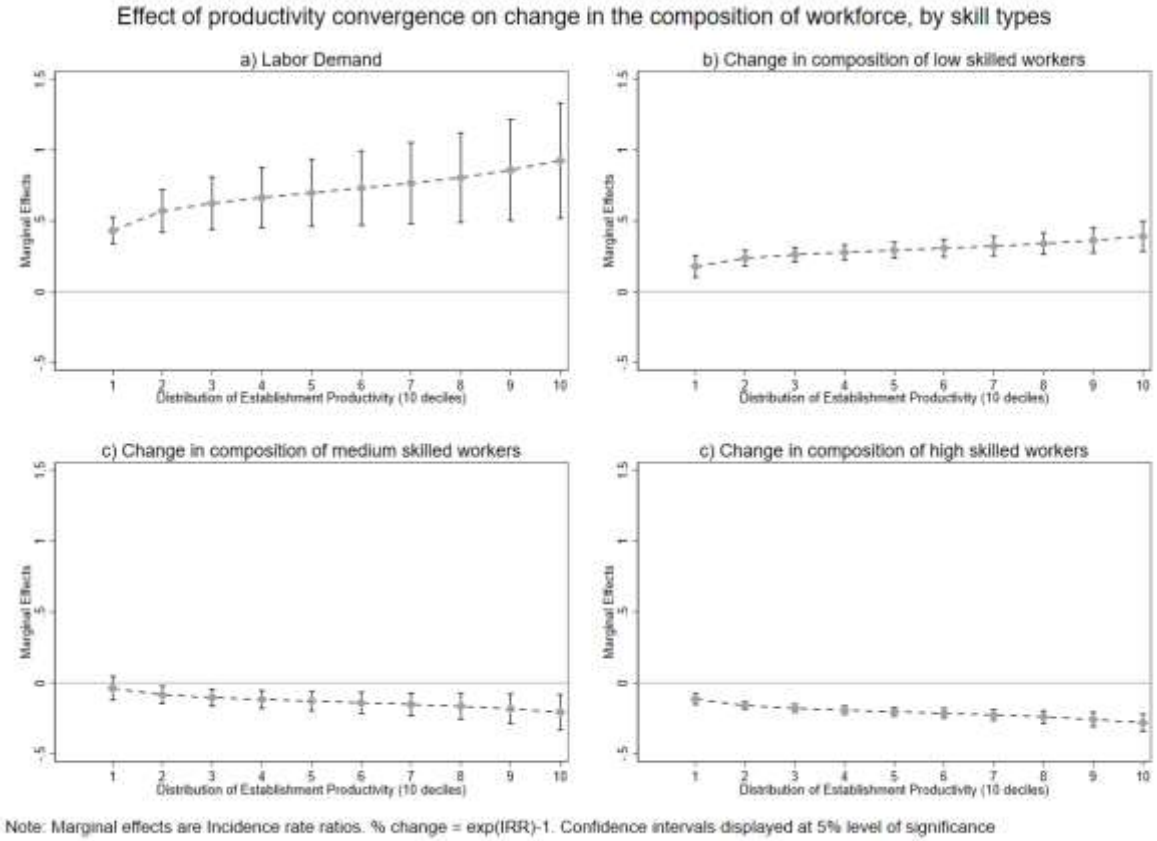


6.3.*Labour substitution effect*

Reports on administrative data on the apprenticeship market recurrently find firm size as a crucial determinant of apprenticeship demand in Germany (BiBB, 2018, 2022). If increase in firm productivity induces an increase in firm size, fall in demand for apprentices must be due to a composition effect. Specifically, firms might want to hire more of

productivity enhancing (high) skilled workers rather than invest in costly training for young labour market entrants (Vandenbussche et al., 2006; Vandenberghe, 2018). Such a *labour substitution effect* is in line with the skill biased technological change literature where skills complementary to technological change, usually found in jobs requiring university degrees, will be more demanded than skills in jobs requiring vocational and technical skills. Since productivity is a measure of disembodied factor augmenting technological change, one can argue that increase in productivity indicate broad based improvement in firm's technology (Hecht, 2018). To test the composition effect, I adapt the methodology in Abowd et al. (2007) and use change in the share of workers by skill type as the dependent variable for the labour substitution effect. For each worker type, I use a linear regression with establishment fixed effects and a two-stage least squares fixed effects regression to 1) control for time invariant effects, and 2) control for plausible endogeneity in CTF (See Table 10).

Figure 6. Non-linear effects of productivity catch-up on overall labor demand and workforce composition



In Figure 6, I show the effects on overall labour demand and composition effects across the productivity distribution. In panel a), dependent variable is the growth rate of establishment workforce in period t. In conformity with the productivity-size relation, productivity catch-up has a positive and increasing effect on employment growth. A 1% increase in closeness to the frontier is associated with a 61% increase in employment growth rate (Column 1 in Table 10). This effect increases from 39% for the firms at the 1st decile of the productivity distribution to 80% for firms at the 9th decile indicating the presence of non-linearity.

Within this overall effect, I find a positive catch-up effect on the composition of low skilled workers but a negative catch-up effect on the composition of medium skilled and high skilled workers. First, the positive effect hints at a potential

replacement effect where establishment substitute roles performed by apprentices by low-skilled workers. In their recent work, Aepli and Kuhn (2021) show how increase in the share of immigrant workers negatively influenced the number of apprentices at the firm level. The increase in low-skilled worker composition supports their conclusion as a plausible explanation of substitution of apprentices for low-skilled workers.

Secondly, the negative effect on medium-skilled worker composition is a potential explanation for the effects I observe for apprenticeship demand. Since firms train apprentices for jobs occupied by medium skilled workers, decline in apprenticeship demand can be the consequence of tasks and work of medium-skilled workers becoming obsolete due to routine biased technological change or being out-sourced/off-shored (Goos et al., 2014). As firms post apprenticeship vacancies with an expectation of their skill demand beyond the training period, decline in medium skill demand in period t is a sign of lower medium skill demand in period $t + s$ (where s is the duration of training) and hence a decline in demand for apprentices in period t. Thirdly, the decrease in high-skilled worker composition is to reduce high human resource costs. The superstar firm model stipulates that firms close to the frontier save on large overhead costs due to higher growth of value added relative to wage costs (Lotti & Sette, 2019; Autor et al., 2020; Stiel & Schiersch, 2022). Such firms invest in advanced technologies like artificial intelligence that replace the tasks of high skilled workers at the cost of manual low skilled tasks.

6.4. Heterogeneous effects

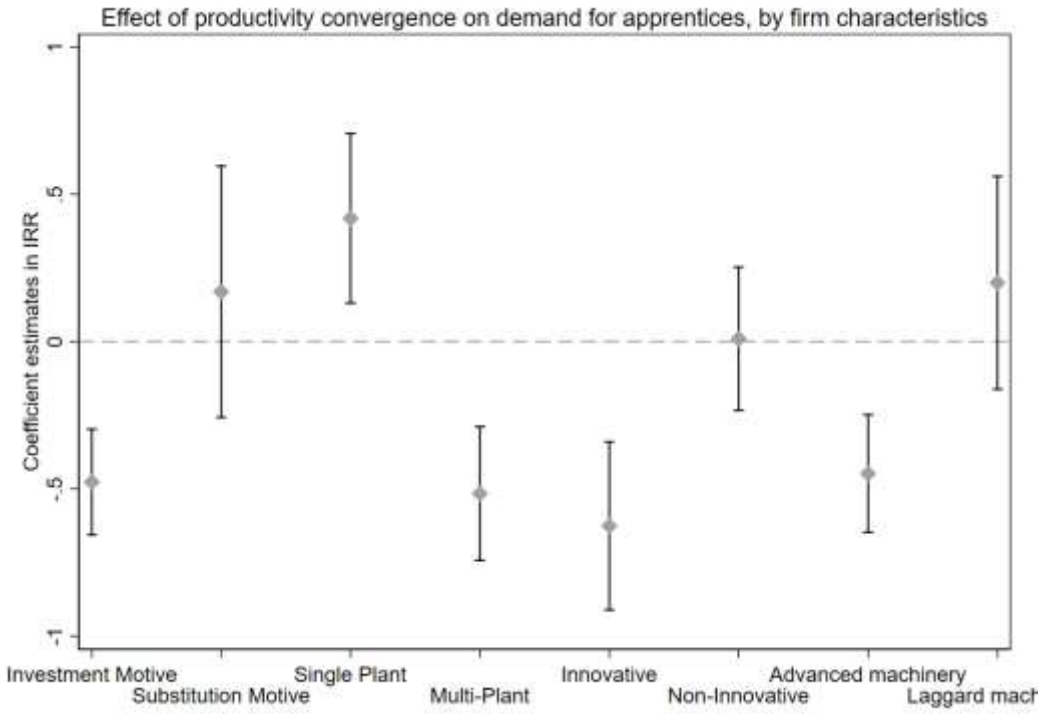
In this section, I provide evidence on the heterogeneity in the main effects by firm characteristics. Specifically, the negative effect of productivity convergence is seen in establishments that have an investment rationale to train apprentices, establishments that are innovative, belong to a multi-plant ownership structure and possess advanced technical equipment.

An established fact in the apprenticeship training literature is that firms train apprentices either with an investment motive or a substitution motive (Mohrenweiser & Backes-Gellner, 2010). Firms that invest in training apprentices and provide high retention rates are categorised by investment motive whereas firms that substitute low-skilled work with apprentices due to the lower costs involved are categorised by substitution motive. In Figure 7, I find productivity convergence having a negative effect for investment oriented firms. For firms with a substitution motive, I do not find significant effects. The negative effect for investment firms might be a sign that apprenticeship training is not complementary to firm's skill demand needs as they improve their productivity.

In Figure 4, I showed how the productivity convergence effect is negative for large firms in Germany. Since large establishments are more likely to be a multi-plant, innovative firm with advanced technical equipment, I would expect a negative catch-up effect for these firm characteristics (Luttmer, 2010; Andrews et al., 2015). In Figure 7 and Table 11, I find that establishments that are part of a multi-plant firm have a negative effect of CTF (β_1=-0.516, p<0.001, column 5) on demand for apprentices whereas single plant establishments have positive effects (β=0.419, p<0.01). Establishments who do either product or process innovation or both appear to reduce their demand for apprentices in year t due to productivity convergence. Similarly, establishments with advanced technical equipment follow the similar pattern as against establishments that do not have new and advanced operating machinery/equipment. Specifically, establishments that are innovative and advanced technically reflect characteristics of a frontier firm. These results buttress the link between productivity and technology in that more productive firms are more technology and innovation intensive, and

lack of complementarity of apprenticeship training with technology growth is potentially the reason for the decline in demand for apprentices.

Figure 7. Heterogeneous effects of productivity convergence on demand for apprentices



6.5. Robustness Checks

6.5.1. Firm's training characteristics

The effect of converging to the productivity frontier on apprenticeship demand might suffer from various issues. First, firms might decrease the number of vacancies if they have a greater ability to fill those vacancies. The decrease in vacancies can also be confounded by a higher retention rates at the firm level. Second, firms might reduce training apprentices for providing further training to its own workers. Moreover, our results might be influenced by measurement errors in productivity estimation, sample selectivity, and influential control variables. In this section, I address these issues by means of robustness checks of the main effects.

In Table 12, I show how the main results of this study are robust to establishment's training characteristics. In Columns 1 and 2, I report the full sample results similar to Columns 4 and 8 in Table 4. I find support for the negative effect of productivity convergence on demand for apprentices across all specifications for which convergence is achieved. However, inability to fill apprentice vacancies is positively associated with firm's demand for apprentices (Columns 3 and 4 in Table 12). Firms that have a higher share of their apprentice vacancies unfilled by trainees post more apprentice vacancies. In Columns 5 and 6, I see that an establishment's ability to retain apprentices negatively influences demand for apprentices. Higher retention rate reflects greater bargaining power, screening ability and quality of training by the firm and subsequently, a higher confidence to fill vacancies (Goeggel & Zwick, 2009; Dummert et al., 2019). Firms with these characteristics are confident in their ability to attract apprentices and thus potentially post fewer vacancies as they have a higher success in filling those vacancies. Although these training characteristics at the establishment level

influence training demand, they do not induce large changes in the magnitude of the convergence effect. This shows that productivity convergence has an independent effect over and above a firm's training characteristics.

To test firms substitution form apprenticeship training to further training, I regress further training intensity on productivity convergence using linear fixed effects regression in Table 13. In Column 1, I focus on the entire sample of establishments that I used for analysing the convergence effect on apprentice demand. I find a positive and statistically significant effect of productivity convergence on share of workers sent for further training. Specifically a 1% increase in closeness to the productivity frontier leads to a 9.5% increase in share of workers sent for further training. However, I do not see strong support for this effect across medium and large establishments and only weak support among small establishments. Therefore, I argue that although convergence effect is negative for apprentice demand, there appears to be some evidence that convergence has a positive effect on in-company further training.

6.5.2. Sample Selection

Threats to internal validity of this study can arise from non-random attrition of establishments from the sample and influential establishments (outliers). Moreover, the definition of the frontier firm can influence the sampling probability of establishments as all the models are restricted to establishments below the productivity frontier. Although I account for attrition bias in productivity measurement via the probability of survival, attrition bias arises if firms in the sample do not train apprentices due to laggard productivity and expectation of plant closure.

In Table 14, I show the full sample estimates in Column 1, balanced sample estimates in Column 2, sample with frontier as the 90th percentile in Column 3, frontier by sector and size in Column 4 and Poisson model regression in Column 5. For the establishments that I consecutively assess between 2009 and 2019 (n=2375), I find productivity convergence has a negative effect on the demand for apprentices. For these establishments, the magnitude of the effect is larger than for the full sample estimates. Specifically, a 1% increase in CTF corresponds with a 34% decrease in the demand for apprentices ($\beta = -0.417, p < 0.05$). This shows that my main results are prone to attrition bias. Similarly, the convergence effect, when the frontier is the 90th percentile of the intra-industry productivity distribution, is also negative. One might argue that within an industry, small firms exist but do not compete with large firms and have their own frontier that they seek to converge to. In Column 4, I define frontier by sector and size, i.e., for each firm within an industry, depending upon its size, it will have a different frontier to converge. Changing the definition of the frontier only decreases the magnitude of the effect, i.e., a 1% increase in CTF leads to a 23% decrease in demand for apprentices ($\beta = -0.261, p < 0.01$). Lastly, count data models can also be modelled using poisson regressions. Although I find a negative convergence effect, this does not appear to be statistically significant. This could be due to the distributional assumptions of the poisson model where it assumed dispersion in the dependent variable to be equal to mean whereas in my case, the dependent variable suffers from over-dispersion.

6.5.3. Productivity Estimation

The choice of using the Levinsohn and Petrin (2003) control function estimator can be criticised due to its assumption about monotonic relationship with productivity. In this section, I use the near-universe of productivity estimation methods to show the robustness of the study to alternative methods of productivity estimation. Specifically, I estimate production functions using Ordinary Least Squares (OLS), Fixed Effects regression (FE), Olley and Pakes (1996) (OP) and the Stochastic Frontier Approach (SFA) (similar to Vandenberghe (2018)) in Table 15. The results in Table 15 show that our

main results are robust to the choice of production function estimation methodology. The results from Levinsohn-Petrin estimator are similar to the ones by pooled OLS ($\beta = -0.331, p < 0.001$, Column 2), Olley-Pakes estimator ($\beta = -0.200, p < 0.05$, Column 4) and the non-parametric estimation method of stochastic frontier approach ($\beta = -0.254, p < 0.001$, Column 6). I get insignificant effects in the fixed effects estimator a large negative effect when productivity is measured as log of value added per worker (labour productivity). The consistency in the negative convergence effect (except in the fixed effects estimator) gives me confidence that the results are robust to the choice of TFP estimation method. In addition to these robustness checks, I also test by controlling for establishment's age, competitive pressure (Heywood et al., 2019) and proxies for supply of apprentices (without school certificate, with intermediate certificate, with university certificates). In all the estimations, I find proof of a negative and statistically significant convergence effect.

7. Conclusion

This paper used a large establishment-level panel data set to analyse the effect of productivity convergence on firm's demand for apprenticeship training. I defined demand for apprenticeship training as the number of apprentice vacancies posted by an establishment in a given year. I address the simultaneity problem between training and establishment productivity using a two stage control function approach. In the first stage, I use time-varying establishment specific shocks to productivity and frontier productivity as instruments for closeness to the productivity frontier. The residuals from this first stage regression are used as a control variable in the second stage regression to control for endogeneity in our main independent variable.

Using negative binomial model, I argue that converging to the productivity frontier leads to a decrease in firm's demand for apprentices. More precisely, a 1% increase in a firm's position in the intra-industry productivity distribution leads to a 25% to 40% decrease in the demand for apprentices. This effect is negative for large firms and positive effect on demand for apprentices for medium sized firms. Furthermore, establishment that have an investment motive of apprenticeship training, belong to a multi-plant ownership structure, innovative firms, and firms with advanced technical equipment show negative convergence effect on apprentice demand. The results shown in the study highlight the heterogeneity with which firms in Germany view apprenticeship training as part of their skill formation strategy.

I explain these results through three mechanisms: 1) convergence effect, 2) cost of training effect, and, 3) labour substitution effects. Firstly, catch-up to the productivity frontier induces technological progress that is skill biased and thus reduces the demand for apprentices. Secondly, to achieve efficiency at the frontier, firms resort to cost-saving measures that induce them to reduce their demand for expensive apprenticeship training. For firms further away from the frontier, apprentices still represent a skill formation strategy to use despite high costs of training. Thirdly, I argue that firms reduce their demand for apprentices due to routine-biased technological change where they reduce their demand for medium skilled workers. This induces firms to reduce demand for routinizable work and since apprentices are trained for these tasks and work, firms subsequently reduce the demand for apprentices.

The primary contribution of this paper is on the apprenticeship training literature which stipulate the positive effect of business cycles and business expectations on firm's demand for apprentices. In this paper, I argue that productivity growth that improves firm's industry position leads to a decrease in its demand for apprentices. My empirical contribution is on the magnitude, heterogeneity and explanation of convergence effects. These effects have important policy implications. Firstly, a reduction in the number of apprentice vacancies reduces the vocational options available to

secondary school graduates in Germany. This nudges students to choose alternate educational tracks or drop out of school and go into low-skilled work. Secondly, inability to find an apprentice position due to lack of vacancies potentially impacts unemployment probability of secondary school students and adversely affects their life-time earnings. Thirdly, since apprenticeship training is a form of general training that constitutes transferable skills, this study provides an empirical contribution to the human capital literature in Becker (1962). This study provides a starting point for further research to investigate the consequences of intra-industry heterogeneity in firm performance on a firm's personnel recruitment strategy, its "make or buy" decision, and incentives to train apprentices.

References

Abowd, J. M., Haltiwanger, J. C., Lane, J., McKinney, K. L., & Sandusky, K. (2007). Technology and the demand for skill: an analysis of within and between firm differences. In: National Bureau of Economic Research Cambridge, Mass., USA.

Acemoglu, D., Aghion, P., & Zilibotti, F. (2006). Distance to Frontier, Selection, and Economic Growth. *Journal of the European Economic Association*, *4*(1), 37-74. https://doi.org/10.1162/jeea.2006.4.1.37

Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics* (Vol. 4, pp. 1043-1171). Elsevier.

Acemoglu, D., & Pischke, J.-S. (1999a). The structure of wages and investment in general training. *Journal of political economy*, *107*(3), 539-572.

Acemoglu, D., & Pischke, J. S. (1999b). Beyond Becker: Training in imperfect labour markets. *The economic journal*, *109*(453), 112-142.

Ackerberg, D., Benkard, C. L., Berry, S., & Pakes, A. (2007). Econometric tools for analyzing market outcomes. *Handbook of econometrics*, *6*, 4171-4276.

Ackerberg, D., Caves, K., & Frazer, G. (2006). Structural identification of production functions.

Ackerberg, D. A., Caves, K., & Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, *83*(6), 2411-2451.

Aepli, M., & Kuhn, A. (2021). Open labor markets and firms' substitution between training apprentices and hiring workers. *Labour economics*, *70*, 101979.

Aghion, P., Boustan, L., Hoxby, C., & Vandenbussche, J. (2009). The causal impact of education on economic growth: evidence from US. *Brookings papers on economic activity*, *1*(1), 1-73.

Allison, P. D. (2009). *Fixed effects regression models*. SAGE publications.

Allison, P. D., & Waterman, R. P. (2002). Fixed–effects negative binomial regression models. *Sociological methodology*, *32*(1), 247-265.

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

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, *58*(2), 277-297.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., & Van Reenen, J. (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly journal of economics*, *135*(2), 645-709.

Autor, D., & Salomons, A. (2017). Robocalypse now: Does productivity growth threaten employment. Proceedings of the ECB Forum on Central Banking: Investment and Growth in Advanced Economie,

Baethge, M. (2007). Das deutsche Bildungs-Schisma Welche Probleme ein vorindustrielles Bildungssystem in einer nach. *Bildung-Lernen: Humanistische Ideale, gesellschaftliche Notwendigkeiten, wissenschaftliche Erkenntnisse*, 93.

Baethge, M., & Wolter, A. (2015). The German skill formation model in transition: from dual system of VET to higher education? *Journal for Labour Market Research*, *48*(2), 97-112.

Bartelsman, E., Dobbelaere, S., & Peters, B. (2015). Allocation of human capital and innovation at the frontier: firm-level evidence on Germany and the Netherlands. *Industrial and Corporate Change*, *24*(5), 875-949.

Bartelsman, E. J., Haskel, J., & Martin, R. (2008). Distance to which frontier? Evidence on productivity convergence from international firm-level data.

Barth, E., Bryson, A., Davis, J. C., & Freeman, R. (2016). It's where you work: Increases in the dispersion of earnings across establishments and individuals in the United States. *Journal of labor economics*, *34*(S2), S67-S97.

Bassanini, A., Booth, A. L., Brunello, G., De Paola, M., & Leuven, E. (2005). Workplace training in Europe. *Available at SSRN 756405*.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of political economy*, *70*(5, Part 2), 9-49.

Bellmann, L., Brunner, L., Ellguth, P., Grunau, P., Hohendanner, C., Kohaut, S., Leber, U., Möller, I., Schwengler, B., Stegmaier, J., & Umkehrer, M. (2022). *IAB Establishment Panel (IAB BP) - Version 9321 v1*. https://doi.org/10.5164/IAB.IABBP9321.de.en.v1

Bellmann, L., Gerner, H.-D., & Leber, U. (2014). Firm-provided training during the great recession. *Jahrbücher Für Nationalökonomie Und Statistik*, *234*(1), 5-22.

Bellmann, L., Grunau, P., Troltsch, K., & Walden, G. (2014). Make or buy: train in-company or recruit from the labour market? *Empirical Research in Vocational Education and Training*, *6*, 1-18.

BiBB. (2018). VET Data Report Germany 2016/2017. Facts and Analyses to Accompany the Federal Government Report on Vocational Education and Training–Selected Findings.

BiBB. (2022). *Datenreport zum Berufsbildungsbericht 2022*. I. u. A. z. E. d. b. Bildung. https://www.bibb.de/dokumente/pdf/Datenreport%202022_20102022_online.pdf

Black, S. E., & Lynch, L. M. (2001). How to compete: the impact of workplace practices and information technology on productivity. *Review of Economics and statistics*, *83*(3), 434-445.

Blatter, M., Muehlemann, S., Schenker, S., & Wolter, S. C. (2016). Hiring costs for skilled workers and the supply of firm-provided training. *Oxford Economic Papers*, *68*(1), 238-257.

Bloom, N., & Van Reenen, J. (2010). Why do management practices differ across firms and countries? *Journal of economic perspectives*, *24*(1), 203-224.

Böheim, R., Schneeweis, N., & Wakolbinger, F. (2009). *Employer provided training in Austria: Productivity, wages and wage inequality*.

Brenke, K. (2011). Ongoing change in the structure of part-time employment. *DIW Economic bulletin*, *1*(6), 13-22.

Brumback, B. A., Zheng, H. W., & Dailey, A. B. (2013). Adjusting for confounding by neighborhood using generalized linear mixed models and complex survey data. *Statistics in medicine*, *32*(8), 1313-1324.

Cabus, S., & Nagy, E. (2021). On the productivity effects of training apprentices in Hungary: evidence from a unique matched employer–employee dataset. *Empirical Economics*, *60*(4), 1685-1718.

Cameron, A. C., & Trivedi, P. K. (1986). Econometric models based on count data. Comparisons and applications of some estimators and tests. *Journal of applied econometrics*, *1*(1), 29-53.

Card, D., Cardoso, A. R., Heining, J., & Kline, P. (2018). Firms and labor market inequality: Evidence and some theory. *Journal of labor economics*, *36*(S1), S13-S70.

Card, D., Kluve, J., & Weber, A. (2018). What works? A meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association*, *16*(3), 894-931.

Carlsson, M., Messina, J., & Skans, O. N. (2016). Wage adjustment and productivity shocks. *The economic journal*, *126*(595), 1739-1773.

Clark, D. (2000). How transferable is German apprenticeship training? IZA Seminar on,

Clark, K. B., & Summers, L. H. (1980). *Demographic differences in cyclical employment variation* (0898-2937).

Comin, D., & Hobiijn, B. (2006). An exploration of technology diffusion (No. w12314). In: National Bureau of Economic Research.

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

Criscuolo, C., Hijzen, A., Schwellnus, C., Chen, W.-H., Fabling, R., Fialho, P., Grabska, K., Kambayashi, R., Leidecker, T., & Nordstrom Skans, O. (2020). Workforce composition, productivity and pay: The role of firms in wage inequality.

Dearden, L., Reed, H., & Van Reenen, J. (2006). The impact of training on productivity and wages: Evidence from British panel data. *Oxford Bulletin of Economics and Statistics*, *68*(4), 397-421.

Dietrich, H., & Gerner, H.-D. (2007). The determinants of apprenticeship training with particular reference tobusiness expectations. *Zeitschrift für ArbeitsmarktForschung-Journal for Labour Market Research*, *40*(2/3), 221-233.

Dionisius, R., Muehlemann, S., Pfeifer, H., Walden, G., Wenzelmann, F., & Wolter, S. C. (2009). Costs and benefits of apprenticeship training. A comparison of Germany and Switzerland. *Applied Economics Quarterly*, *55*(1), 7-37.

Dummert, S., Leber, U., & Schwengler, B. (2019). Unfilled training positions in Germany–regional and establishment-specific determinants. *Jahrbücher Für Nationalökonomie Und Statistik*, *239*(4), 661-701.

Dunne, T., Foster, L., Haltiwanger, J., & Troske, K. R. (2004). Wage and productivity dispersion in United States manufacturing: The role of computer investment. *Journal of labor economics*, *22*(2), 397-429.

Dustmann, C., & Schönberg, U. (2009). Training and union wages. *The Review of Economics and Statistics*, *91*(2), 363-376.

Duval, R., Hong, G. H., & Timmer, Y. (2020). Financial frictions and the great productivity slowdown. *The Review of Financial Studies*, *33*(2), 475-503.

Ellguth, P., Kohaut, S., & Möller, I. (2014). The IAB Establishment Panel—methodological essentials and data quality. *Journal for Labour Market Research*, *47*(1), 27-41.

Faggio, G., Salvanes, K. G., & Van Reenen, J. (2010). The evolution of inequality in productivity and wages: panel data evidence. *Industrial and Corporate Change*, *19*(6), 1919-1951.

Fischer, G., Janik, F., Müller, D., & Schmucker, A. (2009). The IAB Establishment Panel-things users should know. *Journal of Contextual Economics*, *129*(1), 133.

Forster, A. G., & Bol, T. (2018). Vocational education and employment over the life course using a new measure of occupational specificity. *Social Science Research*, *70*, 176-197.

Franz, W., & Soskice, D. (1995). The German apprenticeship system. *Institutional frameworks and labor market performance: comparative views on the US and German economies*, 208-234.

Franz, W., & Zimmermann, V. (2002). The transition from apprenticeship training to work. *International Journal of Manpower*.

Fregin, M.-C., Levels, M., & Van der Velden, R. (2020). Labour market institutions and the challenge of allocating the right people to the right jobs: Evidence on the relation between labour market institutions and optimal skill

matching from 28 industrial countries. *Compare: A Journal of Comparative and International Education*, *50*(2), 257-275.
Goeggel, K., & Zwick, T. (2009). Good Occupation-Bad Occupation? The Quality of Apprenticeship Training. *The Quality of Apprenticeship Training*, 09-024.
Goetgeluk, S., & Vansteelandt, S. (2008). Conditional generalized estimating equations for the analysis of clustered and longitudinal data. *Biometrics*, *64*(3), 772-780.
Goos, M., Manning, A., & Salomons, A. (2014). Explaining job polarization: Routine-biased technological change and offshoring. *American economic review*, *104*(8), 2509-2526.
Guertzgen, N. (2014). Wage insurance within German firms: do institutions matter? *Journal of the Royal Statistical Society Series A: Statistics in Society*, *177*(2), 345-369.
Guimaraes, P. (2008). The fixed effects negative binomial model revisited. *Economics Letters*, *99*(1), 63-66.
Guiso, L., Pistaferri, L., & Schivardi, F. (2005). Insurance within the firm. *Journal of political economy*, *113*(5), 1054-1087.
Guo, Z., & Small, D. S. (2016). Control function instrumental variable estimation of nonlinear causal effect models. *The Journal of Machine Learning Research*, *17*(1), 3448-3482.
Haasler, S. R. (2020). The German system of vocational education and training: challenges of gender, academisation and the integration of low-achieving youth. *Transfer: European Review of Labour and Research*, *26*(1), 57-71.
Hausman, J. A., Hall, B. H., & Griliches, Z. (1984). Econometric models for count data with an application to the patents-R&D relationship. In: national bureau of economic research Cambridge, Mass., USA.
Hecht, J. (2018). Research and development and labour productivity: do high-tech firms exhibit labour-or capital-saving technical change? *Applied Economics*, *50*(16), 1790-1811.
Heß, P., Janssen, S., & Leber, U. (2023). The effect of automation technology on workers' training participation. *Economics of Education Review*, *96*, 102438.
Heywood, J. S., Jirjahn, U., & Pfister, A. (2019). Product market competition and employer provided training in Germany. *Industrial and Corporate Change*, *29*(2), 533-556. https://doi.org/10.1093/icc/dtz053
Huntington-Klein, N. (2021). *The effect: An introduction to research design and causality*. CRC Press.
Jansen, A., Pfeifer, H., Schönfeld, G., & Wenzelmann, F. (2015). Apprenticeship Training in Germany remains investment-focused: results of BIBB Cost-Benefit Survey 2012/13. *Federal Institute for Vocational Education and Training (BIBB) Report*, *1*.
Jansen, A., Pfeifer, H., Schönfeld, G., & Wenzelmann, F. (2017). *BIBB-Cost-Benefit Survey 2012/13.* Version GWA_1.0) Federal Institute for Vocational Education and Training. https://doi.org/10.7803/370.12.1.2.10
Koch, B., Muehlemann, S., & Pfeifer, H. (2019). Do works councils improve the quality of apprenticeship training? Evidence from German workplace data. *Journal of Participation and Employee Ownership*, *2*(1), 47-59.
Kohli, U. (2015). Explaining total factor productivity. *University of Geneva*.
Konings, J. (2008). The Impact of Training on Productivity and Wages-Evidence from Belgian Firm Level Panel Data. *Available at SSRN 1107108*.
Konings, J., & Vanormelingen, S. (2015). The impact of training on productivity and wages: firm-level evidence. *Review of Economics and statistics*, *97*(2), 485-497.
Levinsohn, J., & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies*, *70*(2), 317-341.
Lindley, R. M. (1975). The demand for apprentice recruits by the engineering industry, 1951-71. *Scottish Journal of Political Economy*, *22*(1), 1-24.
Lochner, B., & Schulz, B. (2021). Firm productivity, wages, and sorting.
Lotti, F., & Sette, E. (2019). *Frontier and Superstar Firms in Italy*.
Lüthi, S., & Wolter, S. C. (2020). Are apprenticeships business cycle proof? *Swiss Journal of Economics and Statistics*, *156*(1), 1-11.
Luttmer, E. G. J. (2010). Models of Growth and Firm Heterogeneity. *Annual Review of Economics*, *2*, 547-576. http://www.jstor.org/stable/42940340
Manning, A. (2011). Imperfect competition in the labor market. In *Handbook of labor economics* (Vol. 4, pp. 973-1041). Elsevier.
Mohrenweiser, J., & Backes-Gellner, U. (2008). Apprenticeship training–What for? Investment in human capital or substitution of cheap labour. *Swiss Leading House on the Economics of Education Working Paper*(17).
Mohrenweiser, J., & Backes-Gellner, U. (2010). Apprenticeship training: for investment or substitution? *International Journal of Manpower*.
Mohrenweiser, J., & Zwick, T. (2009). Why do firms train apprentices? The net cost puzzle reconsidered. *Labour economics*, *16*(6), 631-637.
Mohrenweiser, J., Zwick, T., & Backes-Gellner, U. (2019). Poaching and firm-sponsored training. *British Journal of Industrial Relations*, *57*(1), 143-181.

Muehlemann, S., Dietrich, H., Pfann, G., & Pfeifer, H. (2022). Supply shocks in the market for apprenticeship training. *Economics of Education Review*, *86*, 102197.

Muehlemann, S., & Pfeifer, H. (2016). The Structure of Hiring Costs in Germany: Evidence from Firm-Level Data. *Industrial Relations: A Journal of Economy and Society*, *55*(2), 193-218.

Muehlemann, S., Schweri, J., Winkelmann, R., & Wolter, S. C. (2005). A structural model of demand for apprentices. *Available at SSRN 668881*.

Muehlemann, S., & Wolter, S. C. (2014). Return on investment of apprenticeship systems for enterprises: Evidence from cost-benefit analyses. *IZA Journal of Labor Policy*, *3*(1), 25. https://doi.org/10.1186/2193-9004-3-25

Muehlemann, S., Wolter, S. C., & Wueest, A. (2009). Apprenticeship training and the business cycle. *Empirical Research in Vocational Education and Training*, *1*(2), 173-186.

Mühlemann, S. (2016). The cost and benefits of work-based learning.

Müller, S. (2008). Capital stock approximation using firm level panel data. *Jahrbücher Für Nationalökonomie Und Statistik*, *228*(4), 357-371.

Müller, S. (2017). Capital Stock Approximation with the Perpetual Inventory Method.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica: journal of the Econometric Society*, 69-85.

Neuhaus, J. M., & Kalbfleisch, J. D. (1998). Between-and within-cluster covariate effects in the analysis of clustered data. *Biometrics*, 638-645.

OECD. (2011). Measuring Productivity.

Olley, G. S., & Pakes, A. (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, *64*(6), 1263-1297. https://doi.org/10.2307/2171831

Oyer, P., & Schaefer, S. (2010). *Personnel economics: Hiring and incentives*.

Oyer, P., & Schaefer, S. (2011). Chapter 20 - Personnel Economics: Hiring and Incentives**This survey has benefitted from conversations with colleagues too numerous to mention. We thank the editors, Orley Ashenfelter and David Card, for detailed comments.E-mail addresses: oyer_paul@gsb.stanford.edu (Paul Oyer), scott.schaefer@utah.edu (Scott Schaefer). In D. Card & O. Ashenfelter (Eds.), *Handbook of Labor Economics* (Vol. 4, pp. 1769-1823). Elsevier. https://doi.org/https://doi.org/10.1016/S0169-7218(11)02418-X

Pfeifer, H., & Backes-Gellner, U. (2018). Another piece of the puzzle: Firms' investment in training as production of optimal skills inventory. *Empirical Research in Vocational Education and Training*, *10*, 1-17.

Pfeifer, H., Schönfeld, G., & Wenzelmann, F. (2011). How large is the firm-specific component of German apprenticeship training? *Empirical Research in Vocational Education and Training*, *3*, 85-104.

Pfeifer, H., Wenzelmann, F., Schönfeld, G., & Wehner, C. (2021). *BIBB-Cost-Benefit Survey 2017/18* Version GWA_1.0). https://doi.org/10.7803/370.18.1.2.10

Rohrbach-Schmidt, D., & Tiemann, M. (2016). Educational (Mis) match and skill utilization in Germany: Assessing the role of worker and job characteristics. *Journal for Labour Market Research*, *49*(2), 99-119.

Rovigatti, G., & Mollisi, V. (2018). Theory and practice of total-factor productivity estimation: The control function approach using Stata. *The Stata Journal*, *18*(3), 618-662.

Rovigatti, G., & Mollisi, V. (2020). PRODEST: Stata module for production function estimation based on the control function approach.

Rupietta, C., & Backes-Gellner, U. (2019). How firms' participation in apprenticeship training fosters knowledge diffusion and innovation. *Journal of Business Economics*, *89*(5), 569-597.

Ryan, P. (2010). Apprentice strikes, pay structure, and training in twentieth century UK metalworking industry. *Labor in the Era of Globalization, Cambridge: CUP*, 317-353.

Schönfeld, G. (2010). *Kosten und Nutzen der dualen Ausbildung aus Sicht der Betriebe: Ergebnisse der vierten BIBB-Kosten-Nutzen-Erhebung*. W. Bertelsmann Verlag.

Schultheiss, T., & Backes-Gellner, U. (2020). Updated education curricula and accelerated technology diffusion in the workplace: Micro-evidence on the race between education and technology. *Economics of Education Working Paper Series*, *173*.

Schunck, R. (2013). Within and between estimates in random-effects models: Advantages and drawbacks of correlated random effects and hybrid models. *The Stata Journal*, *13*(1), 65-76.

Solga, H., Protsch, P., Ebner, C., & Brzinsky-Fay, C. (2014). *The German vocational education and training system: Its institutional configuration, strengths, and challenges*.

Solow, R. M. (1957). Technical change and the aggregate production function. *The Review of Economics and Statistics*, 312-320.

Stiel, C., & Schiersch, A. (2022). Testing the superstar firm hypothesis. *Journal of Applied Economics*, *25*(1), 583-603.

Stockinger, B., & Zwick, T. (2017). Apprentice poaching in regional labor markets. *ZEW-Centre for European Economic Research Discussion Paper*(17-013).

Syverson, C. (2011). What determines productivity? *Journal of economic literature*, *49*(2), 326-365.

Thelen, K. (2004). *How institutions evolve: The political economy of skills in Germany, Britain, the United States, and Japan*. Cambridge University Press.

Thelen, K. (2007). Contemporary challenges to the German vocational training system. *Regulation & Governance*, *1*(3), 247-260.

Uhly, A. (2020). Alter der Auszubildenden und Ausbildungsbeteiligung der Jugendlichen im dualen System. *BIBB (ed.) Datenreport zum Berufsbildungsbericht*, 162-171.

Van Beveren, I. (2012). Total factor productivity estimation: A practical review. *Journal of economic surveys*, *26*(1), 98-128.

Van Biesebroeck, J. (2005). Firm size matters: Growth and productivity growth in African manufacturing. *Economic Development and cultural change*, *53*(3), 545-583.

Vandenberghe, V. (2018). The Contribution of Educated Workers to Firms' Efficiency Gains: The Key Role of Proximity to the 'Local'Frontier. *De Economist*, *166*(3), 259-283.

Vandenbussche, J., Aghion, P., & Meghir, C. (2006). Growth, distance to frontier and composition of human capital. *Journal of economic growth*, *11*(2), 97-127.

Walden, G., & Troltsch, K. (2011). Apprenticeship training in Germany–still a future-oriented model for recruiting skilled workers? *Journal of Vocational Education & Training*, *63*(3), 305-322.

Wolter, S. C., & Ryan, P. (2011). Apprenticeship. In *Handbook of the Economics of Education* (Vol. 3, pp. 521-576). Elsevier.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Zwick, T. (2006). The impact of training intensity on establishment productivity. *Industrial Relations: A Journal of Economy and Society*, *45*(1), 26-46.

Zwick, T. (2007). Apprenticeship training in Germany-investment or productivity driven? *ZEW-Centre for European Economic Research Discussion Paper*(07-023).

Appendix

1. Total Factor Productivity Estimation

TFP estimation typically relies on the so-called Solow residual which accounts for disembodied technical change in the production process (Solow, 1957; Kohli, 2015). It comes from a "measure of ignorance" which includes both technology factors as well as non-technology factors such as adjustments costs, scale and cyclical effects, pure changes in efficiency and measurement errors (OECD, 2011). In our study, I use the production function approach as it allows us to account for time effects, simultaneity of input choice and attrition bias. A simple Cobb-Douglas production function to estimate TFP looks like the following:

$$Y_{i,t} = A_{i,t} L_{i,t}^{\beta_l} K_{i,t}^{\beta_k} \qquad \text{-(17)-}$$

Where, $Y_{i,t}$ represents gross value added of the establishment i in period t, $L_{i,t}, K_{i,t}$ are inputs of labour and capital (all are observable to the researcher in the IAB-BP) and $A_{i,t}$ is the Hicksian neutral efficiency level of the establishment i in period t (unobservable to the researcher in the IAB-BP).

Taking natural logs of equation (18), I get,

$$y_{i,t} = \beta_0 + \beta_l l_{i,t} + \beta_k k_{i,t} + \epsilon_{i,t}$$

Where lower case letters represent natural logarithms and $\ln(A_{i,t}) = \beta_0 + \epsilon_{i,t}$ where β_0 measures the mean efficiency level across establishments and over time; $\epsilon_{i,t}$ is the time- and establishment specific deviation from β_0. If $\epsilon_{i,t} = v_{i,t} + u_q$, then

$$y_{i,t} = \beta_0 + \beta_l l_{i,t} + \beta_k k_{i,t} + \beta_m m_{i,t} + v_{i,t} + u_q \qquad \text{-(18)-}$$

Where, $\omega_{i,t} = \beta_0 + v_{i,t}$ represents establishment level productivity and u_q is an i.i.d. component representing unexpected deviations from the mean due to measurement error or other external circumstances. Since I use all variables in log form, our TFP estimates are in log forms. Estimating equation (18) yields productivity to be equal to

$$\widehat{\omega}_{i,t} = \hat{\beta}_0 + \hat{v}_{i,t} = y_{i,t} - \hat{\beta}_l l_{i,t} + \hat{\beta}_k k_{i,t} \qquad \text{-(19)-}$$

In this Levinsohn-Petrin (2003) (LP) semi-parametric technique, both selection and simultaneity issues are explicitly taken into account using a control-function approach. Specifically, unobserved productivity ($\omega_{i,t}$) is proxied through a control-function based on material inputs. This function is invertible based on the assumption that material costs are an increasing function of productivity, i.e., $\omega_{i,t} = \beta_0 + s_t(m_{i,t}, k_{it})$ where $m_{i,t}$ represent log values of material costs. Thus, equation (18) becomes,

$$y_{i,t} = \beta_0 + \beta_l l_{i,t} + \beta_k k_{i,t} + \boldsymbol{s_t}(\boldsymbol{m_{i,t}}, \boldsymbol{k_{i,t}}) + u_q \qquad \text{-(20)-}$$

The LP estimation proceeds in two stages. In the first stage, I obtain consistent estimators of labour coefficient holding capital constant under the assumption that changes in labour input influence the decision to invest in capital stock. I estimate the following equation:

Stage 1: Estimate equation (20) where $\phi_{i,t}(m_{i,t}, k_{i,t}) = \beta_0 + \beta_k k_{i,t} + s_t(m_{i,t}, k_{i,t})$

$$y_{i,t} = \beta_l l_{i,t} + \boldsymbol{\phi_{i,t}}(\boldsymbol{m_{i,t}}, \boldsymbol{k_{i,t}}) + u_q \quad \text{-(21}$$

The first stage will provide consistent estimators of labour coefficient (β_l). To recover the coefficient on capital, I use the second stage where productivity is assumed to follow a first-order Markov process, i.e., $s_{i,t+1} = E(\omega_{i,t+1}|\omega_{i,t}) + \xi_{i,t+1}$ where $\xi_{i,t+1}$ represents the news component uncorrelated with productivity and capital in period $t+1$. Conditional on establishment survival, i.e., $\chi_{i,t+1} = 1$ if $\omega_{i,t+1} \geq \underline{\omega}_{i,t+1}$ (productivity in t+1 exceeds a lower bound of productivity), the second stage equation is derived from equation (21):

$$\begin{aligned} y_{i,t+1} - \hat{\beta}_l l_{i,t+1} &= \boldsymbol{\phi_{i,t+1}}(\boldsymbol{m_{i,t+1}}, \boldsymbol{k_{i,t+1}}) + u_{i,t+1} \\ &= \beta_0 + \beta_k k_{i,t+1} + s_{t+1}(m_{i,t+1}, k_{i,t+1}) + u_{i,t+1} \\ &= \beta_0 + \beta_k k_{i,t+1} + E(\omega_{i,t+1}|\omega_{i,t}, \chi_{i,t+1}) + \xi_{i,t+1} + u_{i,t+1} \\ &= \beta_0 + \beta_k k_{i,t+1} + g(P_{i,t}, \phi_t - \beta_k k_{i,t}) + \xi_{i,t+1} + u_{i,t+1} \end{aligned}$$

Where $E(\omega_{i,t+1}|\omega_{i,t}, \chi_{i,t+1}) = g(P_{i,t}, \phi_t - \beta_k k_{i,t})$ follows from the law of motion of productivity shocks and $P_{i,t}$ is the probability of survival of establishment i into the next period, i.e., $P_{i,t} = \Pr[\chi_{i,t+1} = 1]$. For estimation, first stage coefficients of labour are substituted, probability of survival is estimated from equation (21) and the function $g(.)$ is approximated using a higher order polynomial expansion in $P_{i,t}$ and $\phi_t - \beta_k k_{i,t}$ (see Rovigatti and Mollisi (2018))

Stage 2: Estimate the following equation using GMM estimation

$$y_{i,t+1} - \hat{\beta}_l l_{i,t+1} = \beta_0 + \beta_k k_{i,t+1} + g(\phi_{i,t} - \beta_k k_{i,t}) + \xi_{i,t+1} + u_{i,t+1}$$

For simplification, I use the `prodest` command in STATA to perform the LP estimation technique (Rovigatti & Mollisi, 2020). In Table 6, I compare the Levinsohn-Petrin (2003) method with other methods of estimating productivity. I obtain the largest estimation samples with OLS, Fixed Effects and Levinsohn-Petrin method followed by Olley-Pakes (OP) (1996) and Stochastic Frontier Approach. This is because I observe more establishments reporting intermediate costs than investments. Moreover, system GMM uses lags that limits the analysis to firms that are repeatedly observed. In the value added production function, labour input has the highest coefficient in all our specifications. When the frontier is defined as the 95th percentiles of the intra-industry productivity distribution, I see the central tendencies of CTF measures to be similar across TFP estimation methods. For LP, OLS and OP methods, mean CTF is 0.88 with similar standard deviations highlighting considerable overlap in their distributions. Therefore, I choose the LP estimator for its representativeness in our estimation sample, robustness to simultaneity in input choice, and use in the literature.

Table 6. Comparison of Total Factor Productivity Estimation methods

Dependent Variable	Levinsohn-Petrin (2003)	OLS	FE	Olley-Pakes (1996)	Stochastic Frontier Approach
Gross Value Added (in Log)	(1)	(2)	(3)	(4)	(5)
Labor Input	0.995***	0.991***	0.547***	0.968***	1.016***
	(0.000)	(0.005)	(0.012)	(0.000)	(0.005)
Capital Input	0.066***	0.062***	0.005**	0.104***	0.066***
	(0.000)	(0.003)	(0.002)	(0.003)	(0.003)
Control for Attrition	Yes	No	No	Yes	No
N (Establishments)	26888		26888	21152	26889
N	108699	108699	108699	67565	108703
TFP (Mean)	8.941	9.201	11.641	8.933	0.356
TFP (SD)	0.974	0.951	1.453	1.000	0.299
CTF (Mean)	0.882	0.882	0.834	0.883	0.748
CTF (SD)	0.078	0.076	0.098	0.078	0.146

Note: Standard errors in the parentheses. Standard errors obtained using 50 bootstrapped iterations. * if p<0.05, ** if p<0.01, *** if p<0.001.

2. Tables

Table 7. Effect of productivity catch up conditional on gross benefits of training

	Demand for Apprentices			
	(1)	(2)	(3)	(4)
CTF	-0.270**	-0.467***	-0.456	-1.014
	(0.087)	(0.138)	(0.351)	(0.601)
Gross Benefits	0.014***	0.017***	0.002	-0.017
	(0.002)	(0.004)	(0.022)	(0.036)
CTF# Gross Benefits			0.013	0.038
			(0.025)	(0.041)
First Stage Residual (η)		1.939		2.053
		(1.082)		(1.089)
N (Establishments)	18267	8438	18267	8438
N	67951	31184	67951	31184

Note: Benefits measured as benefits of training per apprentice per year in '000 Euros. Establishment controls include number of skilled workers (log), share of highly skilled workers, workforce turnover rate and share of part time workers. Other control variables include establishment size dummies, works council and (log) number of school graduates. All models control for centered values of time varying independent variables. Standard Errors in the parentheses. * if p<0.05, ** if p<0.01, *** if p<0.001.

Table 8. Effect of productivity catch up conditional on gross costs of training

	Demand for Apprentices			
	(1)	(2)	(3)	(4)
CTF	-0.293***	-0.507***	0.939**	2.823***
	(0.087)	(0.138)	(0.354)	(0.562)
Gross Costs	0.003	0.012**	0.052***	0.143***
	(0.002)	(0.004)	(0.014)	(0.022)
CTF# Gross Costs			-0.056***	-0.149***
			(0.016)	(0.024)
First Stage Residual (η)		1.790		3.120**
		(1.082)		(1.102)
N (Establishments)	18267	8438	18267	8438
N	67951	31184	67951	31184

Note: Costs measured as costs of training per apprentice per year in '000 Euros. Establishment controls include number of skilled workers (log), share of highly skilled workers, workforce turnover rate and share of part time workers. Other control variables include establishment size dummies, works council and (log) number of school graduates. All models control for centered values of time varying independent variables. Standard Errors in the parentheses. * if p<0.05, ** if p<0.01, *** if p<0.001.

Table 9. Effect of productivity catch up on demand for apprentices, by establishment size

Dependent Variable	Demand for Apprentices							
	Full Sample		Size<20		20>Size<200		Size>200	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CTF	-0.288***	-0.504***	0.972**	0.776	0.541***	0.859***	-0.558***	-0.998***
	(0.086)	(0.139)	(0.312)	(0.485)	(0.144)	(0.226)	(0.144)	(0.230)
$\overline{CTF}$	0.133	-0.014	-2.143***	-2.168**	-0.365	-1.004*	0.323	-0.616
	(0.175)	(0.295)	(0.414)	(0.672)	(0.244)	(0.395)	(0.351)	(0.622)
First Stage Residual (η)		1.883		4.655		0.291		4.024*
		(1.084)		(3.707)		(1.755)		(1.763)
Augmented Regression Test	163.47***	75.24***	24.58***	8.64***	23.66***	11.90***	32.00***	15.80***
All Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Convergence	Yes	Yes	No	No	Yes	Yes	Yes	Yes
N (Establishments)	18267	8438	8606	3979	8182	3856	2527	1040
N	67951	31184	28851	13520	30519	14112	8581	3552

Note: Establishment controls include number of skilled workers (log), share of hihgly skilled workers, workforce turnover rate and share of part time workers. Other control variables include establishment size dummies, works council and (log) number of school graduates. Standard Errors in the parentheses. * if p<0.05, ** if p<0.01, *** if p<0.001. Augmented Regression test is a test of joint significance of all the between effects in the regression where the null hypothesis represents absence of between effects.

Table 10. Effect of productivity catch up on labor demand and workforce composition

Dependent Variable	Overall Labour Demand		Low skilled Labour Demand		Medium skilled Labour Demand		High skilled labour demand	
	Employment Growth		Change in composition		Change in composition		Change in composition	
	FE	IV-FE	FE	IV-FE	FE	IV-FE	FE	IV-FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CTF	0.615***	0.809***	0.257***	0.265***	-0.102**	-0.140**	-0.177***	-0.132***
	(0.092)	(0.046)	(0.026)	(0.039)	(0.031)	(0.047)	(0.015)	(0.020)
N (Establishments)	18267	8438	14035	8049	14035	8049	14035	8049
N	67951	31184	53601	27647	53601	27647	53601	27647

Note: Controls include establishment size dummies, time dummies, works councils, collective labor agreement and (log) number of school graduates. All regressions control for establishment fixed effects. Standard Errors in the parentheses. * if p<0.05, ** if p<0.01, *** if p<0.001.

Table 11. Effect of productivity catch on demand for apprentices by establishment ownership, innovative ability and state of technology.

Dependent Variable	Demand for Apprentices													
	Full Sample		Single Plant		Multi Plant		Innovative Firms		Non Innovative Firms		Advanced Technology		Laggard Technology	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(7)	(8)	(7)	(8)	(7)	(8)
CTF	-0.288***	-0.504***	0.419**	0.231	-0.516***	-0.692***	-0.625***	-1.044***	0.009	-0.025	-0.447***	-0.768***	0.199	0.645*
	(0.086)	(0.139)	(0.147)	(0.234)	(0.116)	(0.183)	(0.146)	(0.252)	(0.124)	(0.185)	(0.102)	(0.161)	(0.184)	(0.315)
$\overline{CTF}$	0.133	-0.014	-0.953***	-0.897*	0.846**	0.661	0.520*	1.016*	-0.275	-0.626	0.130	0.058	-0.654*	-1.278*
	(0.175)	(0.295)	(0.244)	(0.401)	(0.263)	(0.447)	(0.265)	(0.475)	(0.212)	(0.335)	(0.203)	(0.334)	(0.291)	(0.496)
First Stage Residual (η)		1.883		3.150		2.257		4.529*		1.808		2.436		1.144
		(1.084)		(1.740)		(1.468)		(1.868)		(1.507)		(1.267)		(2.395)
Augmented Regression Test	163.47***	75.24***	88.36***	32.42***	57.91***	34.29***	57.03***	27.49***	101.10***	44.75***	119.81***	57.94***	52.64***	18.14***
All Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Convergence	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N (Establishments)	18267	8438	12789	5872	5478	2566	7636	3071	15521	7487	14100	6465	9129	4294
N	67951	31184	46932	21820	21019	9364	16582	6654	51369	24530	44666	20342	23198	10814

Note: Establishment controls include number of skilled workers (log), share of highly skilled workers, workforce turnover rate and share of part time workers. Other control variables include establishment size dummies, works council and (log) number of school graduates. Standard Errors in the parentheses. * if p<0.05, ** if p<0.01, *** if p<0.001. Augmented Regression test is a test of joint significance of all the between effects in the regression where the null hypothesis represents absence of between effects.

Table 12. Effect of productivity catch up on demand for apprentices controlling for firm's training characteristics

Dependent Variable	Demand for Apprentices							
	Full Sample		Vacancy Filling Ability		Vacancy Retention		Training Endowment	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CTF	-0.288***	-0.504***	-0.335***	-0.417**	-0.281**	-0.503***	-0.188*	0.422**
	(0.086)	(0.139)	(0.089)	(0.139)	(0.090)	(0.139)	(0.079)	(0.132)
CTF_BAR	0.133	-0.014	-0.003	-0.220	0.114	-0.015	0.297*	0.883***
	(0.175)	(0.295)	(0.146)	(0.231)	(0.177)	(0.285)	(0.143)	(0.239)
RES		1.883		1.163		2.073		1.139
		(1.084)		(1.117)		(1.099)		(1.052)
Unfilled Vacancy Rate			0.393***	0.456***				
			(0.021)	(0.031)				
Retention Rate					-0.054	-0.236***		
					(0.029)	(0.046)		
Apprentice Training Intensity							4.869***	4.688***
							(0.084)	(0.132)
Augmented Regression Test	163.47***	75.24***	66.57***	46.90***	81.94***	49.54***	334.26***	145.93***
All Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Convergence	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
N (Establishments)	18267	8438	10168	4715	9980	5462	18267	8438
N	67951	31184	29911	13247	45479	22237	67951	31184

Note: Establishment controls include number of skilled workers (log), share of highly skilled workers, workforce turnover rate and share of part time workers. Other control variables include establishment size dummies, works council and (log) number of school graduates. Standard Errors in the parentheses. * if p<0.05, ** if p<0.01, *** if p<0.001. Augmented Regression test is a test of joint significance of all the between effects in the regression where the null hypothesis represents absence of between effects.

Table 13. Effect of productivity convergence on further training

	Further Training			
	Full Sample	Size<20	20>Size<200	Size>200
	(1)	(2)	(3)	(4)
CTF	0.095**	0.126*	0.039	0.036
	(0.033)	(0.050)	(0.049)	(0.110)
N (Establishments)	17971	8543	8070	2387
N	65951	28472	29716	7763

Note: Establishment controls include number of skilled workers (log), share of highly skilled workers, workforce turnover rate and share of part time workers. Other control variables include establishment size dummies, works council and (log) number of school graduates. Linear fixed effects regression model used controlling for time-invariant unobserved heterogeneity at the establishment level. Standard Errors in the parentheses. * if p<0.05, ** if p<0.01, *** if p<0.001.

Table 14. Effect of productivity catch-up on demand for apprentices, by different sampling characteristics

Dependent Variable	Demand for Apprentices				
	Full Sample	Balanced Sample	Frontier (90th Percentile)	Frontier by Sector-Size	Poisson Model
	(1)	(2)	(3)	(4)	(5)
CTF	-0.288***	-0.417**	-0.279**	-0.261**	-0.303
	(0.086)	(0.157)	(0.084)	(0.082)	(0.248)
$\overline{CTF}$	0.133	0.627	-0.025	0.113	-0.448
	(0.175)	(0.515)	(0.172)	(0.171)	(0.293)
All Controls	Yes	Yes	Yes	Yes	Yes
N (Establishments)	18267	2375	18267	18267	18267
N	67951	19243	67951	67951	67951

Note: Establishment controls include number of skilled workers (log), share of highly skilled workers, workforce turnover rate and share of part time workers. Other control variables include establishment size dummies, works council and (log) number of school graduates. Standard Errors in the parentheses. * if p<0.05, ** if p<0.01, *** if p<0.001.

Table 15. Effect of productivity catch on demand for apprentices, by different TFP estimation methods

Dependent Variable	Levinsohn-Petrin (2003)	OLS	FE	Olley-Pakes (1996)	Labour Productivity	Stochastic Frontier Approach
Demand for Apprentices	(1)	(2)	(3)	(4)	(5)	(6)
CTF	-0.288***	-0.331***	0.164	-0.200*	-0.514***	-0.254***
	(0.086)	(0.090)	(0.123)	(0.086)	(0.102)	(0.055)
$\overline{CTF}$	0.133	0.645***	1.958***	0.043	1.029***	-0.123
	(0.175)	(0.178)	(0.222)	(0.176)	(0.194)	(0.093)
AIC	180044	180038	179932	180050	180020	180007
BIC	180409	180404	180297	180415	180385	180372
N (Establishments)	18267	18267	18267	18267	18267	18267
N	67951	67951	67951	67951	67951	67950

Note: Establishment controls include number of skilled workers (log), share of higgly skilled workers, workforce turnover rate and share of part time workers. Other control variables include establishment size dummies, works council and (log) number of school graduates. Standard Errors in the parentheses. * if p<0.05, ** if p<0.01, *** if p<0.001.