Firm Heterogeneity, Worker Training and Labor Productivity: The Role of Endogenous Self-selection

Sizhong Sun

College of Business Law and Governance
Division of Tropical Environment and Societies
James Cook University

Email: sizhong.sun@jcu.edu.au

[Abstract]

In productivity research with firm data, the existence of endogenous sample attrition is well known. In addition, endogenous sample selection may occur. In a simple model where heterogenous firms consider entry, exit, worker training and price setting in a monopolistically competitive market, I show that firm heterogeneity leads to self-selecting into the market, which in turn dampens the estimate of the marginal effect of worker training. To address the sample selection issue, I devise a generalized method of simulated moments estimator. Estimations with firm data from China's shoe manufacturing industry show that an increase of one standard deviation of worker training expenditure intensity results in an around 5.6% decrease in a firm's labor productivity, larger than the estimate without accounting for endogenous sample selection.

[Key Words]

Firm Heterogeneity, Worker Training, Labor Productivity, Sample Selection, Method of Simulated Moments, Cross-sectional Analysis

[JEL Classification]

D22, J24, L11, C15, C24

[Declarations]

Funding: This study does not receive funding from any sources;

Conflicts of interest/Competing interests: The author, Sizhong Sun, has no conflict of interest to declare;

Availability of data and material: Data are available upon request from the corresponding author;

Code availability: Codes are available upon request from the corresponding author;

Authors' contributions: Not applicable.

1. Introduction

Productivity arguably plays one of the most important roles in economics, and has been studied extensively. Researchers, for example, have tried to estimate productivity from firm level data. In doing so, researchers have recognized and addressed a number of issues. For example, since firms make optimal decisions on input usage and whether to exit the market, estimating the production function from firm level data, and subsequently productivity, is subject to the issues of input endogeneity and endogenous sample attrition (see, among others, Gandhi, Navarro, & Rivers, 2020; Olley & Pakes, 1996).

A less explored issue is that firms also make optimal decision on whether to enter the market. If firms are ex ante homogenous, as economic theories with representative firms will assume, the entry decision results in all firms having equal probability of market entry (i.e., appearing in the sample). Hence, being homogenous, firms' optimal entry decision does not result in an estimation bias.

In contrast, if firms are heterogeneous, their optimal decision of market entry results in different firms having different probabilities of appearing in the sample, leading to endogenous sample selection. Firm heterogeneity refers to the fact that different firms have different endowments of capability/productivity, and as such they behave differently in the market even if everything else is the same, in contrast to the representative firms.

Researchers have found firm heterogeneity to play an important role in the economic modelling of firm behavior. For example, among others, Helpman *et al.* (2004) discover a sorting pattern of serving domestic market, exporting and overseas investment. That is, the least productive firms do not enter the market, and the less productive firms only serve the domestic market, while the more productive firms export and/or invest overseas.

Owing to endogenous sample selection, firm heterogeneity presents an additional layer of challenge for empirical studies. Firms enter the market (and appear in the sample) only if they have a sufficiently high endowment of capability/productivity that results in a non-negative economic profit. Hence, the data that a researcher observes are no longer a random sampling of the underlying population of interest. As the sampling depends on firms themselves, it is necessary to address the issues of endogenous sample selection/incidental truncation. Note this situation cannot be addressed by the classical Heckman sample selection model (Heckman, 1979), as firms that do not enter the market are completely unobserved from the sample.

In this study, I will devise a generalized method of simulated moments (GMSM) estimator to address the endogenous sample selection issue. To do so, I focus on exploring the role of worker training in affecting a firm's average labor productivity, with firm level data from the shoe manufacturing industry in China.

Briefly, heterogenous firms are engaged in a four-stage game. At stage one, they make the decision on whether to enter the market, and at stage two, faced with realized shocks, they contemplate whether to exit or not. Should they decide to stay, they then make an optimal decision on worker training at stage three and set prices at stage four. Solving the model backward, one can derive a set of population moments, which guide the empirical estimations. The moment conditions are a function of the sample selection term, which is computed by the Monte Carlo simulations, resulting in a GMSM estimator. The theoretical modelling considers firms' entry and exit explicitly, addressing both the endogenous sample selection and attrition issues.

The contribution of this study is two-fold, in terms of methodological innovation, to a greater extent, and empirical application, to a lesser extent. Methodologically, I develop an analytical framework to address the endogenous sample selection bias, which can be applied in other similar studies. Regarding the empirical application, the paper explores the impact of worker training on firms' labor productivity, a topic that is of importance to both firms and policymakers. Besides, this study investigates the shoe manufacturing industry in China. In light of its fast economic growth in the past four decades, studies of Chinese manufacturing firms shall present implications to policymakers in other developing countries.

The rest of the paper is organized into five sections. Section 2 briefly surveys the related literature, focusing on those that are directly linked to this study. In Section 3, I construct an analytical framework, where an economic model is established to guide the empirical estimation with the GMSM estimator. Section 4 discusses the data, and the results are presented in Section 5. Section 6 concludes the study.

2. Related Literature

This study is related to two strands of research in the existing literature, one on econometric methods that address the sample selection/incidental truncation issues, and the other on the topic that examines the determinants of productivity, particularly labor productivity. Both strands of research have a substantial number of existing studies, and it is beyond the scope of this paper to comprehensively review them. Instead, I focus on previous studies that are most related to this paper, particularly those on the methods of dealing with sample selection, leaving readers to consult other articles for a broader survey of literature.

On the econometric methods that deal with the sample selection/incidental truncation issue, this strand of research is substantial.[1] Generally, these studies are not specifically

[1] For example, to name a few, Heckman (1979, 1990), Newey et al. (1990), Nawata and Nagase (1996), Kyriazidou (1997), Heckman et al. (1998), Vella (1998), Francesconi and Nicoletti (2006), Zimmer and Trivedi (2006), Madden (2008), Lee (2009), Newey (2009), Greene (2010), van Hasselt (2011), Vossmeyer (2016), and Semykina and Wooldridge (2018).

designed for a firm level study, although the methods can be applied in this context. Two existing studies are closely related to this paper, namely Bloom and Killingsworth (1985) and Nevo (2003).

Bloom and Killingsworth (1985) examine the estimation of a model where both the outcome and truncation equations are linear and the truncation variable is unobserved. Assuming error terms in the two equations are bivariate normally distributed, they develop a maximum likelihood estimator, where the identification relies on the correlation between two error terms. Nevo (2003) utilizes auxiliary information to address the sample selection issue. Specifically, he computes weights from the auxiliary information, which are proportional to the inverse probability of selection. Then he incorporates these weights into a generalized method of moments (GMM) estimator to correct for the sample selection bias.

Similar to Bloom and Killingsworth (1985), the truncation (sample selection) variable in this study is latent. However, unlike their study, the outcome and truncation equations are derived from economic modelling and are nonlinear. For the study of Nevo (2003), in its spirit, this paper also employs a GMM estimator to estimate the parameters of interest. Nevertheless, the approach that I take to address the sample selection bias is different from that of Nevo (2003). In this study, firms' profit maximization implies a structural relationship between the outcome and truncation equations. Utilizing such a structural relationship, rather than auxiliary information, I compute the sample selection term, which is an integral, in the moment conditions by the Monte Carlo simulations, resulting in a GMSM estimator.

Topic-wise, this study contributes to those that examine the determinants of labor productivity, which is also a substantial strand of research.[2] Recent studies on labor

[2] In addition, researchers also survey productivity studies from different dimensions, for example Syverson (2011) on why firms have different productivity levels, Ahmed and Bhatti (2020) on measurement and determinants of multi-factor productivity, and Del Gatto et al. (2011) and Van Beveren (2012) on estimating/measuring productivity/total factor productivity.

productivity include, among others, Zheng *et al.* (2017), Bjuggren (2018), Samargandi (2018), Dua and Garg (2019), He *et al.* (2019), LoPalo (2019), Tombe and Zhu (2019), Yao (2019), Amin and Okou (2020), Battisti et al. (2020), Hoffman and Burks (2020), Ku (2020), Pariboni and Tridico (2020), Zhang *et al.* (2020), Bárány and Siegel (2021), and Gregory-Smith (2021).

Four studies examine the impact of training on labor productivity, and hence are closely related to this paper. Using the Employment Opportunity Pilot Project Survey of Firms in the US in 1980 and 1982, Holzer (1990) finds that hours of training are positively related to productivity growth, but not its level. Using panel data of British industries from 1983 to 1996, Dearden *et al.* (2006) find that work-related training promotes labor productivity. Using data from 21 European countries from 1999 to 2005, Sala and Silva (2013) find that one additional hour of training per worker leads to an around 0.55% increase in labor productivity growth. In Shanghai, China, Ng (2005) observes that such training is more geared towards remedying skill deficiencies, rather than productivity enhancement, with survey data of 800 firms in the manufacturing sector in 2000.

This research differs from these four studies in that it is structural, namely more based on theoretical modelling, and as such makes a contribution to this strand of productivity research. In studying firm productivity, researchers have noted and addressed the endogenous sample attrition issue, due to firms' optimal decision of market exit (see for example, among others, Olley & Pakes, 1996). In addition, this study addresses endogenous sample selection, which arises from heterogenous firms' optimal decision on market entry.

To summarize, this study is linked to two strands of research, namely the productivity studies and econometric methods of correcting the sample selection bias. Within both strands, this study fills a void in the existing literature.

3. Analytical Framework

3.1 Economic Model

Firms are engaged in a four-stage game. At stage one, they decide whether to enter the market, which is monopolistically competitive. Should firms decide to enter, at stage two, they are faced with a shock in each period, which may force them to exit. If they decide to stay, they make a decision on worker training at stage three, and then set prices for their products at stage four.

Firms have complete information except that, in each period, they do not know shocks in the future. However, they have knowledge of their distribution and form expectations accordingly. In particular, at stage one they conjecture the probability of market exit in each period. Besides, as will be made clear later, firms' information set is not completely known to researchers. As such, it is necessary to integrate out unobserved information in the empirical estimations.

In the market, the demand function is $q = \Phi p^{1/(\rho-1)}$, where q and p denote quantity and price of goods, respectively; Φ is aggregate demand, equal to consumer income divided by an aggregate price index, which firms take as given as they are small relative to the market; and ρ is the constant-elasticity-of-substitution (CES) preference parameter ($0 < \rho < 1$).

Upon entry into the market, firms draw a capability endowment $\lambda \in R_+$ from an exogenous distribution (i.i.d. over firms) with a cumulative distribution function (CDF) of $G(\lambda)$. Hence, with probability one, different firms will have different capability draws, and in this sense, firms are heterogenous, in contrast to representative firms.

Production involves a fixed cost (f), for example setting up production facilities such as the assembly line. Then using the production facilities, firms employ one worker, combined with m units of intermediate inputs, to produce s units of outputs. The term s measures the average labour productivity, which depends continuously on the firm endowment index (unobserved to researchers), worker training (τ, the ratio of worker training expenditure against total sales revenue), and an index (h) that captures all the other factors, as follows:

$$s = s(\lambda, \tau, x, \zeta) = \lambda e^{\alpha\tau} h(x, \zeta) \tag{1}$$

where the index h is a continuous function of x (observed firm characteristics) and ζ (factor observed to firms, but not to researchers).

Firms' marginal cost of production is then $c = \frac{w+m}{s}$, where c represents the marginal cost and w is the wage rate. Firms' profit (π) can be written as $\pi = [(1-\tau)p - c]\Phi p^{1/(\rho-1)} - f$. To maximize their profit, at stage four, firms set prices as a markup over their marginal cost of production, namely $p = c/[\rho(1-\tau)]$. Therefore, firms' optimal per-period profit is as follows:

$$\pi^* = (1-\rho)\rho^{\frac{\rho}{1-\rho}}\Phi(w+m)^{\frac{\rho}{\rho-1}}h^{\frac{\rho}{1-\rho}}e^{\frac{\alpha\rho}{1-\rho}\tau}\lambda^{\frac{\rho}{1-\rho}}(1-\tau)^{\frac{1}{1-\rho}} - f$$

where the superscript * denotes optimal profit.

At stage three, there are two regimes for worker training. One is that firms need to train workers to remedy a skill deficiency.[3] In such a case, after training, workers are required to take up new tasks with which they are not familiar, resulting in a decrease of the average

[3] For example, firms may need to train their workers for a particular skill which is needed in the production process but not available from the labor market.

labor productivity,[4] namely $\alpha \leq 0$. Therefore, firms' optimal choice of τ is the minimum level of training that is necessary for remedying the skill deficiency, denoted as $\underline{\tau}$. That is $\tau^* = \underline{\tau}$ where the superscript * denotes optimal training intensity. If $\underline{\tau} = 0$, then a firm is not faced with the skill deficiency problem.

In the second regime, firms use training to improve the worker's skill, in which case $\alpha \geq 0$. Firms' optimal training intensity (τ) is as follows:

$$\tau^* = \begin{cases} \frac{\alpha\rho - 1}{\alpha\rho}, \alpha > \frac{1}{\rho} \\ 0, 0 \leq \alpha \leq \frac{1}{\rho} \end{cases}.$$

Note if the marginal benefit of worker training is small ($0 \leq \alpha \leq \frac{1}{\rho}$), it is not worthwhile for firms to invest in worker training ($\tau^* = 0$). Firms, but not the researchers, observe which regime they are in.

At stage two, in each period, firms are subject to shocks, either from the production side (e.g., as those in Gandhi et al., 2020; Olley & Pakes, 1996) or the demand side (e.g., as in De Loecker, 2011). Responding to the shocks, firms make a decision on whether to exit the market. If they exit, they can collect a liquidation value of ϕ. In contrast, the value of stay is $V = \pi^* + \eta max\{\phi, E[V']\}$ where η is the discount rate; V' is the next period's stay value; and the expectation is taken over the next period's shocks and transition of states. Hence, firms' optimal decision on whether to exit (χ_2) is $\chi_2 = 1(V < \phi)$, where $1(\cdot)$ is the indicator function. Note firms' optimal decision on whether to exit the market results in endogenous sample attrition in the empirical estimations.

[4] Note for individual workers who receive training, their skill sets will expand. However, it can occur that the firm's average labor productivity decreases.

At stage one, in contemplating whether to enter the market, firms form beliefs of the probability of market exit in each period, denoted as $1 - \delta_t$ where the subscript t represents time. Hence the expected per-period profit is $\delta_t \pi_t^*$, where I append the subscript t to π^* to make it clear that optimal per-period profit can vary across time. Then firms' optimal decision on market entry (χ_1) is as follows:

$$\chi_1 = 1\left(\sum_{t\geq 0} \eta^t \left(\prod_{k=0}^{t-1} \delta_k\right) [\delta_t \pi_t^* + (1-\delta_t)\phi] \geq 0\right) = 1\left(\frac{1}{1-\eta\delta}[\delta\pi^* + (1-\delta)\phi] \geq 0\right)$$
$$= 1(\pi^* \geq 0)$$

where the second equality is obtained by imposing a steady-state restriction ($\delta_t = \delta$ and $\pi_t^* = \pi^*$, $t \geq 0$) and the third equality imposes a normalization that $\phi = 0$.

The optimal per-period profit (π^*) is a monotone increasing function of the capability endowment (λ) and $\pi^*(0) = -f < 0$, where the argument of π^* is the lower bound of the capability endowment. Hence, there exists a threshold (cut-off) capability endowment ($\underline{\lambda}$) below which firms will earn a negative profit (and a negative entry value). The cut-off capability endowment can be derived as follows:

$$\underline{\lambda} = (1-\rho)^{-\frac{1-\rho}{\rho}} \rho^{-1} \Phi^{-\frac{1-\rho}{\rho}} (w+m) h^{-1} e^{-\alpha\tau} (1-\tau)^{-\frac{1}{\rho}} f^{\frac{1-\rho}{\rho}} \tag{2}$$

Firms' optimal revenue is as follows:

$$r^* = \rho^{\frac{\rho}{1-\rho}} \Phi (w+m)^{\frac{\rho}{\rho-1}} s^{\frac{\rho}{1-\rho}} (1-\tau)^{\frac{\rho}{1-\rho}} = \rho^{\frac{\rho}{1-\rho}} \Phi (w+m)^{\frac{\rho}{\rho-1}} h^{\frac{\rho}{1-\rho}} e^{\frac{\alpha\rho}{1-\rho}\tau} (1-\tau)^{\frac{\rho}{1-\rho}} \lambda^{\frac{\rho}{1-\rho}} \tag{3}$$

where r^* represents optimal revenue.

The equilibrium is characterised as a distribution of optimal prices that firms charge, where firms have a capability endowment above the cut-off level and they have correct beliefs ($\delta = 1 - E[\chi_2]$). Hence, heterogenous firms self-select into the market. They are observed in the data only if their capability endowment is above the cut-off level. That is, the firm data are not a random sampling of the underlying population, in contrast to the prediction of theories with representative firms. The self-selection issue cannot be addressed by the classical Heckman sample selection model, as firms with a capability endowment below the cut-off level are unobserved completely from the sample. Proposition 1 summarizes the implication of the self-selection issue for empirical estimation of firm labor productivity.

Proposition 1: Suppose firm-level data are generated by profit maximizing firms in a monopolistically competitive market, as described by the four-stage game, and the quantity of interest is labor productivity (s). Then, (1) using firm-level data to estimate s as a function of τ and x is subject to an endogenous sample selection bias; and (2) the sample selection bias dampens the estimate of marginal impacts of τ and x.

Proof: By the definition of s, $lns = \alpha\tau + ln[h(x,\zeta)] + ln(\lambda)$. Since λ is a firm's capability endowment with CDF $G(\lambda)$, we have the following conditional population moment, $E[lns|x,\tau] = \alpha\tau + H(x) + \int_0^\infty ln\lambda dG(\lambda)$ where $H(x) = \int ln[h(x,\zeta)]dG_\zeta(\zeta)$ and $G_\zeta(\zeta)$ is the CDF of ζ. In contrast, using the observed firm-level data for estimation, the conditional population moment is as follows:

$$E[lns|x,\tau,\chi_1 = 1, \chi_2 = 1\,] = E\left[lns\middle|x,\tau,\lambda \geq \underline{\lambda}\,\right] = \alpha\tau + H(x) + \int_{\underline{\lambda}}^\infty ln\lambda \frac{1}{1-G(\underline{\lambda})} dG(\lambda) \quad (4)$$

where $\underline{\lambda}$ is a function of τ and x. Comparing the two conditional population moments, it is clear that the estimation is subject to an endogenous sample selection bias, due to the presence of the term $\underline{\lambda}$.

Differentiating the population moment with respect to x, we obtain $\frac{\partial E[lns|x,\tau,\lambda\geq\underline{\lambda}]}{\partial x} = \frac{\partial H}{\partial x} + \frac{\partial\underline{\lambda}}{\partial x}\frac{g(\underline{\lambda})}{[1-G(\underline{\lambda})]^2}\int_{\underline{\lambda}}^{\infty}(ln\lambda - ln\underline{\lambda})dG(\lambda)$, where $g(\underline{\lambda})$ is the probability density function. Similarly, $\frac{\partial E[lns|x,\tau,\lambda\geq\underline{\lambda}]}{\partial \tau} = \alpha + \frac{\partial\underline{\lambda}}{\partial \tau}\frac{g(\underline{\lambda})}{[1-G(\underline{\lambda})]^2}\int_{\underline{\lambda}}^{\infty}(ln\lambda - ln\underline{\lambda})dG(\lambda)$. From Equation (2), if $\frac{\partial H}{\partial x} > 0$ ($\alpha > 0$), then $\frac{\partial\underline{\lambda}}{\partial x} < 0$ ($\frac{\partial\underline{\lambda}}{\partial \tau} < 0$), and vice versa. So, the sample selection bias dampens the estimate of marginal impacts.

□

Remarks: A number of remarks are due here. First, in the theoretical modelling, firms make optimal decisions on entry, exit, worker training and prices. In addition, the theoretical framework can be extended to additional layers of decision making, and adapted for other purposes. For example, after entry and exit decisions, firms can engage in a two-stage game, at stage one deciding R&D and stage two price. Proposition 1 continues to hold with such kind of extension. Second, the decision making is almost static in the sense that I focus on the steady state. Accounting for firm dynamics more elaborately is a direction for future research. Third, firms' self-selection into the market results in an incidental truncation of the data. Profit maximization leads to a structural relationship between optimal profit (and subsequently value of entry), which determines the truncation of data, and the quantity of interest (s). I will utilize this structural relationship in estimations, with a method of simulated moments (MSM) estimator.

3.2 Econometric Method

The quantity of interest is average labor productivity (Equation 1). For the purpose of estimation, I further parameterize Equation (1) as follows:

$$lns = \alpha\tau + ln[h(x,\zeta)] + ln\lambda = \alpha\tau + x\beta + \zeta + ln\lambda \quad (5)$$

where x is a $(K-1)\times 1$ vector of observed factors, which will be specified later, and ζ captures all factors that are unobserved to researchers and has mean zero.

With Equation (5), the conditional population moment (Equation 4) can be re-written as $E\left[lns \middle| x,\tau,\lambda \geq \underline{\lambda}\right] = \alpha\tau + x\beta + \int_{\underline{\lambda}}^{\infty} ln\lambda \frac{1}{1-G(\underline{\lambda})} dG(\lambda)$.[5] Subsequently, the unconditional population moments are as follows:

$$E\left[\left(lns - \alpha\tau - x\beta - \int_{\underline{\lambda}}^{\infty} ln\lambda \frac{1}{1-G(\underline{\lambda})} dG(\lambda)\right) z \middle| \lambda \geq \underline{\lambda}\right] = E\left[\left(\frac{1-\rho}{\rho} lnr - ln\rho - \frac{1-\rho}{\rho} ln\Phi + ln(w+m) - \ln(1-\tau) - \alpha\tau - x\beta - \int_{\underline{\lambda}}^{\infty} ln\lambda \frac{1}{1-G(\underline{\lambda})} dG(\lambda)\right) z \middle| \lambda \geq \underline{\lambda}\right] = E\left[\left(\frac{1-\rho}{\rho} lnr - \frac{1-\rho}{\rho} lnI + ln(w+m) - \ln(1-\tau) - x\beta - \int_{\underline{\lambda}}^{\infty} ln\lambda \frac{1}{1-G(\underline{\lambda})} dG(\lambda)\right) z \middle| \lambda \geq \underline{\lambda}\right] = 0 \tag{6}$$

where z is an $L\times 1$ vector of instruments that can include x and τ; in the first equality, I use Equation (3) to substitute lns and for notational convenience I omit the superscript (*) of r; and in the second equality, as aggregate demand equals consumer income divided by an aggregate price index, I parameterize it as $ln\Phi = \alpha_0 + \alpha_1 t + lnI$ where t denotes time and I is consumer income. The terms $-\alpha\tau$, $-ln\rho - \frac{1-\rho}{\rho}\alpha_0$ and $-\frac{\rho}{1-\rho}\alpha_1 t$ are absorbed into $x\beta$.[6]

The first step of the estimations is to estimate the CES preference parameter (ρ) by using the relationship between revenue and total variable cost implied by firm profit maximization, using a sample of non-exporting firms that do not have working training, as follows:

---

[5] The expectation is taken over both ζ and λ. Besides, note that in taking expectation with respect to ζ, it is possible that x and τ are endogenous, namely $E\left[\zeta \middle| x,\tau,\lambda \geq \underline{\lambda}\right] = \int \zeta dG_{\zeta}(\zeta|x,\tau)$ is a function of x and τ. For τ, I later also use excluded instruments and report the results in Section 5.3. For x, assuming $\int \zeta dG_{\zeta}(\zeta|x,\tau)$ is a linear function of x, it is then absorbed into the term $x\beta$.

[6] Note in Equation (6), as I focus on the steady state, by conditioning on $\lambda \geq \underline{\lambda}$, one also captures endogenous sample attrition. Future research can account for sample attrition (firm dynamics) more elaborately while addressing the sample selection issue.

$$TVC = \rho r + \epsilon \tag{7}$$

where *TVC* denotes total variable cost and ϵ is an error term that captures measurement errors. This identification strategy has been used in previous studies, for example Aw *et al.* (2011).

With an estimate of ρ, I then proceed to estimate β. Let $g_i = \int_{\underline{\lambda}_i}^{\infty} ln\lambda \frac{1}{1-G(\underline{\lambda}_i)} dG(\lambda)$, $y_i = \frac{1-\rho}{\rho} lnr_i - \frac{1-\rho}{\rho} lnI_i + ln(w_i + m_i) - \ln(1 - \tau_i) - g_i$, and $Q = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i\beta)z_i$,[7] where the subscript *i* indexes observations and *n* denotes the number of observations. Should researchers observe the sample selection term $g \equiv (g_1 \quad \cdots \quad g_n)'$, the generalized method of moments (GMM) estimator of β is $\hat{\beta} = \underset{\beta}{\operatorname{argmin}}\, Q'Q = (X'ZZ'X)^{-1}X'ZZ'Y$, where X is the $n \times K$ matrix of explanatory variables, namely $X \equiv (x_1' \cdots x_n')'$; Z is the $n \times L$ matrix of instruments, namely $Z \equiv (z_1' \cdots z_n')'$; and Y is the $n \times 1$ vector of the dependent variable, namely $Y \equiv \left(y_1 \cdots y_n\right)'$. Note it contains the integral $\int_{\underline{\lambda}_i}^{\infty} ln\lambda \frac{1}{1-G(\underline{\lambda}_i)} dG(\lambda)$, which in turn depends on data and β. For the GMM estimator to be feasible, I utilize the Monte Carlo simulation to compute the integral in the estimation process, which leads to the following generalized method of simulated moments (GMSM) estimator:

$$\hat{\beta} = (X'ZZ'X)^{-1}X'ZZ'\hat{Y} \tag{8}$$

where $\hat{Y}$ denotes the estimate of Y with $\hat{y}_i = \frac{1-\rho}{\rho} lnr_i - \frac{1-\rho}{\rho} lnI_i + ln(w_i + m_i) - \ln(1 - \tau_i) - \hat{g}_i$, $i = 1, \dots, n$. In more detail, I use the following algorithm[8] to implement the GMSM estimator:

The Algorithm:

[7] Note x and z are the same as in Equation (6).
[8] A similar algorithm has been used by Sun and Anwar (2022) to estimate product quality.

Step 1: To draw $J = 10{,}000$ random numbers for $ln\lambda$ and ζ from a standard normal distribution.[9]

Step 2: With $\hat{\beta}^{v-1}$, where $\hat{\beta}^0 = 0$ and the superscript v denotes the iteration loop, for each observation i, compute $g_i = \int_{\underline{\lambda}_i}^{\infty} ln\lambda \frac{1}{1-G(\underline{\lambda}_i)} dG(\lambda)$ by Monte Carlo simulations, namely $\hat{g}_i = \frac{\sum_{j=1}^{J} ln\tilde{\lambda}_j 1(ln\tilde{\lambda}_j \geq ln\underline{\lambda}_i^{v-1})}{\sum_{j=1}^{J} 1(ln\tilde{\lambda}_j \geq ln\underline{\lambda}_i^{v-1})}$ where $ln\tilde{\lambda}$ is the random draws and $\underline{\lambda}_i^{v-1}$ is the cut-off capability (Equation 2) evaluated at $\hat{\beta}^{v-1}$. With the simulated $\hat{g}$, then solve Equation (8) for $\hat{\beta}^v$.

Step 3: If $\left\|\hat{\beta}^v - \hat{\beta}^{v-1}\right\| < lt$, where lt is a pre-defined tolerance level, at which the algorithm stops. Otherwise, go to Step 2.

By the Law of Large Numbers, $\operatorname*{plim}_{J\to\infty} \hat{g}_i = g_i$, and consequently $\operatorname*{plim}_{J\to\infty} \hat{Y} = Y$ and $\operatorname*{plim}_{J,n\to\infty} \hat{\beta} = \beta$. With Equation (8), the variance-covariance matrix of $\hat{\beta}$ can be written as $E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right] = (X'ZZ'X)^{-1}X'ZZ'E[uu']ZZ'X(X'ZZ'X)^{-1}$, where u is an $n \times 1$ vector with $u_i = \zeta_i + ln\lambda_i - \hat{g}_i$, $i = 1, \dots, n$. In order to estimate $E[uu']$, I first predict the residual $\hat{\zeta}_i + ln\hat{\lambda}_i$ and compute $\hat{u}_i = \hat{\zeta}_i + ln\hat{\lambda}_i - \hat{g}_i$, where $\hat{g}_i$ is obtained from the simulation. Note $\hat{u}_i$ contains the effect of simulations. Then, the estimate of $E[uu']$ is an $n \times n$ matrix where the diagonal elements are $\hat{u}_i^2$, and the off-diagonal elements are $\hat{u}_i\hat{u}_j$ if observations i and j are from the same firm in adjacent years and 0 otherwise, $i, j = 1, \dots, n$. Therefore, the standard errors reported later are robust to arbitrary heteroskedasticity and first-order autocorrelation.

[9] Here $ln\lambda$ and ζ can be combined into one term. I use the normal distribution, given its prevalence. One can also use other distributions. A frequently used distribution is the Pareto distribution. However, one disadvantage of using the Pareto distribution is that the sample selection effect cancels out the direct effect of x and τ, namely $E\left[lns\middle|x, \tau, \lambda \geq \underline{\lambda}\right]$ does not depend on x and τ, which suggests that the Pareto distribution assumption is likely to be too strong for empirical estimations.

Later in estimations, I gradually increase the number of control variables in *x*. Such a step-wise building-up approach allows me to examine the sensitivity of the coefficient estimate of worker training expenditure intensity (τ). For the instruments (z), I first assume *x* and τ are exogenous (namely *z* is *x* and τ), and then allow τ to be endogenous (namely *z* is *x* and two excluded instruments that will be discussed in Section 4).

4. Data and Variables

The data are sourced from the National Bureau of Statistics (NBS), China, and contain 1,493 firms in 2005, 1,672 firms in 2006, and 1,850 firms in 2007 in the shoe manufacturing industry (including both textile fabric and leather shoes). Monetary values in the dataset are deflated by the producer price index, obtained from *China Statistical Yearbook 2008* (base year: 2005).

Firms report their expenditure on worker training and education, from which I calculate the expenditure intensity (τ), namely the ratio of expenditure against total sales revenue. For labor productivity, as shown by Equation (6), it consists of two components, one that is computable from data (denoted as $\widehat{lns}$, the computed component) and the other that depends on parameters to be estimated. That is, $lns = \left\{\frac{1-\rho}{\rho} lnr - \frac{1-\rho}{\rho} ln\mathrm{I} + ln(w+m) - \ln(1-\tau)\right\} + \left\{-ln\rho - \frac{1-\rho}{\rho}\alpha_0 - \frac{\rho}{1-\rho}\alpha_1 t\right\} = \widehat{lns} + \left\{-ln\rho - \frac{1-\rho}{\rho}\alpha_0 - \frac{\rho}{1-\rho}\alpha_1 t\right\}$, where r is domestic sales revenue, I is average consumer income, $(w+m)$ equals total variable costs divided by the number of workers, and τ is the training intensity. I compute $\widehat{lns}$ and report in Table 1. Total variable costs consist of total wage payment and intermediate inputs. I utilize Equation (7) to estimate ρ, and find a point estimate of 0.7697 with a standard error of 0.0019. The point estimate implies a substitution elasticity of 4.34, approximately in line with estimates in previous studies (see for example Aw et al., 2011; Das, Roberts, & Tybout, 2007).[10]

In the estimations later, I control for the other factors that potentially affect firm labor productivity (the vector x in the theoretical model), which are chosen both following previous studies and based on the conceptual link between the variables and labor productivity. The control variables (x) include time trend, firm size, capital intensity, R&D intensity, ownership structure, whether a firm exports, whether a firm is foreign-invested, age, FDI presence, unemployment insurance intensity, pension and health insurance intensity, housing provident fund and subsidies, and the industry worker training intensity. In addition, I also control for a supply-side factor in the labor market, namely the government's educational expense (as a share of GDP).

It is not surprising that firm labor productivity can exhibit a time trend. In addition, the time trend also captures supply-side factors in the labor market, for example it is highly correlated with the number of graduates from technical schools (correlation > 0.99). Workers are found to be more productive in larger firms (Idson & Oi, 1999), suggesting firm size plays a role. I measure firm size by the fixed assets annual net average (in natural logarithm). Similarly, capital intensity is likely to affect labor productivity, which is measured as the capital stock per worker (in natural logarithm). R&D is another important determinant of labor productivity (see for example Baumann & Kritikos, 2016; Raymond, Mairesse, Mohnen, & Palm, 2015), and is measured as the share of R&D spending in a firm's total sales revenue (%).

Ownership structure, whether a firm exports, and whether a firm is foreign-invested are three firm characteristics that can affect labor productivity, which are dummy variables

[10] For example, with a sample of 3703 firms in Taiwanese electronics industry in the years 2000 and 2002-2004, Aw et al. Aw et al. (2011) estimate an ρ of 0.8432 in the domestic market.

that take a value of one if a firm is privately owned, exports, and is foreign-invested, respectively. Firm age captures a firm's experience in operation, and FDI presence controls for possible productivity spillover from FDI. A large number of previous studies have found FDI productivity spillovers (for example, among others, Anwar & Sun, 2014; Sun, 2011; Xu & Sheng, 2012). Firm age is measured as the number of years since the business started, and FDI presence is the share of employment of foreign-invested firms in the industry.

The unemployment insurance intensity, pension and health insurance intensity, and housing provident fund and subsidies intend to control for the incentives firms provide to their workers, and are the shares of expenditure in a firm's total sales revenue (%). To control for the spillover effect of worker training, I include the industry worker training intensity, namely the ratio of total industry worker training expenditure against total industry sales (%). The government's educational expense is likely to affect supply of workers in the labor market, and hence is included as a control.

In implementing the GMSM estimator, data of fixed cost of production are needed. To compute the fixed cost, I first construct the capital stock. That is, I compute firms' investment in fixed assets as the difference of the fixed assets' original value, reported in the dataset, from its previous year and deflate it by the fixed asset investment price index obtained from the *China Statistical Yearbook 2008*. Assuming the deflated fixed assets' original value of 2005 as the capital stock of 2005, I then calculate the capital stocks of 2006 and 2007 as the sum of depreciated capital stock of the previous year (depreciation rate of 5%) and the investment. The fixed cost[11] is calculated as the real interest rate times the capital stock, where the real interest rate is the lending interest rate minus the inflation rate.

[11] Alternatively, one can draw the fixed cost from a given distribution.

Later in our estimations, the instruments are x and τ themselves. In addition, as a robustness check, I also use alternative instruments for the worker training expenditure intensity (τ), namely the average worker training expenditure intensity by other firms in the same industry-province (excluding the firm itself) and the share of firms in the same industry-province that have non-zero spending on worker training (also excluding the firm itself). On the one hand, these two instruments are correlated with the worker training expenditure intensity, while on the other hand, they shall not affect a firm's labor productivity directly as firms are small relative to the market and do not behave strategically in a monopolistically competitive market. In addition, the potential spillover effect from other firms' worker training is controlled by the industry worker training intensity.

Table 1 reports the summary statistics of these variables. The sample exhibits substantial variations. For example, the standard deviation of labor productivity is around 20% of its mean. The variation in firm size is similar to that of labor productivity, with a standard deviation around 20% of its mean. Around 31% firm-year observations report positive spending on worker training and education. Nevertheless, firms appear to allocate a small proportion of their sales revenue to worker training and expenditure, with an average intensity of only 0.05%. The R&D intensity, unemployment insurance, pension and health insurance, and housing fund and subsidies exhibit a similar pattern.

A majority of firms are privately owned (83%), and a sizeable proportion of firms export and are foreign-invested (31 and 28%, respectively). Firms are around 7 years old on average, where 391 firms are newly established, accounting for 7.8% of total sample size. The presence of FDI-invested firms is substantial, accounting for around 62.7% of total industry employment. A high level of FDI presence can result in fierce competition with domestic firms, which in turn may harm domestic firms in the short run. The shoe manufacturing industry is a labor-intensive industry, where the negative minimum value of capital intensity is due to natural logarithm. Interestingly, its standard deviation is nearly 50% of its mean, suggesting a wider spread than those of labor productivity and firm size.

Table 1. Summary Statistics

Variable	Mean	S. D.	Min	Max
Labor productivity, computed component ($\widehat{lns}$)	4.8763	0.9910	1.0088	9.2849
Worker training (%)	0.0467	0.2698	0	12.3922
Firm size	7.8584	1.4928	1.5495	13.9373
Capital intensity	2.1168	1.0789	-3.6925	6.2106
R&D intensity (%)	0.0652	0.3796	0	9.0620
Age	7.2431	4.9876	1	53
FDI	0.6272	0.0426	0.5216	0.6769
Unemployment insurance (%)	0.1730	0.5401	0	12.0528
Pension and health insurance (%)	0.6250	1.4870	0	61.7084
Housing fund and subsidies (%)	0.1602	0.5749	0	12.6995
Industry training intensity (%)	0.0468	0.0063	0.0243	0.0539
Education expenditure (%)	4.4970	0.0192	4.4813	4.5263
Instrument 1	0.0475	0.0928	0	2.6069
Instrument 2	0.3065	0.2617	0	1
Fixed cost (ln*f*)	4.0856	1.4764	-2.0931	9.8828
Ownership	0.8305			
Whether export	0.3087			
Whether foreign-invested	0.2774			

Note: N = 5,015. Instrument 1 is the average worker training expenditure intensity by other firms in the same industry-province (excluding the firm itself). Instrument 2 is the share of firms in the same industry-province that have non-zero spending on worker training (excluding the firm itself).
Source: NBS, China.

5. Results

In order to estimate the impact of worker training on firm labor productivity in the shoe manufacturing industry, I implement a GMSM estimator by utilizing the analytical framework in Section 3. In particular, I use a step-wise approach, starting with a minimum specification where only worker training expenditure intensity and constant are included in estimation, and then gradually building up the specification by including additional control variables. Such a building-up approach allows one to check the stability of the parameter estimate, and should the estimate of marginal impact of worker training be robust, one shall not observe substantial variations in different specifications. For the purpose of examining the effect of sample selection on the parameter estimate, I also conduct GMM estimations without considering the sample selection issue. Table 2 reports the estimation results, and Table 3 presents the results of the GMSM estimations.

5.1 The effect of sample self-selection

Comparing Table 2 with Table 3, one can observe differences in the point estimates. For example, in column [1] of Table 2, the point estimate of the coefficient of worker training expenditure intensity is -0.1426, with a robust standard error of 0.0736 (statistically significant at the 5% level). In contrast, the point estimate of the same specification in Table 3 (column [1]), where sample selection is accounted for, is -0.3014, with a robust standard error of 0.0383, which is also statistically significant at the 5% level. The point estimate in column [1] of Table 2 is outside four standard errors of the point estimate in column [1] of Table 3, while the latter is outside two standard errors of the former estimate, suggesting significant differences in the two estimations.

Proposition 1 predicts that if a factor negatively affects labor productivity, only more capable firms can enter the market (and appear in the sample), as less capable firms will suffer an economic loss if they enter the market. Consequently, with the self-selected sample, the estimated marginal impact of the factor is biased towards zero, namely less negative. A similar situation occurs if the factor positively affects labor productivity. Therefore, such kind of sample self-selection tends to dampen the estimate of the true marginal effect. The estimation results in Tables 2 and 3 are largely consistent with this prediction. As a result, the interpretation of results will be based on estimates in Table 3.

Table 2. Estimation Results with Self-selected Sample

Variables		[1]	[2]	[3]	[4]	[5]
constant	β_0	4.8829	4.5816	3.4587	3.3274	20.4584
		(0.0144)	(0.0370)	(0.3348)	(0.3239)	(5.3672)
worker training	β_1	-0.1426	-0.1455	-0.1616	-0.1288	-0.1295
		(0.0736)	(0.0712)	(0.0854)	(0.0697)	(0.0705)
time trend	β_2		0.1456	0.1285	0.1560	0.0793
			(0.0170)	(0.0238)	(0.0232)	(0.0375)
firm size	β_3			0.1481	0.1557	0.1572
				(0.0130)	(0.0125)	(0.0125)

capital intensity	β_4			0.0939	0.0828	0.0822
				(0.0173)	(0.0167)	(0.0168)
R&D intensity	β_5			0.0250	0.0202	0.0212
				(0.0419)	(0.0382)	(0.0385)
ownership	β_6			0.0030	0.0002	-0.0008
				(0.0357)	(0.0351)	(0.0350)
whether export	β_7			-0.3771	-0.4073	-0.4104
				(0.0316)	(0.0309)	(0.0309)
whether foreign-invested	β_8			-0.1718	-0.0967	-0.0962
				(0.0335)	(0.0348)	(0.0348)
age	β_9			-0.0071	-0.0031	-0.0030
				(0.0033)	(0.0031)	(0.0031)
FDI	β_{10}			0.0128	0.1775	-0.0379
				(0.4650)	(0.4491)	(0.5352)
unemployment insurance	β_{11}				-0.1094	-0.1068
					(0.0602)	(0.0602)
pension and health insurance	β_{12}				-0.1153	-0.1168
					(0.0441)	(0.0446)
housing fund and subsidies	β_{13}				-0.0876	-0.0989
					(0.0335)	(0.0343)
industry training intensity	β_{14}					-0.7650
						(2.5507)
education expenditure	β_{15}					-3.7375
						(1.1621)
Observation	N	5,015	5,015	5,015	5,015	5,015

Note: Robust standard errors in brackets.
Source: The author's estimation with data from NBS, China.

5.2 Worker training and education

The parameter of interest is the coefficient estimate of worker training expenditure intensity. In all five specifications (columns [1]-[5] in Table 3), the point estimates are negative and statistically significant at the 5% level, with the estimates ranging between -0.3014 and -0.2060. Hence, the estimates appear to be robust to different specifications.

Using the point estimate in the full specification (column [5] in Table 3), a one standard deviation increase (0.27%) in the worker training expenditure intensity results in an around 5.6% decrease in a firm’s labor productivity. The statistically significant and negative estimate suggests that firms are in the regime of remedying skill deficiencies (the first regime in the theoretical model). Consistent with the observation by Ng (2005), such kind of worker training and education intends to remedy skill deficiencies, rather than improve workers' skill. Workers trained with new skills are likely to be assigned with new tasks, which in turn reduces firms' labor productivity since they are unfamiliar with the new tasks and their new skills may not be utilized efficiently. As a result, an increase in worker training and education intensity leads to deterioration of labor productivity.

Table 3. Estimation Results after Accounting for Sample Selection Issue

Variables		[1]	[2]	[3]	[4]	[5]
constant	β_0	4.4102	3.9643	3.3646	3.1811	13.9794
		(0.0123)	(0.0256)	(0.2916)	(0.2847)	(2.6167)
worker training	β_1	-0.3014	-0.2986	-0.2854	-0.2063	-0.2060
		(0.0383)	(0.0381)	(0.0387)	(0.0333)	(0.0335)
time trend	β_2		0.2134	0.2072	0.2652	0.2208
			(0.0116)	(0.0186)	(0.0182)	(0.0241)
firm size	β_3			0.1138	0.1270	0.1279
				(0.0111)	(0.0105)	(0.0105)
capital intensity	β_4			-0.0669	-0.0811	-0.0813
				(0.0143)	(0.0138)	(0.0139)
R&D intensity	β_5			0.0454	0.0446	0.0456
				(0.0323)	(0.0283)	(0.0285)
ownership	β_6			0.0043	-0.0015	-0.0024
				(0.0311)	(0.0300)	(0.0299)
whether export	β_7			-0.8322	-0.8616	-0.8636
				(0.0267)	(0.0258)	(0.0258)
whether foreign-invested	β_8			-0.3010	-0.1391	-0.1372
				(0.0292)	(0.0281)	(0.0281)
age	β_9			-0.0102	-0.0032	-0.0032
				(0.0028)	(0.0025)	(0.0025)
FDI	β_{10}			0.3540	0.5598	0.4924
				(0.4023)	(0.3914)	(0.4047)
unemployment insurance	β_{11}				-0.1530	-0.1505
					(0.0234)	(0.0233)
pension and health insurance	β_{12}				-0.3390	-0.3422
					(0.0125)	(0.0124)
housing provident fund and subsidies	β_{13}				0.0364	0.0305
					(0.0213)	(0.0209)
industry training intensity	β_{14}					-1.0026
						(1.9067)
education expenditure	β_{15}					-2.3615
						(0.5524)
Observation	N	5015	5015	5015	5009	5009

Note: Robust standard errors in brackets.
Source: The author's estimation with data from NBS, China.

The significantly negative effect of worker training and education on labor productivity points towards a potential firm level resource misallocation. Conceptually, one expects that the labor market shall provide sufficient skills that firms can employ, and if not, the market will adjust its equilibrium wage so as to induce more supply of the skills, and in addition, policymakers can implement appropriate policies to further enhance the supply. For example, if skills of operating a specialized machine in the shoe manufacturing industry are in demand, the increased wages due to high demand of the skill encourage workers to acquire the skills. Furthermore, policy makers can incentivize technical schools to train more graduates on this aspect.

Compared with firms themselves, it is more efficient for the labor market and policymakers to address the skill deficiency problem, due to the advantage of specialization. In this sense, it is a resource misallocation that firms address the skill deficiency problem through their own worker training and education. Given the significant negative impact on labor productivity, it is appropriate for policymakers to discourage worker training and education that remedies skill deficiency. Instead, should there be a skill deficiency problem, a well-functioning labor market will address the issue by adjusting the equilibrium wage rate, and as such policymakers shall aim to establish a strong institution that help alleviate/remove the labor market distortions.

As for other control variables,[12] the coefficient of the time trend is positive and statistically significant at the 5% level, suggesting improvement of labor productivity over time. Bigger firms appear to have a higher level of labor productivity, possibly due to economies of scale. In contrast, capital intensity is not associated with a higher level of labor

[12] Note, as discussed in Footnote 5, estimates of the coefficients of x can pick up the effect of x being endogenous. That is, while factors in x directly affect lns, they can also indirectly affect lns through ζ.

productivity, which can occur as the shoe manufacturing industry is traditionally more labor intensive. The estimate of the coefficient of R&D intensity is not statistically significant at the 5% level, despite the positive point estimate. Privately owned firms and state and collectively owned firms do not appear to have different labor productivity, as the coefficient estimate of ownership structure is statistically insignificant at the 5% level. Similarly, firm age and FDI presence also do not significantly affect firm labor productivity.

Surprisingly, exporting experience dampens a firm's labor productivity. This occurs possibly because the shoe manufacturing industry is a traditional exporting industry, for which China has comparative advantage, and as such learning by exporting is not important for firms. Compared with domestic firms, foreign-invested firms have lower labor productivity. Firms' welfare payment significantly affects their labor productivity. Unemployment insurance dampens labor productivity, which is not surprising as it reduces workers' incentive to work hard. The pension and health insurance similarly discourages firms' labor productivity, while the effect of housing provident fund and subsidies is not statistically significant at the 5% level. The industry training expenditure intensity appears not to have a statistically significant effect, suggesting little spillover effect from worker training by other firms. The government's education expenditure, a supply-side factor, significantly reduces firms' labor productivity. An increase in education expenditure leads to less labor supply in the market, which in turn reduces the incentive for workers to work hard.

5.3 Robustness

In Tables 2 and 3, I step-wisely build up the model specification in the estimations, where the coefficient estimates of worker training expenditure intensity exhibit only small variations across different specifications. Therefore, the estimate is robust to different specifications of the empirical model.

Table 4. Alternative Estimation Results

Variables		[1]	[2]	[3]	[4]	[5]
constant	β_0	4.4488	3.9601	3.5582	3.2026	2.4857
		(0.0301)	(0.0258)	(0.3946)	(0.3928)	(0.9222)
worker training	β_1	-1.3811	-0.3129	-0.2861	-0.2072	-0.2009
		(0.6476)	(0.0379)	(0.0391)	(0.0333)	(0.0332)
time trend	β_2		0.2153	0.1966	0.2639	0.3469
			(0.0118)	(0.0238)	(0.0238)	(0.1135)
firm size	β_3			0.1147	0.1277	0.1297
				(0.0111)	(0.0106)	(0.0122)
capital intensity	β_4			-0.0679	-0.0815	-0.0840
				(0.0144)	(0.0139)	(0.0147)
R&D intensity	β_5			0.0452	0.0442	0.0464
				(0.0323)	(0.0283)	(0.0287)
ownership	β_6			0.0050	0.0000	-0.0105
				(0.0314)	(0.0302)	(0.0314)
whether export	β_7			-0.8341	-0.8621	-0.8594
				(0.0268)	(0.0258)	(0.0260)
whether foreign-invested	β_8			-0.3007	-0.1401	-0.1344
				(0.0293)	(0.0281)	(0.0284)
age	β_9			-0.0102	-0.0031	-0.0038
				(0.0028)	(0.0025)	(0.0027)
FDI	β_{10}			0.0734	0.5194	2.7669
				(0.5393)	(0.5354)	(3.2657)
unemployment insurance	β_{11}				-0.1535	-0.1493
					(0.0234)	(0.0242)
pension and health insurance	β_{12}				-0.3377	-0.3395
					(0.0122)	(0.0122)
housing provident fund and subsidies	β_{13}				0.0357	0.0276
					(0.0211)	(0.0253)
industry training intensity	β_{14}					-18.4276
						(30.2471)
Observation	N	5010	5015	5015	5008	5008

Note: Robust standard errors in brackets. Education expense not included due to multicollinearity.
Source: The author's estimation with data from NBS, China.

Table 4 reports the estimation results, where I use the average worker training expenditure intensity by other firms in the same industry-province (excluding the firm itself) and the share of firms in the same industry-province that have non-zero spending on worker training (also excluding the firm itself) to instrument the worker training expenditure intensity. It can be observed that the coefficient estimate of worker training expenditure intensity remains negative, despite the magnitude of the point estimate exhibits changes.

Hence, the estimate of a negative coefficient of worker training expenditure intensity is largely robust to alternative instruments.

6. Concluding Remarks

The paper explores endogenous sample selection in firm-level productivity research, by investigating whether firms' expenditure on worker training and education promotes their labor productivity. For this purpose, I first set up a simple economic model where heterogenous firms are engaged in a four-stage game, namely to make decisions on market entry, exit, worker training and prices. Firms' optimal decision on market exit results in endogenous sample attrition. In addition, firms' optimal decision on market entry, together with firm heterogeneity, results in firms endogenously self-selecting into the market (and the sample). I show that such kind of self-selection dampens the estimate of the marginal impact of worker training expenditure intensity, namely the coefficient estimate is biased towards zero, if not appropriately addressed.

To address the endogenous sample selection issue, I design a GMSM estimator, where I utilize the structural relationship among different variables of interest, implied by firms' profit maximization, to simulate the sample selection term by the Monte Carlo method. The analytical framework established in this paper can be used for firm level analysis of other purposes.

Applying the GMSM estimator with firm level data from China's shoe manufacturing industry from 2005 to 2007, I find that firms' worker training expenditure intensity generates a significantly negative effect on their labor productivity, which is larger than the estimate without accounting for endogenous sample selection. The negative effect occurs because such training and education is more geared towards remedying skill deficiencies, rather than enhancing labor productivity. What is worse, after the training, workers are likely to be assigned with new tasks, which subsequently reduces their productivity due to unfamiliarity with both the new tasks and newly acquired skills.

The finding of a significantly negative impact points towards a potential resource misallocation issue. Compared with firms, the labor market and policymakers are in a better position to remedy the skill deficiency problem, for example through wage rate adjustment and implementation of appropriate incentive policies in the market. Hence, policymakers shall discourage worker training and education that aims at remedying skill deficiencies, and instead remove market distortions and promote a well-functioning labor market.

References:

Ahmed, T., & Bhatti, A. A. (2020). MEASUREMENT AND DETERMINANTS OF MULTI-FACTOR PRODUCTIVITY: A SURVEY OF LITERATURE. *Journal of Economic Surveys, 34*(2), 293-319. doi:https://doi.org/10.1111/joes.12360

Amin, M., & Okou, C. (2020). Casting a shadow: Productivity of formal firms and informality. *Review of Development Economics, 24*(4), 1610-1630. doi:https://doi.org/10.1111/rode.12697

Anwar, S., & Sun, S. (2014). Heterogeneity and curvilinearity of FDI-related productivity spillovers in China's manufacturing sector. *Economic Modelling, 41*, 23-32. doi:https://doi.org/10.1016/j.econmod.2014.03.021

Aw, B. Y., Roberts, M. J., & Xu, D. Y. (2011). R&D Investment, Exporting, and Productivity Dynamics. *American Economic Review, 101*(4), 1312-1344. doi:doi: 10.1257/aer.101.4.1312

Bárány, Z. L., & Siegel, C. (2021). Engines of sectoral labor productivity growth. *Review of Economic Dynamics, 39*, 304-343. doi:https://doi.org/10.1016/j.red.2020.07.007

Battisti, M., Belloc, F., & Del Gatto, M. (2020). Labor productivity and firm-level TFP with technology-specific production functions. *Review of Economic Dynamics, 35*, 283-300. doi:https://doi.org/10.1016/j.red.2019.07.003

Baumann, J., & Kritikos, A. S. (2016). The link between R&D, innovation and productivity: Are micro firms different? *Research Policy, 45*(6), 1263-1274. doi:https://doi.org/10.1016/j.respol.2016.03.008

Bjuggren, C. M. (2018). Employment protection and labor productivity. *Journal of Public Economics, 157*, 138-157. doi:https://doi.org/10.1016/j.jpubeco.2017.11.007

Bloom, D. E., & Killingsworth, M. R. (1985). Correcting for truncation bias caused by a latent truncation variable. *Journal of Econometrics, 27*(1), 131-135. doi:https://doi.org/10.1016/0304-4076(85)90048-X

Das, S., Roberts, M. J., & Tybout, J. R. (2007). Market Entry Costs, Producer Heterogeneity, and Export Dynamics. *Econometrica, 75*(3), 837-873. doi:10.1111/j.1468-0262.2007.00769.x

De Loecker, J. (2011). Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity. *Econometrica, 79*(5), 1407-1451. doi:https://doi.org/10.3982/ECTA7617

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. doi:https://doi.org/10.1111/j.1468-0084.2006.00170.x

Del Gatto, M., Di Liberto, A., & Petraglia, C. (2011). MEASURING PRODUCTIVITY. *Journal of Economic Surveys, 25*(5), 952-1008. doi:https://doi.org/10.1111/j.1467-6419.2009.00620.x

Dua, P., & Garg, N. K. (2019). Determinants of labour productivity: Comparison between developing and developed countries of Asia-Pacific. *Pacific Economic Review, 24*(5), 686-704. doi:https://doi.org/10.1111/1468-0106.12294

Francesconi, M., & Nicoletti, C. (2006). Intergenerational mobility and sample selection in short panels. *Journal of Applied Econometrics, 21*(8), 1265-1293. doi:https://doi.org/10.1002/jae.910

Gandhi, A., Navarro, S., & Rivers, D. A. (2020). On the Identification of Gross Output Production Functions. *Journal of Political Economy, 128*(8), 2973-3016. doi:10.1086/707736

Greene, W. (2010). A stochastic frontier model with correction for sample selection. *Journal of Productivity Analysis, 34*(1), 15-24. doi:10.1007/s11123-009-0159-1

Gregory-Smith, I. (2021). WAGES AND LABOR PRODUCTIVITY: EVIDENCE FROM INJURIES IN THE NATIONAL FOOTBALL LEAGUE. *Economic Inquiry, 59*(2), 829-847. doi:https://doi.org/10.1111/ecin.12960

He, J., Liu, H., & Salvo, A. (2019). Severe Air Pollution and Labor Productivity: Evidence from Industrial Towns in China. *American Economic Journal: Applied Economics, 11*(1), 173-201. doi:10.1257/app.20170286

Heckman, J. (1979). Sample Selection Bias as a Specification Error. *Econometrica, 47*(1), 153-161. doi:10.2307/1912352

Heckman, J. (1990). Selection Bias and Self-selection. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *The New Palgrave: A Dictionary of Economics*. London: Palgrave Macmillan.

Heckman, J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing Selection Bias Using Experimental Data. *Econometrica, 66*(5), 1017-1098. doi:10.2307/2999630

Helpman, E., Melitz, M. J., & Yeaple, S. R. (2004). Export Versus FDI with Heterogeneous Firms. *American Economic Review, 94*(1), 300-316. doi:10.1257/000282804322970814

Hoffman, M., & Burks, S. V. (2020). Worker overconfidence: Field evidence and implications for employee turnover and firm profits. *Quantitative Economics, 11*(1), 315-348. doi:https://doi.org/10.3982/QE834

Holzer, H. J. (1990). The Determinants of Employee Productivity and Earnings. *Industrial Relations: A Journal of Economy and Society, 29*(3), 403-422. doi:https://doi.org/10.1111/j.1468-232X.1990.tb00761.x

Idson, T. L., & Oi, W. Y. (1999). Workers Are More Productive in Large Firms. *The American Economic Review, 89*(2), 104-108. Retrieved from http://www.jstor.org/stable/117089

Ku, H. (2020). Does Minimum Wage Increase Labor Productivity? Evidence from Piece Rate Workers. *IZA Discussion Paper*(13369). Retrieved from Available at SSRN: https://ssrn.com/abstract=3628244

Kyriazidou, E. (1997). Estimation of a Panel Data Sample Selection Model. *Econometrica, 65*(6), 1335-1364. doi:10.2307/2171739

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies, 76*(3), 1071-1102. doi:10.1111/j.1467-937X.2009.00536.x

LoPalo, M. C. (2019). *Essays on the Determinants of Worker Productivity and Labor Market Outcomes.* (PhD). University of Texas at Austin, Austin.

Madden, D. (2008). Sample selection versus two-part models revisited: The case of female smoking and drinking. *Journal of Health Economics, 27*(2), 300-307. doi:https://doi.org/10.1016/j.jhealeco.2007.07.001

Nawata, K., & Nagase, N. (1996). Estimation of sample selection bias models. *Econometric Reviews, 15*(4), 387-400. doi:10.1080/07474939608800363

Nevo, A. (2003). Using Weights to Adjust for Sample Selection When Auxiliary Information Is Available. *Journal of Business & Economic Statistics, 21*(1), 43-52. doi:10.1198/073500102288618748

Newey, W. K. (2009). Two-step series estimation of sample selection models. *The Econometrics Journal, 12*(suppl_1), S217-S229. doi:10.1111/j.1368-423X.2008.00263.x

Newey, W. K., Powell, J. L., & Walker, J. R. (1990). Semiparametric Estimation of Selection Models: Some Empirical Results. *The American Economic Review, 80*(2), 324-328. Retrieved from http://www.jstor.org/stable/2006593

Ng, Y. C. (2005). Training determinants and productivity impact of training in China: a case of Shanghai. *Economics of Education Review, 24*(3), 275-295. doi:https://doi.org/10.1016/j.econedurev.2004.05.005

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

Pariboni, R., & Tridico, P. (2020). Structural change, institutions and the dynamics of labor productivity in Europe. *Journal of Evolutionary Economics, 30*(5), 1275-1300. doi:10.1007/s00191-019-00641-y

Raymond, W., Mairesse, J., Mohnen, P., & Palm, F. (2015). Dynamic models of R & D, innovation and productivity: Panel data evidence for Dutch and French manufacturing. *European Economic Review, 78*, 285-306. doi:https://doi.org/10.1016/j.euroecorev.2015.06.002

Sala, H., & Silva, J. I. (2013). Labor productivity and vocational training: evidence from Europe. *Journal of Productivity Analysis, 40*(1), 31-41. doi:10.1007/s11123-012-0304-0

Samargandi, N. (2018). Determinants of Labor Productivity in MENA Countries. *Emerging Markets Finance and Trade, 54*(5), 1063-1081. doi:10.1080/1540496X.2017.1418658

Semykina, A., & Wooldridge, J. M. (2018). Binary response panel data models with sample selection and self-selection. *Journal of Applied Econometrics, 33*(2), 179-197. doi:https://doi.org/10.1002/jae.2592

Sun, S. (2011). Foreign Direct Investment and Technology Spillovers in China's Manufacturing Sector. *The Chinese Economy, 44*(2), 25-42. doi:10.2753/1097-1475440202

Sun, S., & Anwar, S. (2022). Estimation of product quality in China's food processing and manufacturing industries. *Economic Modelling, 107*, 105681. doi:https://doi.org/10.1016/j.econmod.2021.105681

Syverson, C. (2011). What Determines Productivity? *Journal of Economic Literature, 49*(2), 326-365. doi:10.1257/jel.49.2.326

Tombe, T., & Zhu, X. (2019). Trade, Migration, and Productivity: A Quantitative Analysis of China. *American Economic Review, 109*(5), 1843-1872. doi:10.1257/aer.20150811

Van Beveren, I. (2012). TOTAL FACTOR PRODUCTIVITY ESTIMATION: A PRACTICAL REVIEW. *Journal of Economic Surveys, 26*(1), 98-128. doi:https://doi.org/10.1111/j.1467-6419.2010.00631.x

van Hasselt, M. (2011). Bayesian inference in a sample selection model. *Journal of Econometrics, 165*(2), 221-232. doi:https://doi.org/10.1016/j.jeconom.2011.08.003

Vella, F. (1998). Estimating Models with Sample Selection Bias: A Survey. *The Journal of Human Resources, 33*(1), 127-169. doi:10.2307/146317

Vossmeyer, A. (2016). Sample Selection and Treatment Effect Estimation of Lender of Last Resort Policies. *Journal of Business & Economic Statistics, 34*(2), 197-212. doi:10.1080/07350015.2015.1024837

Xu, X., & Sheng, Y. (2012). Productivity Spillovers from Foreign Direct Investment: Firm-Level Evidence from China. *World Development, 40*(1), 62-74. doi:https://doi.org/10.1016/j.worlddev.2011.05.006

Yao, Y. (2019). Does higher education expansion enhance productivity? *Journal of Macroeconomics, 59*, 169-194. doi:https://doi.org/10.1016/j.jmacro.2018.11.009

Zhang, J., Mishra, A. K., Zhu, P., & Li, X. (2020). Land rental market and agricultural labor productivity in rural China: A mediation analysis. *World Development, 135*, 105089. doi:https://doi.org/10.1016/j.worlddev.2020.105089

Zheng, L., Batuo, M. E., & Shepherd, D. (2017). The Impact of Regional and Institutional Factors on Labor Productive Performance—Evidence from the Township and Village Enterprise Sector in China. *World Development, 96*, 591-598. doi:https://doi.org/10.1016/j.worlddev.2017.04.006

Zimmer, D. M., & Trivedi, P. K. (2006). Using Trivariate Copulas to Model Sample Selection and Treatment Effects. *Journal of Business & Economic Statistics, 24*(1), 63-76. doi:10.1198/073500105000000153