

Digitalization, Wage Inequality, and Employment*

Yi Lee and Chu Ping Lo[†]

This paper presents a simple model demonstrating how digitalization drives inter-industry resource reallocation, shifting from low digital-intensive industries to high digital-intensive industries. While digitalization substitutes workers, it also lowers productivity thresholds, enabling less productive firms to overcome entry barriers, thereby increasing sectoral employment — a phenomenon more pronounced in high digital-intensive sectors. Consequently, digitalization might raise employment in high digital-intensive sectors without contributing to sectoral wage inequality. Empirical testing of Taiwan’s city-level data from 2001 to 2019 supports our predictions, showing that both digitalization and digital infrastructure are associated with increased firm turnover, higher employment, elevated wages, and reduced wage inequality.

Key Words: Digitalization; Digital intensity; Trade.

JEL Classification Numbers: F12, F13.

1. INTRODUCTION

Rapidly advancing digital technologies, such as the Internet and Artificial Intelligence (AI), have significantly increased enterprise efficiency (e.g., Bajari et al., 2015; Reavie, 2018), as they replace routine tasks with greater precision and cognitive task modality (Georgieff and Hye, 2021). Digitalization can improve production efficiency and help enterprises break through trade barriers by reducing transaction and information costs. For instance, some empirical analyses suggest that the Internet, as a data channel, improves trade (Freund and Weinhold, 2002; Clarke and Wallsten, 2006; Liu and Nath, 2013). This creates new opportunities to participate in more domestic and international markets.

* Acknowledges: This research is supported by the Ministry of Science and Technology, Taiwan (NSC 111-2410-H-002-158-MY2).

[†] Lee: Department of Economics, National Tsing Hua University, Hsinchu, Taiwan. Email: ylee@mx.nthu.edu.tw; Lo: Corresponding author. Department of Agricultural Economics, National Taiwan University, Taipei, Taiwan. E-mail: cplo@ntu.edu.tw. Acknowledgment: This research is supported by the Ministry of Science and Technology, Taiwan (111-2410-H-002-158-MY2).

Another worldwide perception is that greater exposure to digitalization is linked to higher unemployment level (Franken and Wattenberg, 2019). Thus, governments may be hesitant to invest in digitalization, given that it displaces human labor, potentially leading to massive job losses and worsening wage inequality. To the best of our knowledge, the literature has rarely addressed the labor-related impacts of digitalization. To fill this gap, the primary purpose of this investigation is to provide evidence as to whether and how digitalization affects labor and wage inequality. However, this paper has no ambition to cover all digital technologies such as AI, but instead targets digitalization's impact, especially the development of a country's digital infrastructure on sectoral wage inequality and employment.

To illustrate a positive feedback loop between data and production, we develop a simple model by integrating digitalization into the frameworks of Melitz (2003) and Helpman et al. (2010). These prominent models assume that firms are differentiated within a representative industry. In contrast, this paper diverges from their models by allowing for differences in digital intensity across industries. In this model, countries possess unique digital infrastructure, industries vary in digital intensity, firms differ in productivity, and production factors encompass not only labor but also data factors.

This paper suggests that greater digital (infrastructure) development within a country enables its firms to access and utilize more data, generating positive production externalities. Here, digitalization refers to the extent of data utilization in production. This increased data, when integrated as a production factor, contributes to boosting output in the associated sectors. We argue that, digitalization lowers productivity thresholds for new entrants, increasing sectoral employment without exacerbating sectoral wage inequality. Furthermore, the effect of digitalization on sectoral employment is more pronounced when firms are in high digital-intensive sectors than in low digital-intensive sectors. As a result, a country's improvement in digitalization helps drive inter-industry resource reallocation from low digital-intensive industries to high digital-intensive industries.

The model's predictions mentioned above warrants empirical testing. Given constraints associated with global country-level digital data, we opt for city-level data from Taiwan. We consider Taiwan as a distinctive and finely-tuned "global" economy, where its diverse cities can be likened to small economic entities engaging in free trade among themselves. Given the presumptions that countries have distinct digitization infrastructure, industries differ in digital intensity, and firms exhibit varying levels of productivity, using Taiwan's city-level data for hypothesis testing is justifiable for two primary reasons. Firstly, digital infrastructure (e.g., computer usage) have increased over time. However, these changes are not uniform

across all Taiwanese cities, resulting in varying digital infrastructure between them. Secondly, Taiwan encompasses six municipalities and thirteen counties, each with its unique industry development focus, spanning from high to low digital-intensive sectors.¹

Our empirical results align with our model's predictions, showing that cities with high levels of digital development have increased labor employment and higher labor wages compared to cities with lower digital development. Additionally, highly digitized cities tend to have higher firm turnover rates, while advanced digitalization is associated with lower levels of wage inequality.

The rest of this paper proceeds as follows. Section 2 sets up the model. Section 3 shows the equilibrium. Section 4 addresses how digitalization impacts sectoral wage inequality. Section 5 focuses on sectoral employment. Section 6 presents the empirical implements. An alternative measure of digitalization is adopted as the robust test in Section 7 and Section 8 concludes.

2. THE MODEL

In a world of characterized by diverse countries and industries, production factors consist of not only labor but also data. Labor is heterogeneous in its abilities, particularly in its digital ability. While labor is dedicated to production, it also plays a role in transforming information from production activities into another production factor: data. The rate of data digitization varies by country and by industry. It is positively influenced by the digital infrastructure available in the firm's location and is largely dependent on the average ability of the workforce in the associated industry. Therefore, firms in an industry employ and compensate their workforce based not only on firm-specific productivity but also on the sectoral average ability of workers in the industry.

Suppose that consumers in each country have identical preferences. The aggregate demand for a representative differentiated good within a sector is expressed as a constant elasticity substitution function, reflecting the

¹For example, the financial services and high-tech industries are mainly concentrated in the northern region of Taiwan, with various service industries predominating in Taipei City. Computer, electronics, and optical product manufacturing cluster in New Taipei City, while the semiconductor industry centers around Hsinchu City. Electronic component manufacturing is prominent in Taoyuan City, water transportation in Keelung City, and chemical materials manufacturing in Yilan County. The central region of Taiwan traditionally hosts the machinery industry (e.g., Taichung City) and the petrochemical industry (e.g., Yunlin County). The southern region is a hub for heavy industries, with Kaohsiung City specializing in petrochemicals and steel production. Tainan City focuses on the optoelectronic and semiconductor industries, while Pingtung County specializes in food and feed manufacturing.

demand for a range of horizontally differentiated varieties within the sector:

$$Q = \left[\int_{j \in J} q(j)^\beta dj \right]^{1/\beta},$$

where $0 < \beta < 1$ is the elasticity of substitution parameter, j indexes varieties, J denotes the set of varieties within the sector and $q(j)$ is consumption of variety j . By providing only one of a continuum of varieties, each firm in a country takes the aggregate sectoral demand (E) and aggregate price index (P) as given. Hence, the firm's equilibrium revenue is:

$$r(j) = p(j)q(j) = Bq(j)^\beta, \quad (1)$$

where $B = E^{1-\beta} P^\beta$.

Similar to the Melitz (2003) model, a potential producer pays an entry cost of $f_e > 0$ units of the numeraire to draw its productivity from a distribution. Axtell (2001) studied the size distribution of U.S. firms and found that it follows a Pareto distribution, as did Luttmer (2007) when he further discussed firm dynamics and the selection process. Although our focus is on firm size, the underlying productivity distribution can be inferred, as firm size is often correlated with productivity. Thus, we specify firm's productivity distribution as the Pareto distribution $G(\theta) = 1 - (\theta_{\min}/\theta)^z$, for $\theta \geq \theta_{\min} \geq 1$ and $z > 1$, where θ_{\min} is minimum productivity threshold for a firm to survive. The differentiated good from the associated industry is produced by heterogenous firms, which find a suitable workforce with search frictions.

We argue that raw data form a production externality that can serve as a production factor when it has been constructed into effective data.² The production function of a representative heterogeneous firm in country i is given by:

$$y_i(\theta) = \theta l_i^{1-\mu} d_i^\mu, \quad 1 > \mu \geq 0, \quad (2)$$

where l_i is the size of the workforce employed by a heterogenous firm, d_i denotes an effective data (information) input of the firm, and μ is the share of data in production in country i that indicates the digital intensity of the industry. Effective data is essential for optimizing processes, reducing waste, and enhancing productivity through the optimization of supply chains, production schedules, and resource allocation. Demand equals supply in the equilibrium, as implied in equations (1) and (2), we obtain $y_i = q_i, \forall i$.

Being a production externality, the volume of raw data should positively correlate with a firm's output. On the other hand, with better digital

²Some literature perceive data as a production factor (e.g., Varian, 2019).

infrastructure as a data conduit, firms in that country tend to have a higher digitization rate, transforming raw data into effective constructed data. To construct and transfer raw data into production factors (effective data) for use in production, firms must first collect, clean, manage, mine, maintain, and analyze these raw data. In the digital era, relevant industrial-level data provide the diverse and comprehensive datasets necessary to develop accurate and robust prediction models. For instance, to refine artificial intelligence algorithms, firms typically expose them to a wide range of scenarios, improving their ability to generalize and perform well in real-world applications. Thus, a firm needs not only its own data but also industrial-level data to fully realize its potential in accuracy, innovation, operational efficiency, and competitiveness.

Consequently, a firm in an industry can benefit not only from its own production externality but also from the production externality of its associated industry, while data from other industries are considered irrelevant to its production. The above argument suggests that a firm's capability for data factor generation highly depends on the average digital ability of labor in its associated industry. The higher the average ability of labor in an industry, the more relevant industry-level data is available for firms' usage in the industry. Conclusively, the effective data factor of a firm increases with its firm-specific raw data, the industry-specific average ability of labor in generating data, and the country-specific development of digital infrastructure.

To present the heterogenous digital ability among labor, we follow Helpman et al. (2010) and assume that a continuum of workers exists in a country who independently draw their ability from a Pareto distribution $G_a(a) = 1 - (a_{\min}/a)^\lambda$ for $a \geq a_{\min} > 0$, where $\lambda > 1$ is a shape parameter that governs the distribution of labor ability and a_{\min} represents the minimum (digital) ability of labor in the country. For simplicity, we normalize the minimum (digital) ability of labor to unity.

Suppose that there exists an arbitrary screening cost in hiring talent. It is then reasonable to assume that labor in an industry μ faces an industry-specific screening threshold $a_c(\mu)$ during hiring, and $a_c(\mu) \geq a_{\min}$, where a_{\min} is the minimum ability of workers. Here, the μ represents the level of digitalization within an industry, and its associated effects on workforce's ability is denoted by $a'_c(\mu) > 0$. From the ability distribution mentioned above, we can then obtain an average digital ability in an industry, denoted by $\bar{a}_c(\mu) = a_c(\mu)\lambda/(\lambda - 1)$.

Acemoglu and Restrepo (2018) argue that digitalization and automation increase the demand for workers with advanced technical skills, leading to higher wages and productivity thresholds in those sectors. Their argument suggests that digitalization compels firms to raise their screening thresholds

to hire workers with greater digital competencies. We will elaborate on this in more detail later.

Building on the insights of Acemoglu and Restrepo (2018), it is reasonable to argue that a firm's effective data generation is positively related not only to its available raw data but also to its in-house digital talent. Therefore, a firm's data generation function can be expressed as:

$$d_i = \phi_i y_i(\theta) \bar{a}_c(\mu)^\gamma, \quad (3)$$

where ϕ_i represents a country's development of digital infrastructure that will be explained in more detail later. The measure of raw data is positively associated with the equilibrium output of a firm with productivity θ , and for simplicity, this measure is assumed to be the output. Here, we regard labor ability as equivalent to digital capability, so that $\bar{a}_c(\mu)$ also represents the average ability of labor in data construction in industry μ of country i . For simplicity, we drop the country specific subscript i to save notation hereafter.

Following Helpman et al. (2010), with a Pareto distribution of worker ability, a firm in industry μ chooses its industry-specific screening threshold $a_c(\mu)$ hires a measure $l(\phi, \mu) = n(a_{\min}/a_c(\mu))^\lambda$ of workers with average ability $\bar{a}_c(\mu)$, where n is a measure of potential workers a firm with productivity θ will hire in industry μ .³

In equation (3), a country's development of digital infrastructure (e.g., internet, broadband, AI) is supposed to have a positive impact on its digitization rate for all firms in the country. In general, better digital infrastructure enables a country's firms to generate and leverage more data to fine-tune product features and substitute more labor in production. Thus, we presume $\phi > 1$ to catch the idea that a final good generally requires at least one stage of processing, and each stage of processing generates its corresponding data. In other words, multiple processing activities result in a multiplicative increase in relevant data.

Given equations (2) and (3), we obtain the production function of a firm with productivity θ in industry μ in a country with a digital development level ϕ as:

$$y = \bar{\lambda} \theta^{\frac{1}{1-\mu}} \phi^{\frac{\mu}{1-\mu}} \bar{a}_c(\mu)^{\frac{\mu\gamma}{1-\mu} - \lambda} n, \quad (4)$$

where $\bar{\lambda} = [a_{\min} \lambda / (\lambda - 1)]^\lambda$ is a constant term. To highlight the importance of the dataset scale in production, which is reflected by the digital ability of

³Given the Pareto distribution of worker ability as $G_a(a) = 1 - (a_{\min}/a)^\lambda$ and an industry-specific screening threshold $a_c(\mu)$, we can calculate the average ability of workers a firm in the industry will hire is as $\bar{a}_c(\mu) = \int_{a_c}^{\infty} dG_a(a) = \int_{a_c}^{\infty} -\lambda a^{-\lambda-1} da = (a_{\min}/a_c(\mu))^\lambda$. With the measure of potential workers n for this firm, the firm's employment can be expressed as $n(a_{\min}/a_c)^\lambda$.

the workforce, we presume a minimum digital-intensity threshold $\mu = \frac{\lambda}{\gamma+\lambda}$ and assume that all industries have a greater digital-intensity as $\mu > \underline{\mu}$. This ensures that, ceteris paribus, a firm's output increases with the digital-intensity threshold. To further simplify the notation, we drop the industry symbol μ and refer to it as a representative industry hereafter.

With equations (1) and (4), the revenue of a firm with productivity θ is as:

$$r_x(\theta) = r_d(\theta) + r_e(\theta) = [1 + I_x(\theta)\tau^{-\beta/(1-\beta)}]^{1-\beta} r_d(\theta), \quad (5)$$

where τ denotes trade costs between the two countries, $\tau \geq 1$, and $I_x(\theta)$ is a dummy variable representing market access to foreign markets.⁴ Without loss of generality, we presume that all countries are similar in size in the initial stage, such as $B = B^*$. Suppose that when a firm only serves the domestic market while $I_x(\theta) = 1$ if the firm serves both the domestic and foreign markets. Here, $r_d(\theta) = By(\theta)^\beta$ and $r_e(\theta) = B^*y^*(\theta)^\beta\tau^{-\beta}$ represents revenues from domestic sales and exporting, respectively.

In the standard Diamond-Mortensen-Pissarides framework, there exists searching and matching frictions in the labor market (Pissarides, 1974; Diamond, 1982a, 1982b). Following Helpman et al. (2010), we let $b > 0$ represents the unit searching cost of matching potential workers in a country. However, we simplify Helpman et al.'s (2010) framework by assuming that all firms in an industry take the sectoral bar of ability (e.g., $a_c(\mu)$) as given, so that no firm-specific screening costs are involved. As a result, firms incur the same screening costs to identify workers with abilities below the sectoral bar, as this standard is universally accepted and applied across the industry.

Each firm and its hired workers engage in strategic bargaining over the division of revenue from production. This bargaining requires that the marginal surplus of the firm equals the marginal surplus of labor. It turns out that labor obtains a share of $\beta/(1+\beta)$ revenue, while the remaining share $1/(1+\beta)$ is attributed to the firm.⁵ With equations (4) and (5), a representative firm's profits are then given by

$$\begin{aligned} \pi &\equiv \max_{n \geq 0, a_c \geq a_{\min}, I_x \in [0,1]} \left\{ \frac{r_x(\theta)}{1+\beta} - bn - f_d - I_x f_x \right\} b \\ &= \max_{n \geq 0, a_c \geq a_{\min}, I_x \in [0,1]} \left\{ \frac{(\bar{\lambda}\theta^{\frac{1}{1-\mu}} \phi^{\frac{\mu}{1-\mu}} a_c(\mu)^{\frac{\mu\gamma}{1-\mu} - \lambda})^\beta}{1+\beta} - bn - f_d - I_x f_x \right\}, \quad (6) \end{aligned}$$

where $f_x > 0$ represents an additional fixed cost of exploring foreign markets.

⁴In monopolistic competition, exporting firms equal the marginal revenues in the two separate markets, leading to $[y_x(\theta)/y_d(\theta)]^{1-\beta} = \tau^{-\beta}(B^*/B) = \tau^{-\beta}$ if we presume $B = B^*$.

⁵The solution to the differential equation $\frac{\partial(r-wI)}{\partial I} = w$ is given by $w = \frac{\beta}{1+\beta} \frac{r}{I}$.

Taking the first order conditions of equation (6) with respect to the measure of workers sampled (n), we obtain:

$$\left(\frac{\beta}{1+\beta}\right)r_x(\theta) = bn. \quad (7)$$

Plugging equilibrium in equation (7) to (5), we obtain the optimal revenue of a firm:

$$r^*(\theta) = \eta[\theta^{\frac{1}{1-\mu}}\phi^{\frac{\mu}{1-\mu}}a_c(\mu)^{\frac{\mu\gamma}{1-\mu}-\lambda}]^{\frac{\beta}{1-\beta}}, \quad (8)$$

where $\eta = \left[b^{-1}\left(\frac{\beta}{1+\beta}\right)\bar{\lambda}\right]^{\frac{\beta}{1-\beta}}$ is a constant term. As implied in equation (8), a firm's revenue increases with its productivity (θ), with the country-specific digital development (ϕ), and with the industry-specific average digital ability of workers ($\bar{a}_c(\mu)$). From above, we obtain the optimal profit of an exporting firm as:

$$\pi_x(\theta) = \frac{1-\beta}{1+\beta}r_x(\theta) - f_d - f_x. \quad (9)$$

Equations (8) and (9) determine the productivity cutoffs. Specifically, firms with a productivity cutoff θ_d make zero profit in the domestic market:

$$\theta_d = \left(\frac{1+\beta}{1-\beta}\frac{f_d}{\eta}\right)^{\left(\frac{1-\beta}{\beta}\right)(1-\mu)}\phi^{-\mu}a_c(\mu)^{-\mu\gamma+\lambda(1-\mu)}. \quad (10)$$

The cut-off productivity for exporting firms is given by:

$$\theta_x = [(1+\tau^{-\beta/(1-\beta)})^{-(1-\beta)}(1+f_x/f_d)]^{\frac{1-\beta}{\beta}(1-\mu)}\theta_d, \quad (11)$$

where we assume f_x/f_d is sufficiently larger than τ to ensure $\theta_x > \theta_d$. Firms able to serve both domestic and foreign markets must have a productivity level that is greater than θ_x , while firms with a productivity level lower than θ_d exit the market, and firms with productivity along $[\theta_d, \theta_x]$ only serve the domestic market.

As implied in equation (10) and (11), digitalization helps small firms overcome entry barriers, thereby increasing employment and product variety. This phenomenon is more pronounced in high digital-intensive sectors

⁶With revenue of $r_d(\theta) = By(\theta)^\beta$, $l(\mu) = n(a_{\min}/a_c(\mu))^\lambda$, and $y = \bar{\lambda}\theta^{\frac{1}{1-\mu}}n\phi^{\frac{\mu}{1-\mu}}a_c(\mu)^{\frac{\mu\gamma}{1-\mu}-\lambda}$ in equation (4), we obtain that

$y = \bar{\lambda}a_{\min}^{-\lambda}\theta^{\frac{1}{1-\mu}}\phi^{\frac{\mu}{1-\mu}}a_c(\mu)^{\frac{\mu\gamma}{1-\mu}}l$.
⁷In (9), $\pi_d(\theta) = \frac{1-\beta}{1+\beta}r_d(\theta) - f_d = 0$ leads to $\theta_d = \left(\frac{1+\beta}{1-\beta}\frac{f_d}{\eta}\right)^{\frac{1-\beta}{\beta}(1-\mu)}\phi^{-\mu}a_c(\mu)^{-[\mu\gamma-(1-\mu)\lambda]}$.

than in low digital-intensive sectors. These observations lead to the following proposition:

PROPOSITION 1. *Digitalization helps smaller, less productive firms overcome productivity barriers to market entry, thereby increasing sectoral employment and the variety of products available in those industries. However, this phenomenon is more pronounced in high digital-intensive sectors than in low digital-intensive sectors.*

Proof. In equation (10), we obtain the derivation $\frac{\partial \ln \theta_d}{\partial \phi} = -\mu \phi^{-1} < 0$. This implies that a country's improvement in digital development reduces the sectoral productivity threshold, with the effect being more pronounced in industries that are relatively high in digital intensity (i.e., a greater μ leads to a larger reduction in θ). ■

Free Entry Condition

Suppose f_e defines the entry cost for all firms. Consequently, the free entry condition requires:

$$\begin{aligned} & f_d \int_{\theta_x}^{\infty} \theta_d [(\theta/\theta_d)^{\beta/1-\beta} - 1] f G_{\theta}(\theta) \\ & + f_x \int_{\theta_x}^{\infty} [(\theta/\theta_d)^{\beta/1-\beta} - 1] (1-\phi)^{-(1-\gamma)\beta/(1-\beta)} dG_{\theta}(\theta) = f_e. \end{aligned} \quad (12)$$

The free entry and zero profit conditions constitute the equilibrium of threshold productivity. With equations (9) and (12), we obtain the sectoral total income as

$$E = M \int_{\theta_x}^{\infty} \theta_d w(\theta) l(\theta) dG(\theta) = \kappa M z f_e, \quad (13)$$

where $\kappa = \left(\frac{\beta}{1+\beta}\right) \left(\frac{1}{1-\mu} \frac{\beta}{1-\beta}\right)^{-1}$ and M denotes the total number of differentiated firms in the sector. After some algebra, we obtain $M = \kappa E / z f_e$.

3. EQUILIBRIUM

With equations (5), (7), (8), and (11), the employment of a firm with productivity θ in sector μ of a country with digital development ϕ is given by $l(\phi, \theta) = \left[b^{-1} \left(\frac{\beta}{1+\beta} \right) \bar{\lambda} \right]^{\frac{1}{1-\mu}} \left[\theta^{\frac{1}{1-\mu}} \phi^{\frac{\mu}{1-\mu}} a_c(\mu)^{\frac{\mu\gamma}{1-\mu} - \frac{\lambda}{\beta}} \right]^{\frac{\beta}{1-\beta}}$.⁸ Combining the

⁸We have $l(\phi, \mu) = n(a_{\min}/a_c(\mu))^{\lambda} = \left[b^{-1} \left(\frac{\beta}{1+\beta} \right) \bar{\lambda} \right]^{\frac{1}{1-\mu}} \left[\theta^{\frac{1}{1-\mu}} \phi^{\frac{\mu}{1-\mu}} a_c(\mu)^{\frac{\mu\gamma}{1-\mu} - \frac{\lambda}{\beta}} \right]^{\frac{\beta}{1-\beta}}$.

above equation with equation (10), for firms that engage in exporting their employment can be rewritten as

$$l_x(\phi, \theta) = (1 + I_x \tau^{-\beta/(1-\beta)})^{(1-\beta)} (\theta/\theta_d)^{(\frac{1}{1-\mu})(\frac{\beta}{1-\beta})} a_c(\mu)^{(\frac{\beta}{1-\beta})(\mu - \frac{\lambda}{\beta(\gamma+\lambda)})} l_d, \quad (14)$$

where $l_d = \left(\frac{\beta}{1-\beta} \frac{f_d}{b}\right) \bar{\lambda}$ is a constant.

4. SECTORAL WAGE INEQUALITY

From equations (8), (10), (14), the wage of a firm with productivity θ in sector μ in a country with digital development ϕ is:

$$w_x(\phi, \mu) = (1 + I_x \tau^{-\beta/(1-\beta)})^{(1-\beta)} a_c(\mu)^\lambda w_d, \quad (15)$$

where $w_d = b\bar{\lambda}^{-1}$. Equation (15) suggests that all firms in an industry pay the wages to their workforce when recruiting based on the sectoral digital intensity. In this model, although there is no wage inequality within an industry, firms in relatively high digital-intensive sectors tend to, ceteris paribus, generate higher revenue, pay higher wages, and increase employment more than firms in relatively low digital-intensive sectors.¹⁰ We summarize these arguments in the following proposition:

PROPOSITION 2. *Digital development does not result in sectoral wage inequality. Firms in high digital-intensive industries tend to offer higher wages to their workforce compared to those in low digital-intensive industries.*

Proof. In equation (15), we have $\partial \ln w_x / \partial \phi = 0$, implying that digital development does not result in wage inequality within an industry. We also obtain $\partial \ln w_x / \partial \mu = \lambda \bar{a}'_c(\mu) / \bar{a}_c(\mu) > 0$ from equation (15). This indicates that, within a country, firms in relatively high digital-intensive industries, ceteris paribus, tend to pay higher wages to their workforce compared to those in low digital-intensive industries. Consequently, although there is no wage inequality within an industry, wage inequality exists across different industrial sectors. ■

⁹Given $w = \frac{\beta}{1+\beta} \frac{r}{l}$, we obtain $w = \frac{\beta}{1+\beta} \frac{r}{l} = a_c(\mu)^\lambda w_d$.

¹⁰In equation (8), we obtain $r'_x(\mu) > 0$. We also have $r'_x(\theta) > 0$ when μ is sufficiently large.

5. SECTORAL EMPLOYMENT

With equations (11), (12), (13), and (14), a country's total sectoral employment is:

$$L_x(\phi, \mu) = \{1 + [(1 + \tau^{-\beta/(1+\beta)})^{1-\beta} - 1](\theta_x/\theta_d)^{-z}\}L(\mu), \quad (16)$$

where $L(\phi, \mu) = Ml_d z(z - \frac{1}{1-\mu} \frac{\beta}{1-\beta})^{-1} \phi^{\mu z} a_c(\mu)^{[\mu\gamma - \lambda(1-\mu)]z} (\frac{\theta_x}{\theta_d} - 1)$ denotes total sectoral employment of non-exporters.¹¹ With equation (16), it is easy to obtain $\frac{\partial \ln L_x}{\partial \phi} = \frac{\partial \ln L}{\partial \phi} = \mu \frac{z}{\phi} > 0$, implying that an improvement in digital development promote sectoral employment, particularly in high digital-intensive industries (a greater μ). This suggests that an improvement in digital infrastructure acts as a productivity enhancer, enabling firms, particularly those with lower productivity, to overcome entry barriers and increase their workforce. However, the impact is asymmetrical, disproportionately benefiting high digital-intensive industrial sectors, where the enhanced digital capabilities amplify efficiency and demand for labor.

In the current model, we assume that the industry-specific data intensity parameter μ is exogenous across countries. However, in the digital era, data utilization in production has become increasingly prominent, representing a fundamental shift in how goods and services are created. Compared to traditional production models, where inputs were generally limited to labor, capital, and human capital, data is gradually serving as a critical factor driving efficiency, innovation, and competitiveness. To reflect this evolution, here we relax the assumption of exogenous digital intensity across sectors. Instead, we propose that sectoral digital intensity is increasing with the growing importance of information technology and digital infrastructure in shaping modern economies — a process we term digitalization.

Then, we can derivate equation (16) to analyze how sectoral employment changes with digitalization as $\frac{\partial \ln L_x}{\partial D} = \frac{\partial \ln L}{\partial D} = (\gamma + \lambda)z(\mu - \frac{\lambda}{\gamma + \lambda})(\frac{a'_c(\mu)}{a_c(\mu)})\mu'(D)$, where D denotes the level of digitalization. This expression suggests that when a sector's digital intensity is sufficiently high as $\mu > \frac{\lambda}{\gamma + \lambda}$, the sectoral employment increases with digitalization as $\frac{\partial \ln L_x}{\partial D} > 0$. However, when a sector's digital intensity is sufficiently low as $\mu < \frac{\lambda}{\gamma + \lambda}$, the sectoral employment decreases with digitalization as $\frac{\partial \ln L_x}{\partial D} < 0$.

The above arguments emphasize how improvements in digital infrastructure and data utilization have an asymmetric impact on sectoral employment, reallocating resources from low digital-intensive industries to high

¹¹In equation (14), given $G(\theta)$, $l(\phi, \mu)$ and θ_d that we have calculated, we obtain the solution: $L(\phi, \mu) = M \int_{\theta_d}^{\theta_x} l(\theta) dG(\theta) = Ml_d z(z - \frac{1}{1-\mu} \frac{\beta}{1-\beta})^{-1} \phi^{\mu z} a_c(\mu)^{[\mu\gamma - \lambda(1-\mu)]z} (\frac{\theta_x}{\theta_d} - 1)$

digital-intensive industries. The following proposition summarizes this observation.

PROPOSITION 3. Digitalization drives inter-industry resource reallocation, increasing employment in high digital-intensive industries while decreasing it in low digital-intensive industries.

This implication is striking. It has been perceived in the literature that technology adoption, such as computation or digitalization, generally leads to labor substitution across all sectors. However, our model suggests that sectoral employment in a country might instead increase with the country's digital development if the associated sectors are sufficiently high digital-intensive. The intuition behind this finding is twofold. Firstly, digital development lowers the productivity threshold, allowing small productive firms to overcome barriers to market entry, hereby employing workers. Secondly, in practice, firm productivity typically follows a distribution like the Pareto distribution, where only a small portion of firms exhibit high productivity, while the majority have low productivity.

On one hand, low-productivity firms, which constitute the majority, have the potential to hire a substantial number of workers, thereby boosting sectoral employment. On the other hand, digitalization favors high-productivity firms in high digital-intensive industries, leading to a substitution effect that replaces some low-digital-ability labor. On balance, sectoral employment is not necessarily reduced when the substitution effect is offset by the entry of more low-productivity firms into the market, facilitated by digital development lowering entry barriers and enabling these firms to increase employment.

Final, we attribute wage increases and workforce expansion mainly to screening costs, which enhance a firm's digital capabilities and, in turn, boost productivity, wages, and employment. However, an endogenous data-driven productivity mechanism, where high data utilization improves efficiency and raises demand for skilled labor, may also influence these outcomes. Exploring this endogenous mechanism is beyond the scope of this lengthy paper and is left for future research.

6. EMPIRICAL IMPLEMENTATION

The implication necessitates a comprehensive examination using actual data. We thus analyze Taiwan's city-level data to empirically explore whether variations in digital intensity among cities influenced labor market outcomes from 2001 to 2019. Even though the theoretical model focuses on the effects of digitalization on the labor outcomes across industries of different digital intensities, the variations of industrial structures across cities

allow us to test whether the shocks of digitalization have a high impact on cities that are dominated by high digital intensive industries. Table 1 reports the share of industrial sales within a city according to the industrial level of digitalization.¹² The first column shows the share of sales of industries with digitalization index below the 25th percentile (the lowest digital intensive industries). The second column reports the share of sales of industries with digitalization index in the range of the 25th percentile and the median. The third column shows the share of sales of industries with digitalization index above the median and below the 75th percentile. The last column reports the share of sales of industries with digitalization indices above the 75th percentile (the highest digital intensive industries). For instance, industries in Keelung, Taipei and New Taipei cities are likely to be high-digital intensive industries, while Yilan county is dominated by low-digital intensive industries.

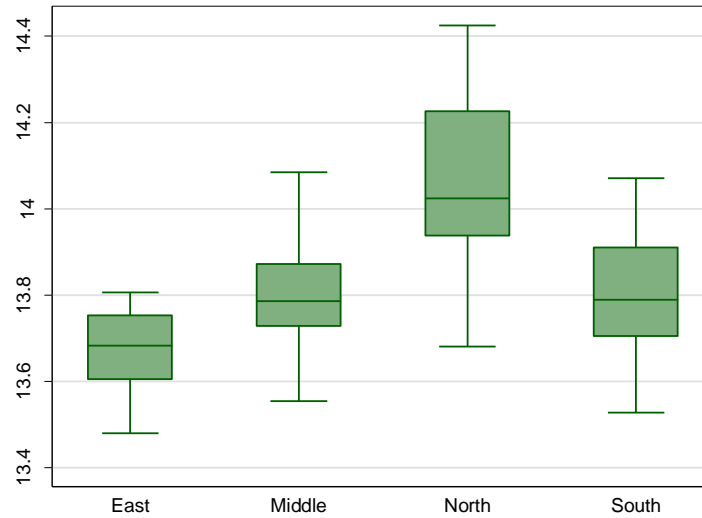
TABLE 1.
Distribution of Industrial Sales in a City, by Digitalization Level

Region	City/County	Share of Industrial Sales with Digital index in the Range			
		(0, 25%]	(25%, 50%]	[50%,75%]	(75%,100%]
North	Keelung City	14.0	18.3	4.4	63.4
	Taipei City	5.4	7.1	6.5	81.0
	New Taipei City	13.9	16.2	10.3	59.6
	Yilan County	21.7	56.0	7.3	15.0
	Taoyuan City	18.1	19.0	7.0	55.9
	Hsinchu County	4.5	17.2	5.0	73.3
	Hsinchu City	1.5	3.5	4.2	90.8
Central	Miaoli County	16.5	29.2	7.4	46.9
	Taichung City	25.7	16.4	20.6	37.3
	Changhua County	29.7	42.4	9.3	18.6
	Nantou County	27.7	24.3	8.3	39.7
	Yunlin County	45.7	45.0	4.1	5.2
South	Chiayi County	20.4	61.7	7.3	10.5
	Chiayi City	46.7	15.6	23.3	14.4
	Kaohsiung City	45.8	29.9	4.9	19.4
	Tainan City	28.4	23.0	6.4	42.2
	Pingtung County	51.8	19.9	6.1	22.2
East	Taitung County	48.7	25.3	10.4	15.6
	Hualien County	13.5	83.0	3.1	0.5

¹²The digitalization index is based on Mucha and Seppala (2021) which is reported in Table 2.

The advantage of using city-level data within one country is that there is less heterogeneity in the labor market conditions that affect employment and labor wages (e.g., labor standards) within a country, when making international comparisons across countries. However, there may be some local shocks impacting labor market outcomes and the digital development. We include a set of region dummy variables to control for unobserved local shocks in the regression model, since income distributions in Taiwan primarily vary across regions rather than across cities within one region.

FIG. 1. Distribution of City-level Household Income of Taiwan, by Regions



The North region of Taiwan has the highest average household income, and which is around 1.3 million NTD (USD 40,000). The East region has the lowest income level, accounting for only 67 percent of the income level of the North region. Figure 1 displays the box plot of the distribution of city-level income (in logs) across regions (North, South, Middle, and East) during the period of 2001-2019. The lower quartile of the income level in the North region is significantly higher than the upper quartile of income in the other regions, suggesting that variations in household income are stronger across regions than across cities within one region.

6.1. Data

In this subsection, we introduce the data used in the empirical examination.

B. Measures of Labor Participation and Wages

We use data of labor participation and wage during the period of 2001 to 2019 from the Taiwan Manpower Survey published by the government's Directorate-General of Budget, Accounting and Statistics (DGBAS). The Taiwan Manpower Survey collects individual citizens' labor participation status, their wage level, the industry that the individual works for, and the city where the individual lives. According to this information, we can calculate the total number of laborers ($L_{c,t}$) and the average wages across individuals ($W_{c,t}$) in city c of year t . Wage inequality in each city is measured by the standard deviation of wages ($SDW_{c,t}$) across individuals in city c of year t . We separately construct city-level employment, wage, and wage inequality data for the manufacturing and service sectors.

C. Measure of Firm Turnover Rate

The Ministry of Economic Affairs (MOEA) reports the number of new establishments, the number of firms that shut down, and the number of active establishments in each city in Taiwan. Based on this information, we calculate the firm entry rate and exit rate in a city by

$$En_{c,t} = NP_{c,t}/CP_{c,t} \text{ and } Ex_{c,t} = EP_{c,t}/CP_{c,t},$$

where $En_{c,t}$ is the firm entry rate in city c of year t , and $Ex_{c,t}$ is the firm exit rate in city c of year t . $NP_{c,t}$ is the number of new firms in city c of year t , $CP_{c,t}$ is the number of active firms in city c of year t , and $EP_{c,t}$ is the number of firms that shut down in city c of year t . In the service sector, we calculate firm entry and exit rates in each city. This firm turnover rate is calculated by

$$turnover_{c,t} = \frac{En_{c,t} + Ex_{c,t}}{2}.$$

D. Measure of Digital Development

There are some variables in the following that represent digital development of a city. A high usage of computers or mobile phones among citizens in a city reflects a high level of digital development. The number of personal computers per one hundred households and the average number of mobile phones in one household are two measures of city-level digital development applied here. The usages of personal computers and mobile phones are obtained from the Family Income and Expenditure Survey (DGBAS).

In addition to the measure of digital development at the household level, we also adopt another measure of digital development that represents the structure of industries associated with digital intensities. Production technology differs in digital intensity across industries. For example, production technology in the computer industry is digitally intensive, but the food manufacturing industry has a relatively low level of digital-intensity. Based on Calvio et al.'s (2018) method, Mucha and Seppala (2021) construct an indicator measuring industrial digital intensity. Table 2 reports Mucha

TABLE 2.
Sectoral Measurement of Digital Intensity

Sector	ISIC Code (rev. 4)	Digital Intensity Score
Agriculture, forestry, fishing	01-03	0.0463
Mining and quarrying	05-09	0.2361
Food products, beverages and tobacco	10-12	0.3254
Textiles, wearing apparel, leather	13-15	0.4246
Wood and paper products, and printing	16-18	0.4563
Coke and refined petroleum products	19	0.3532
Chemicals and chemical products	20	0.4087
Pharmaceutical products	21	0.3651
Rubber and plastics products	22-23	0.4365
Basic metals and fabricated metal products	24-25	0.369
Computer, electronic and optical products	26	0.5648
Electrical equipment	27	0.5185
Machinery and equipment (not elsewhere classified)	28	0.5324
Transport equipment	29-30	0.6157
Furniture; other manufacturing; repairs of computers	31-33	0.5754
Electricity, gas, steam and air cond.	35	0.3016
Water supply; sewerage, waste management	36-39	0.3016
Construction	41-43	0.2698
Wholesale and retail trade, repair	45-47	0.5926
Transportation and storage	49-53	0.3194
Accommodation and food service activities	55-56	0.287
Publishing, audiovisual and broadcasting	58-60	0.6157
Telecommunications	61	0.8796
IT and other information services	62-63	0.8241
Finance and insurance	64-66	0.8222
Real estate	68	0.0741
Legal and accounting activities, etc.	69-71	0.662
Scientific research and development	72	0.6204
Advertising and market research; other business services	73-75	0.6806
Administrative and support service activities	77-82	0.6528
Public administration and defense	84	0.5333
Education	85	0.3944
Human health activities	86	0.4333
Residential care and social work activities	87-88	0.4111
Arts, entertainment and recreation	90-93	0.4889
Other service activities	94-96	0.6167

Source: Mucha and Seppala (2021)

and Seppala’s (2021) measurement of digital intensity across industries, in which some are high digital-intensive sectors (e.g., IT service) and some are low digital-intensive sectors (e.g., agricultural and real estate industries).

We construct the city-level digital indicator that captures the industrial structures reflecting the digital intensities of a city as a whole. The digital index is constructed by:

$$WD_{c,t} = \sum_{s=1}^n S_{sc,t} \times DI_s,$$

where $S_{sc,t}$ is the labor share of sector s over total employment in city c and DI_s is defined as the sectoral digital intensity of sector $s = 1, 2, \dots, n$. A city has a high digital index ($WD_{c,t}$) if a large fraction of laborers works in the high digital-intensive sectors. Table 3 reports the average value for the measures of digital development in the four regions (North, Central, South and East) of Taiwan, which serves as the digital index of that region.¹³ The digital levels vary across regions. Cities in the North region are more likely to have a high level of digital development for the three measures. Cities in the East region have the lowest level of digital development.

TABLE 3.
Measures of Digital Levels, by Regions

Region	Digital Index	No. of Computers	No. of Mobile Phones
East	0.46	52.39	163.07
Central	0.49	70.89	196.70
North	0.51	100.30	228.32
South	0.48	68.62	189.59

Note: No. of computers/ phones is the number of personal computers/phones per one hundred households.

6.2. Empirical Models and Empirical Results

In this section, we introduce empirical models for testing the three hypotheses.

6.2.1. Firm Turnover Rate and Digital Development

Proposition 1 states that increased digital development helps lower productivity thresholds for new entrants, suggesting that cities with a high

¹³Keelung City, Taipei City, New Taipei City, Taoyuan City, Yilan City, Hsinchu County, and Hsinchu City are in the North region. Miaoli County, Taichung City, Changhua County, Nantou County, and Yunlin County are in the Central region. Chiayi County, Chiayi City, Tainan City, Kaohsiung City, and Pingtung County are in the South region. Hualien County and Taitung County are in the East region.

level of digital development are more likely to have a low productivity threshold for new entrants. A low productivity threshold for firm entry implies that potential entrants face a high probability of entering the market and incumbent firms face a high probability of exiting the market. Therefore, the model predicts that economies (e.g., cities) with a high level of digital development are likely to have high entry and exit rates. We test the first hypothesis by running the following regression model:

$$\ln(\textit{turnover}_{c,t}) = \alpha_0 + \alpha_1 X_{c,t-1} + \alpha_2 \ln(\textit{pop}_{c,t}) + \alpha_3 \ln(\textit{firm}_{c,t-1}) + D_g + D_t + \varepsilon_{c,t}, \quad (17)$$

where $\textit{turnover}_{c,t}$ as mentioned previously is the firm turnover rate in city c at year t , and $X_{c,t-1}$ is the measure of digital levels of city c in year $t - 1$. The digital level consists of digital intensities, represented by the digital index ($WD_{c,t}$) of city c that prioritizes specific digital-intensive sectors, and $I(HWD_{c,t-1})$ is an indicator variable which is equal to one if the digital index is above the median level. That is, $I(HWD_{c,t-1})$ is equal to one for the high digital-intensive city. The digital development, represented by both the number of computers per one hundred households ($\textit{compu}_{c,t-1}$) and the number of mobile phones per one hundred households ($\textit{mob}_{c,t-1}$) in the city. Here, $\textit{pop}_{c,t}$ is the population in city c that captures the market size of a city.

In equation (17), $\ln(\textit{firm}_{c,t-1})$ denotes the number of firms in logarithm. Nevertheless, we presume $\textit{firm}_{c,t-1} = \textit{plant}_{c,t-1}$ when $\textit{plant}_{c,t-1}$ denotes the number of plants in the manufacturing sectors located in city c , and $\textit{firm}_{c,t-1} = \textit{company}_{c,t-1}$ when $\textit{company}_{c,t-1}$ denotes the number of service providers in the city. Firms face a small local competition effect if the number of competitors in the local market is small. D_g includes a set of dummy variables controlling regional shocks, while D_t controls for the year effect.

Table 4 presents the estimation results.¹⁴ In columns (1), (2), (5) and (6) the specifications utilize the digital index as the measure of digital intensity for industries located in the city. In general, the positive coefficients of the digital index suggest that firms in cities with a significant presence of digital-intensive industries tend to experience higher turnover rates. Columns (3), (4), (7), and (8) report the estimation results when city-level digital development is gauged by the prevalence of computer and mobile phone usage among households.

¹⁴There is no data on firm turnover rates for years 2006, 2011 and 2016.

TABLE 4.
Firm Turnover Rate and Digitalization

Variables	Manufacturing				Service			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$WD_{c,t-1}$	0.443				2.042***			
	(1.022)				(0.761)			
$I(HWD_{c,t-1})$		0.103*				0.104***		
		(0.060)				(0.033)		
$\ln(compu_{c,t-1})$			0.069				0.301***	
			(0.128)				(0.077)	
$\ln(mob_{c,t-1})$				0.138				0.545***
				(0.275)				(0.189)
$\ln(plant_{c,t-1})$	-0.0395	-0.0418	-0.0388	-0.0386				
	(0.037)	(0.036)	(0.037)	(0.037)				
$\ln(pop_{c,t})$	0.235***	0.222***	0.229***	0.228***	0.613***	0.647***	0.634***	0.653***
	(0.055)	(0.054)	(0.054)	(0.055)	(0.057)	(0.056)	(0.056)	(0.057)
$\ln(company_{c,t-1})$					-0.660***	-0.701***	-0.693***	-0.712***
					(0.063)	(0.064)	(0.062)	(0.065)
Constant	-1.910***	-1.547***	-1.903***	-2.321*	-1.048***	-0.068	-1.163***	-2.794***
	(0.700)	(0.532)	(0.635)	(1.324)	(0.362)	(0.258)	(0.322)	(0.889)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	205	205	205	205	227	227	227	227
R-squared	0.467	0.475	0.468	0.468	0.735	0.738	0.744	0.736

Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$

Firm turnover rates are elevated in cities with a high usage rate of personal computers and/or mobile phones. However, these positive correlations between turnover rates and the level of digital development are only statistically significant in the service sector. Furthermore, high firm turnover rates are observed in cities with large populations and fewer competitors. This aligns with the idea that a large market can accommodate a greater number of firms and facilitates easier entry.¹⁵

¹⁵Bresnahan and Reiss (1991) indicate that there are more stores in regions with a large market size, when compared to small-sized regions. Furthermore, Berry (1992) finds that the probability of firm entry in a local market is negatively correlated to the number of rivals.

6.2.2. Labor Wage and Digitalization

Proposition 2 suggests that economies (e.g., cities) that dominated by more highly digital-intensive sectors tend to exhibit higher wage levels. This effect is enhanced when the economies (e.g., cities) are strongly developed in digital infrastructure. The relationship between the wage and digitalization is expressed as

$$\begin{aligned} \ln(w_{c,t}) = & \beta_0 + \beta_1 X_{c,t-1} + \beta_2 \ln(\text{sales}_{c,t-1}) + \beta_3 \ln(\text{firm}_{c,t-1}) + \beta_4 \text{Hed}_{c,t} \\ & + \beta_5 \ln(\text{econdep}_{c,t}) + D_g + D_t + \varepsilon_{c,t}, \end{aligned} \quad (18)$$

where $w_{c,t}$ is the average wage rate of city c in year t . Equation (18) incorporates city-level sectoral sales revenue because wages are expected to depend on sales revenues,¹⁶ in which we use $\ln(\text{sales}_{c,t-1})$ to denote the total sales of city c in year $(t-1)$. When sales revenue data for the service sector are unavailable, we utilize the city's population as a measure of sectoral sales, as a larger population typically indicates greater consumption. While the number of firms (companies/plants) in the market may affect labor demand and thus the equilibrium wage, we also control the number of firms $\ln(\text{firm}_{c,t-1})$ of city c in year $(t-1)$ in equation (18). The quality of labor influences the equilibrium wage. Here, $\text{Hed}_{c,t}$ is a dummy variable that is equal to one if the fraction of citizens who have a college degree is higher than 40%. $\text{econdep}_{c,t}$ is the city-level governmental expenditures on the economic activities.

The residuals in the regression model may be correlated to the measure of digitalization. For instance, cities with high wages may have more resources to invest in digital infrastructure, leading to a high digitalization level. We adopt the instrumental variable method to solve the endogeneity problem. The instrumental variable needs to be correlated to the endogenous variable, the measure of digitalization, but uncorrelated to the residuals of the labor wage equation. The road intensity at the city-level is chosen as the instrumental variable for the measure of digitalization. The road intensity is measured by the total length of roads divided by the area of the city. The transportation costs are low in cities with high road intensity. Firms have high intensives to invest in digital equipment due to the low transportation costs. While the road intensity is correlated to the digital levels of a city, it is less likely to be correlated to the labor market outcome. The 2SLS method is adopted and the estimation results of the first stage are reported in Table 5. The positive coefficients on the road intensity indi-

¹⁶Total sales in a city for the manufacturing sector are collected from the Ministry of Economics Affairs (MOEA).

cate that cities with high road intensities are likely to have high digital measures. We use the same instrumental variable, road intensity, for the following wage inequality and labor employment equations.

TABLE 5.

First Stage of the 2SLS of the Wage Equation

Variables	Manufacturing				Service			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\ln(Road_{c,t-1})$	0.004*** (0.001)	0.067*** (0.017)	0.049*** (0.007)	0.013** (0.004)	0.001 (0.001)	0.069*** (0.020)	0.021*** (0.008)	0.007* (0.004)
$\ln(sales_{c,t-1})$	0.013** (0.002)	0.139*** (0.034)	0.079*** (0.014)	0.013 (0.009)				
$\ln(plant_{c,t-1})$	-0.003 (0.002)	-0.023 (0.039)	0.031* (0.016)	0.029** (0.010)				
$Hed_{c,t}$	0.008* (0.004)	-0.005 (0.086)	0.162*** (0.034)	0.025 (0.022)	0.034*** (0.003)	0.345*** (0.097)	0.165*** (0.050)	0.020 (0.019)
$\ln(pop_{c,t})$					0.005 (0.004)	-0.142 (0.117)	0.079 (0.048)	0.005 (0.023)
$\ln(company_{c,t-1})$					0.001 (0.004)	0.467*** (0.112)	0.133** (0.047)	0.098*** (0.022)
$\ln(econdep_{c,t-1})$	-0.003* (0.002)	0.104** (0.036)	0.002 (0.014)	0.012 (0.009)	0.007** (0.002)	0.027 (0.068)	-0.086** (0.028)	-0.036** (0.014)
Constant	0.238*** (0.026)	-2.933*** (0.453)	1.906*** (0.182)	4.196*** (0.115)	0.372*** (0.017)	-2.903*** (0.482)	2.276*** (0.200)	4.328*** (0.096)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R-squared	0.63	0.49	0.84	0.82	0.62	0.44	0.73	0.74
Observations	285	285	285	285	247	247	247	247

Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Table 6 provides the estimation results of (18). The positive coefficients on the three digital measures support the theoretical prediction that wages in an economy rise with the economy's digital development, and the phenomenon is more pronounced in relatively high digital-intensive industries. The positive correlation between the levels of wage and digital measures is significant in both the manufacturing and service sectors.

In general, the total sales and the education level did not have a significant effect on labor wages. The coefficients on the number of plants (companies) are negative and statistically significant. Since sectoral sales revenue is also included in the regression model, one interpretation of the parameter is that cities with a large number of plants tend to offer lower

TABLE 6.
Wage and Digitalization (2nd Stage of 2SLS)

Variables	Manufacturing				Service			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$WD_{c,t-1}$	7.616***				14.504			
	(1.927)				(11.964)			
$I(HWD_{c,t-1})$		0.438***				0.172***		
		(0.106)				(0.058)		
$\ln(compu_{c,t-1})$			0.598***				0.556***	
			(0.098)				(0.212)	
$\ln(mob_{c,t-1})$				2.239***				1.793*
				(0.735)				(1.009)
$\ln(sales_{c,t-1})$	-0.049**	-0.009	0.005	0.024				
	(0.025)	(0.017)	(0.010)	(0.019)				
$\ln(plant_{c,t-1})$	-0.019	-0.033**	-0.061***	-0.108***				
	(0.017)	(0.016)	(0.012)	(0.034)				
$Hed_{c,t}$	0.025	0.090**	-0.009	0.032	-0.461	-0.029	-0.062	-0.006
	(0.049)	(0.036)	(0.037)	(0.062)	(0.441)	(0.037)	(0.053)	(0.049)
$\ln(pop_{c,t})$					-0.069	0.031	-0.038	-0.003
					(0.094)	(0.023)	(0.035)	(0.040)
$\ln(company_{c,t-1})$					-0.041	-0.113***	-0.106***	-0.208**
					(0.053)	(0.029)	(0.031)	(0.093)
$\ln(econdep_{c,t-1})$	0.069***	-0.002	0.042***	0.017	-0.061	0.038***	0.090***	0.107**
	(0.019)	(0.017)	(0.010)	(0.021)	(0.088)	(0.013)	(0.025)	(0.045)
Constant	7.607***	10.702***	8.278***	0.025	4.958	10.849***	9.085***	2.589
	(0.576)	(0.285)	(0.271)	(3.209)	(4.489)	(0.181)	(0.510)	(4.392)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	285	285	285	285	247	247	247	247

Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

wages at a given level of sales revenue. Furthermore, a large number of plants implies that each plant is small in size with low labor demand. The positive coefficients on $\ln(econdep)$ suggest that local government expenditures on economic development have positive effects on labor wages. Local government provides funds to help firms develop and innovate new technology that may increase firms' profit and labor wage.

Proposition 2 also implies that there is no wage inequality within an industrial sector. However, wage inequality does exist across sectors. More specifically, wages in relatively high digital-intensive industries tend to be

higher than in low digital-intensive industries within an economy (or city). We test this hypothesis by examining the following equation:

$$\ln(SDW_{c,t}) = \gamma_0 + \gamma_1 X_{c,t-1} + \gamma_2 \ln(sales_{c,t-1}) + \gamma_3 \ln(firm_{c,t-1}) + \gamma_4 \ln(econdep_{c,t}) + D_g + D_t + \varepsilon_{c,t}, \quad (19)$$

where $SDW_{c,t}$ is the standard deviation of labors' wage in city c . Being similar to the wage equation, we also use the road intensity to control the endogeneity problem. The estimation results are reported in Table 7.

TABLE 7.
Wage Inequality and Digitalization (2nd Stage of 2SLS)

Variables	Manufacturing				Service			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$WD_{c,t-1}$	2.778 (2.611)				-5.059*** (1.255)			
$I(HWD_{c,t-1})$		0.201 (0.187)				-0.249*** (0.067)		
$\ln(compu_{c,t-1})$			0.196 (0.179)				-0.652*** (0.186)	
$\ln(mob_{c,t-1})$				0.834 (0.790)				-3.127*** (1.193)
$\ln(sales_{c,t-1})$	-0.209*** (0.038)	-0.199*** (0.033)	-0.187*** (0.029)	-0.182*** (0.030)				
$\ln(plant_{c,t-1})$	-0.162*** (0.036)	-0.167*** (0.036)	-0.176*** (0.037)	-0.195*** (0.046)				
$\ln(pop_{c,t})$					-0.450*** (0.064)	-0.494*** (0.068)	-0.412*** (0.076)	-0.439*** (0.099)
$\ln(company_{c,t-1})$					-0.046 (0.059)	0.043 (0.068)	0.0191 (0.069)	0.226* (0.136)
$\ln(econdep_{c,t-1})$	-0.076** (0.034)	-0.104*** (0.037)	-0.087*** (0.032)	-0.096*** (0.035)	0.039 (0.038)	0.008 (0.041)	-0.055 (0.048)	-0.112 (0.076)
Constant	9.083*** (0.869)	10.312*** (0.532)	9.379*** (0.613)	6.242* (3.498)	12.748*** (0.535)	10.194*** (0.339)	12.389*** (0.526)	24.467*** (5.201)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	285	285	285	285	247	247	247	247

Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

There is no significant effect of digitalization on wage inequality in the manufacturing sector. However, there are negative coefficients on the digi-

tal index in the service sector, suggesting that wage inequality is less pronounced in cities that prioritize high digital-intensive industries in the service sector. In addition, cities with large level of sectoral sales and large populations tend to have less wage inequality. The negative coefficients on $\ln(econdep_{c,t})$ suggest that cities with high local government expenditures on economic development have low wage inequality. However, these effects are only significant in the manufacturing sector.

We argue that while digital-intensive industries generally offer higher wages, cities that emphasize digital development may invest more in education and infrastructure, leading to a more equitable distribution of digital skills across the workforce. This hence reduces the wage premium and mitigates the overall wage inequality in these cities. Finally, the wage distribution tends to be more equalized in cities with higher industrial sales and a larger number of plants in the manufacturing sector.

6.2.3. *Employment and Digitalization*

Proposition 3 suggests that digitalization increases sectoral employment in high digital-intensive sectors, but decreases it in low digital-intensive sectors. Given that each of Taiwan's cities has distinct focus on industry development, ranging from high to low digital-intensive sectors, we would expect that cities prioritizing high digital-intensive sectors and having advanced digital development are likely to experience higher levels of employment. This hypothesis can be tested by the following regression model:

$$\begin{aligned} \ln(L_{c,t}) = & \delta_0 + \delta_1 X_{c,t-1} + \delta_2 \ln(sales_{c,t-1}) + \delta_3 \ln(firm_{c,t-1}) \\ & + \delta_4 \ln(econdep_{c,t}) + D_g + D_t + \varepsilon_{c,t}, \end{aligned} \quad (20)$$

where $L_{c,t}$ is the employment level of city c in year t . The estimation results are reported in Table 8.

In Table 8, the positive coefficients for both the industrial digital intensity and digital development parameters aligns with our theoretical model's predictions. Cities with a prevalence of digital-intensive industries are likely to have higher employment levels. Additionally, cities with widespread computer and mobile phone usage also exhibit higher levels of employment. The positive correlations between sectoral sales with labor employment indicate that a large market drives higher demand for labor. Moreover, the positive coefficients on industrial sales and population reflect that large markets have high labor demand. Lastly, cities with a substantial number of plants or companies tend to have higher levels of employment.

TABLE 8.Labor Employment and Digitalization (2nd Stage of 2SLS)

Variables	Manufacturing				Service			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$WD_{c,t-1}$	20.045*** (3.027)				23.146*** (2.055)			
$I(HWD_{c,t-1})$		1.453*** (0.305)				1.141*** (0.135)		
$\ln(compu_{c,t-1})$			1.415*** (0.139)				2.984*** (0.343)	
$\ln(mob_{c,t-1})$				6.019*** (1.175)				14.310*** (3.672)
$\ln(sales_{c,t-1})$	0.125*** (0.044)	0.193*** (0.053)	0.279*** (0.023)	0.316*** (0.045)				
$\ln(plant_{c,t-1})$	0.492*** (0.042)	0.454*** (0.058)	0.390*** (0.029)	0.253*** (0.069)				
$\ln(pop_{c,t})$					0.747*** (0.104)	0.951*** (0.138)	0.573*** (0.140)	0.696** (0.304)
$\ln(company_{c,t-1})$					0.175* (0.096)	-0.229* (0.136)	-0.122 (0.129)	-1.066** (0.420)
$\ln(econdep_{c,t-1})$	0.370*** (0.040)	0.168*** (0.061)	0.291*** (0.025)	0.231*** (0.052)	-0.047 (0.061)	0.094 (0.082)	0.380*** (0.088)	0.644*** (0.235)
Constant	-7.530*** (1.007)	1.336 (0.866)	-5.397*** (0.477)	-28.035*** (5.202)	-11.158*** (0.875)	0.529 (0.683)	-9.517*** (0.975)	-64.781*** (16.009)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	285	285	285	285	247	247	247	247

Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

7. ALTERNATIVE MEASURE OF DIGITALIZATION

In this section, we use the fraction of firms that adopt digital technology as an alternative measure of the digitalization of a city. Taiwan's government surveys every firm in the manufacturing and service sectors every five years and collects the information about whether the firm uses computers or the internet in production. During the sample periods, we have data on the fraction of firms that use computers or internet for the manufacturing and service sectors in 2006, 2011 and 2016. We use this fraction as the

TABLE 9.
Firm Digital Technology Adoption and Labor Market Outcome

	1st Stage			2nd Stage		
	(1)	(2)		(3)	(4)	(5)
VARIABLES	$\ln FD$	$\ln FD$	VARIABLES	$\ln W$	$\ln SDW$	$\ln L$
	$(\ln W)$	$(\ln SDW / \ln L)$				
$\ln(Road_{c,t-1})$	0.036**	0.042***	$\ln FD_{c,t}$	0.435**	-1.030**	3.408***
	(0.011)	(0.008)		(0.185)	(0.452)	(0.650)
$Hed_{c,t}$	0.040		$Hed_{c,t}$	0.064		
	(0.058)			(0.040)		
$\ln(econdep_{c,t-1})$	0.020	0.023	$\ln(econdep_{c,t-1})$	0.035**	-0.026	0.198**
	(0.024)	(0.024)		(0.014)	(0.054)	(0.078)
$\ln(sales_{c,t})$	0.070**	0.070**	$\ln(sales_{c,t})$	0.013	-0.206***	0.227***
$(\ln(pop_{c,t}))$	(0.026)	(0.059)	$(\ln(pop_{c,t}))$	(0.018)	(0.059)	(0.085)
$\ln(plant_{c,t-1})$	0.045	0.045	$\ln(plant_{c,t-1})$	-0.064***	-0.031	0.267**
$(\ln(company_{c,t-1}))$	(0.030)	(0.030)	$(\ln(company_{c,t-1}))$	(0.020)	(0.074)	(0.106)
D_s	-0.108	-0.109	D_s	0.501***	-1.798***	2.963***
	(0.230)	(0.229)		(0.143)	(0.539)	(0.776)
Constant	1.827***	1.803***	Constant	8.726***	12.459***	-9.820***
	(0.324)	(0.321)		(0.442)	(1.273)	(1.833)
Year FE	Yes	Yes	Year FE	Yes	Yes	Yes
Region FE	Yes	Yes	Region FE	Yes	Yes	Yes
Observations	95	95	Observations	95	95	95

Note: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

measure of digitalization of firms and run the following regression model:

$$\begin{aligned}
n(Y_{c,t} = & \theta_0 + \theta_1 X_{c,t-1} + \theta_2 \ln(sales_{c,t-1}) + \theta_3 \ln(firm_{c,t-1}) \\
& + \theta_4 \ln(econdep_{c,t}) + D_g + D_t + \varepsilon_{c,t}, \tag{21}
\end{aligned}$$

where $Y_{c,t}$ represents labor wages ($W_{c,t}$), wage inequality ($SDW_{c,t}$), and labor employment ($L_{c,t}$).¹⁷ $FD_{c,t}$ is the fraction of firms in city c and year t that use computers or the internet in production. Since we only have data for three years, we pool the manufacturing and service sectors together and include a dummy variable (D_s) which is equal to one for the service sector to reflect the variation of labor market outcome in these two sectors. Similar to the regression model in Sections 6.2.2 and 6.2.3, we use

¹⁷There is no data on the firm turnover rates in years 2006, 2011 and 2016 and we cannot estimate the firm-turnover equation (17).

the road intensity as the instrument for $FD_{c,t}$ to solve the endogeneity problem. The estimation results are reported in Table 9.

Columns (1) and (2) report the estimation results of the 1st stage of the 2SLS model. The positive and significant coefficients on the road intensity suggest that cities with a high road intensity are likely to have a high fraction of firms adopting digital technology. Columns (3)-(5) show the 2nd stage of the estimation results for the labor wage, wage inequality and labor employment equations, respectively. The positive coefficients on firms' adoption of digital technology ($\ln FD$) in the wage and labor employment equations suggest that cities occupied by high digital-intensive industries are likely to have high wages and employment levels. The negative coefficient of $\ln FD$ in the wage inequality equation indicates that the distribution of labor wage is more equal in cities with high digital intensities. These results are consistent with those in Section 6.

8. CONCLUSIONS

This paper presents a simple model suggesting that a country can enhance its welfare by promoting digital development, which lowers productivity thresholds for new entrants, enabling the generation of a greater variety of goods. However, higher digital-intensive sectors tend to benefit more from digital development, experiencing greater increases in wages and employment.

Contrary to the common belief that technology adoption leads to job losses, the current model suggests that digitalization can boost sectoral employment across sectors, particularly in highly digital-intensive sectors. While digitalization may exert a substitution effect that displaces some workers, improved digital development can offset this effect by lowering productivity thresholds for new entrants. These new entrants can subsequently increase sectoral employment. Consequently, compared to low digital-intensive industries, the overall outcome may be an increase in sectoral employment in high digital-intensive industries, as their enhanced digital capabilities amplify efficiency and labor demand.

We empirically test the predictions of the theoretical model by examining the relation between digitalization and labor market across cities in Taiwan. These empirical results align with our theoretical model, where digital development lowers the productivity threshold for market entry, enhances labor employment and wages, and mitigates wage inequality. Cities with high digital intensity tend to exhibit higher firm turnover rates, increased employment, higher wage levels, and reduced wage inequality.

REFERENCES

- Acemoglu, Daron, and Pascual Restrepo, 2018. The Race Between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment. *American Economic Review* **108(6)**, 1488-1542.
- Axtell, Robert L., 2001. Zipf Distribution of U.S. Firm Sizes. *Science* **293**, 1818-1820.
- Bajari, Patrick, Denis Nekipelov, Stephen P. Ryan, and Miaoyu Yang, 2015. Machine Learning Methods for Demand Estimation. *American Economic Review: Papers and Proceedings* **105(5)**, 481-485.
- Baldwin, R. E., and G. G. Cain, 2000. Shifts in US Relative Wages: The Role of Trade, Technology, and Factor Endowments. *Review of Economics and Statistics* **82**, 580-595.
- Berry, Steven, 1992. Estimation of a Model of Entry in the Airline Industry. *Econometrica* **60(4)**, 889-917.
- Bessen, James E., 2017. AI and Jobs: The Role of Demand. Boston Univ. School of Law, Law and Economics Research Paper No. 17-46.
- Bresnahan, T., and P. Reiss, 1991. Entry and Competition in Concentrated Markets. *Journal of Political Economy* **99(5)**, 977-1009.
- Brynjolfsson, E., and A. McAfee, 2014. The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. W. W. Norton & Co.
- Clarke, G. R. G., and S. J. Wallsten, 2006. Has the Internet Increased Trade? Developed and Developing Country Evidence. *Economic Inquiry* **44(3)**, 465-484.
- Diamond, Peter A., 1982a. Aggregate Demand Management in Search Equilibrium. *Journal of Political Economy* **90**, 881-894.
- Diamond, Peter A., 1982b. Wage Determination and Efficiency in Search Equilibrium. *Review of Economic Studies* **49**, 217-227.
- Franken, S., and M. Wattenberg, 2019. The Impact of AI on Employment and Organization in the Industrial Working Environment of the Future. Proceedings of the 1st European Conference on the Impact of Artificial Intelligence and Robotics (ECI-AIR19), 141-148.
- Freund, C., and D. Weinhold, 2002. The Internet and International Trade in Services. *American Economic Review* **92(2)**, 236-240.
- Georgieff, Alexandre, and Raphaela Hye, 2021. Artificial Intelligence and Employment: New Cross-country Evidence. OECD Social, Employment and Migration Working Papers No. 265.
- Giovanni, Julian Di, and Andrei A. Levchenko, 2009. Trade Openness and Volatility. *Review of Economics and Statistics* **3**, 558-585.
- Haskel, J., and M. J. Slaughter, 2001. Trade, Technology and UK Wage Inequality. *Economic Journal* **111**, 163-187.
- Helpman, Elhanan, Oleg Itskhoki, and Stephen Redding, 2010. Inequality and Unemployment in a Global Economy. *Econometrica* **78(4)**, 1239-1283.
- Leamer, E. E., 1996. Wage Inequality from International Competition and Technological Change: Theory and Country Experience. *American Economic Review* **86**, 309-314.
- Liu, Lirong, and Hiranya K. Nath, 2013. Information and Communications Technology and Trade in Emerging Market Economies. *Emerging Markets Finance and Trade* **49(6)**, 67-87.

- Luttmer, Erzo G. J., 2007. Selection, Growth, and the Size Distribution of Firms. *The Quarterly Journal of Economics* **122(3)**, 1103-1144.
- Melitz, M. J., 2003. The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica* **71**, 1695-1725.
- Mucha, Tomasz Marcin, and Timo Seppala, 2021. Managing Firm Resources at the Digital Technology Frontier. ICIS 2021 TREOs, 51.
- Nigai, Sergey, 2017. A Tale of Two Tails: Productivity Distribution and the Gains from Trade. *Journal of International Economics* **104(C)**, 44-62.
- Pissarides, Christopher A., 1974. Risk, Job Search, and Income Distribution. *Journal of Political Economy* **82**, 1255-1268.
- Reavie, Vance, 2018. Do You Know the Difference Between Data Analytics and AI Machine Learning? Forbes.
- Varian, Hal, 2019. Artificial Intelligence, Economics, and Industrial Organization. In: *The economics of artificial intelligence*.
- World Economic Forum, 2020. The Future of Jobs Report 2020. World Economic Forum.