Essays on Employments and Wages in The Online Labour Market

Ngoc Dieu Linh Vi

Submitted to the University of Birmingham in fulfilment of the requirements for the Degree of Doctor of Philosophy in Economics

University of Birmingham

2022

Acknowledgment

My PhD journey has finally come to its end. First of all, I would like to thank my principal supervisor, Professor Oleksandr (Sasha) Talavera. He has provided me with excellent support, guidance, motivation, encouragement and positive criticism throughout the years. His advice did not only help me as a student but also as a human being. His efforts have provided me with essential skills to develop my future career.

Second, I would like to thank Professor Kimberly Scharf as my second supervisor. She has always provided me with prompt support, guidance and advice when I need it.

Third, I would like to thank my two examiners, Dr. Michael Henry and Dr. Sarah Bridges for their insightful and excellent comments for the thesis.

Lastly, I also want to express my deepest gratitude to all my family and friends for their unconditional support and love. I especially want to thank my Mum, Dad, little sister, husband, son, Anne-Marie and Rachel for supporting me not only during this journey but in my whole life. Thank you for raising me up when I am down. Thank you for making me believe in myself when I had doubts.

Abstract

This thesis consists of three empirical papers, which focus on jobs and wage setting in the online labour market. The first paper employs a sample of 38,945 gender-targeted online job vacancies in Vietnam from February 2019 to July 2020 to investigate gender differences in returns to physical attractiveness. In particular, we compare the monthly offered wage in matched vacancies with and without beauty preferences of the same characteristics among job ads directed at men and women separately. We find evidence that better-looking women enjoy a wage premium of up to 5.2 percent, whereas better-looking men do not. Earning premia for attractive women is present not only among occupations with intensive social interactions where beauty is expected to enhance performance but also among other occupations. We also document substantial heterogeneity in returns to look across education, experience levels, job positions and job locations.

The second paper uses a unique dataset of online job vacancies from February 2019 to July 2020 to examine wage dollarisation – the use of foreign currency for salary setting. We first document that employers advertise wages in US dollars to search for highly-skilled workers. Additionally, we employ matching techniques to investigate the relationship between wage and exchange rate benefits offered by jobs quoting salary in US dollars. Our results reveal that there is a complementarity between wage and exchange rate benefits. The positive wage effect of exchange rate benefits is highest in lower-position jobs as well as jobs of lower education and experience levels.

The third paper examines the wage gap between the formal and informal job sectors using online job vacancies data from February 2019 to July 2020. First, we identify a job posting as informal if it shows discrimination against the worker's (i) marital status, (ii) gender and (iii) disability. Second, we employ various matching techniques and estimate the wage gap between formal and informal employment in the matched samples. Our findings suggest that informal jobs tend to offer a lower salary than their formal counterparts. In particular, the study shows significant wage premia of 1.6 to 1.9 percent for formal jobs and wage penalties of 0.8 to 21.4 percent for informal jobs. Furthermore, we find that the wage gap between formal and informal jobs is highly heterogeneous across different job types.

Table of Contents

List of Tables

List of Figures

Chapter 1. Thesis Overview

1.1. Introduction

Over the last decade, many firms and job candidates have been relying on online job boards to advertise and find jobs. As a result, economics researchers have growingly considered the online job market as a novel source of data for studying labour market dynamics. Compared to more traditional sources, online job vacancy data offer the benefit of time and cost-effectiveness (Wade and Parent, 2001; Mang, 2012; Steinmetz et al., 2014). Particularly, while the results of official surveys on the labour market could take approximately a quarter or a year to become available, online vacancies are real-time data, which can be gathered in a much shorter time at a lower cost. Another main benefit of online job data is that the content of vacancies is more likely to provide more detailed information than traditional data sources (Kureková et al., 2015). Hence, with web-based data, we can have better insights into the labour market dynamics, job characteristics and employers' requirements for job candidates.

Various research papers seek to create a skills taxonomy from online job postings to analyse the labour market dynamics. For instance, Deming and Kahn (2018) identify 10,000 unique skill keywords and sort them into ten commonly observed and recognisable skill groups (i.e., social, cognitive, finance, software etc.). Using machine learning methods, Djumalieva and Sleeman (2018) automatically generate a network of skills and group them into skill clusters based on the frequency of co-occurrences of skills in the same job postings and their shared context.

Researchers can address deeper labour market questions from the extracted skills, such as changes in skills demand and the nature of work. For instance, Hershbein and Kahn (2018) provide evidence of upskilling – employers demanding higher-skilled workers - following the Great Recession. They also show that this upskilling, which is most pronounced in routine-cognitive occupations, is associated with rises in capital investments. Using a supervised machine learning method, Adams et al. (2020) classify types of work arrangements described in UK job vacancies. They find that the employers need flextime to reduce labour costs, and a large and unexpected increase in the national minimum wage can lead to a growth in the share of flexible and non-salaried vacancies at low-wage jobs.

Using online vacancies data, researchers can timely update how the labour market responds to exogenous shocks. For example, Forsythe et al. (2020) show a dramatic contraction in the number of US job postings over March and April 2020 due to the Covid-19 crisis across nearly all industries and occupations, regardless of the timing of stay-at-home orders. They also point out that spikes in unemployment insurance initial claims match patterns of job postings decline. Similarly, Hensvik et al. (2021) document a substantial drop of 40% in newly posted vacancies in Sweden during the three months after the Covid-19 pandemic. In additional, Javorcik et al. (2019) study the effect of the Brexit vote on labour demand in the UK. They find a substantial decline in higher-skilled online job postings after the Brexit referendum and no decline during the negotiation period.

Other studies use online vacancies to document the impacts of technological developments on the labour market. For example, employing the Burning Glass Technology dataset, Acemoglu et al. (2020) find that highly AI-exposed firms are reducing hiring in non-AI positions as well as expanding AI positions hiring. Yet, they find no significant impact of AI exposure on employment or wages at more aggregated levels, suggesting that AI is still in its infancy and not yet having large labour market consequences. Alekseeva et al. (2021) document the dramatic rise in demand for AI-related skills in job vacancies across occupations. At the firm level, they find that higher demand for AI skills is linked with higher market capitalisation, higher investments in R&D and higher cash holdings. They also find a substantial earning premium for vacancies that require AI skills and for non-AI vacancies posted by firms with a large share of AI job postings. Deming and Noray (2018) explain the shortage of STEM (i.e., Science, Technology, Engineering, and Math) workers, high initial economic returns and exit from the field over time by the technology boom, which creates new job tasks and makes old ones outdated.

Although the advantages of using online job vacancy data in economic research are undeniable, some reservations are still associated with this data source. First, it is pointed out that online job vacancies mainly target high-skilled jobs, white-collar positions and STEM (sciences, technology, engineering and mathematics) (see, for example, Gosling et al., 2004; Stefánik, 2012; Carnevale et al., 2014). Additionally, online job postings data is criticised for being lack of salary information (see Brenčič, 2012; Hall and Krueger, 2012; Marinescu and Wolthoff, 2020). Particularly, only a small proportion of job advertisements, typically between one-fifth and one-third, contain the explicitly offered salary (Faberman and Kudlyak, 2016). This lack of salary information might be because the wage can be negotiable (Hall and Krueger, 2012) or due to the employers' strategic reasons (Brenčič, 2012).

However, those concerns are less likely to be problematic in this thesis. In fact, the dataset used in this thesis consists of online vacancies collected from a leading job site in Vietnam. These job ads cover a broad range of occupations and job categories/industries at high-skilled and low-skilled levels. Moreover, the majority (i.e., 60%) of the job ads in our original dataset contain the monthly wage information.

This thesis consists of three empirical studies, which examine various aspects of job advertisements and different factors that can affect the salary offered by recruiting firms in the online labour market in Vietnam, including the requirement for good looks, the provision of a fringe benefit (i.e., exchange rate benefit) and the existence of the informal job sector.

The first study, "Gender Differences in Returns to Beauty" documents differentials in the effects of beauty requirements on the offered salary between women-targeted and men-targeted job postings. Our dataset consists of online job postings collected from a well-known job board in Vietnam from February 2019 to July 2020. There are 38,945 gender-targeted job vacancies covering 97 job titles belonging to both high-skilled and low-skilled job segments. Each job vacancy contains detailed information on the required education level, work experience, skill requirements, job position, job location and gender preference. This wealthy and well-structured dataset enables us to match jobs based on a wide range of job characteristics and to examine the heterogeneous beauty effect across different job types.

To ensure that we compare apples to apples, not apples to pears, we rely on different matching techniques to make our vacancies more homogeneous, hence more comparable. These matching techniques include propensity score matching, coarsened exact matching and text-based matching. Our regression estimations show a significant beauty premium of up to 5.2 percent for women and no beauty effect or a significant beauty penalty of up to 4.8 percent for men. This finding suggests that good-looking women are more advantaged than their men counterparts in terms of labour market outcome.

We further find that the wage premium for beautiful women exists not only among occupations with high interpersonal/social interactions, where beauty is expected to enhance job performance, but also among occupations of low interpersonal/social interactions. Moreover, we document that the beauty effect is considerably heterogeneous across different types of jobs. More specifically, beautiful women might be more advantaged in jobs with a low level of education and work experience. Such advantage gradually vanishes and no longer exists among jobs at higher skill levels.

The second study, titled "Wage Dollarisation: Evidence from Online Job Vacancy Data" examines the wage dollarisation in the online labour market, which is the use of foreign currency (i.e., US dollars) in advertising wages. The use of foreign currency in wage setting is a way to offer a non-pecuniary benefit to the employees to hedge against domestic currency depreciation or so-called exchange rate benefit. Hence, this paper aims to examine the relationship between wage and exchange rate benefits as well as the job characteristics and skill requirements associated with such benefit provision. In this research, we employ the same online vacancies dataset as the previous study, but over a longer period of time - between February 2019 and December 2020. In our sample, there are more than 170,000 job postings for more than 170 narrowly-defined occupations. Detailed information for each vacancy, including the offered salary, required education qualification, work experience, job position, job location and skill requirements, are extracted from the job ad text and included in our analysis.

The paper first aims to investigate the link between wage and non-wage compensations (i.e., exchange rate benefits). The estimation results show that there is a complementarity between wage and exchange rate benefits. Particularly, on average, a dollarised job can offer a hedge against exchange rate risks and 40% to 46% higher salary in comparison with a local-currency job. The findings are consistent when using different matching techniques (i.e., propensity score matching, coarsened exact matching and text matching). After controlling for the potential foreign firm wage effect, our results for the complementarity between wage and non-wage compensations remain robust. This finding can be explained by the efficiency wage theory, which argues that firms are willing to pay higher wages and benefits to raise the productivity of workers.

The study then proceeds to investigate potential heterogeneities in the wage effect of non-pecuniary benefits. The results reveal that while there are no significant differences in the relationship between wage and non-wage benefits across required education levels, there are substantial heterogeneities across experience categories and job positions. That is, the complementarity between wage and benefits is most prominent among vacancies requiring little work experience and at the lowest levels.

We further investigate job ads' characteristics and skill requirements offering exchange rate benefits. We find that advertising wage in US dollars is similar to hiding wages in online job postings in the sense that both strategies aim to attract high-skilled workers. More specifically, dollarised jobs are more likely to require higher education qualifications, more work experience and managerial positions. In addition, we show that wage dollarisation is mostly present in the two largest cities. In addition, firms that set wages in US dollars tend to look for workers who possess foreign language, writing and artistic skills.

The third study, titled "Formal versus Informal Employment: Evidence from Online Job Vacancy Data", aims to identify the informal employment as well as the wage gap between the formal and informal job sectors. The data used in this study is similar to that of the previous study on wage dollarisation. While there is no consensus in the literature on how to define the job in/formality, our dataset allows us to exploit four distinct and comprehensive measures to identify formal and informal jobs. In particular, we define a job vacancy as formal if the employer explicitly mentions their payment and contribution to compulsory social insurance in the job description text. Additionally, we identify a job posting as informal if it shows discrimination against the worker's (i) marital status, (ii) gender and (iii) disability.

Using a similar approach as the two previous papers, we employ various matching techniques, namely the propensity score matching, coarsened exact matching and text-based matching with the Jaccard similarity score to match job vacancies to examine the wage gap between formal and informal employment. Our findings suggest that informal jobs tend to offer a lower salary than their formal counterparts. In particular, we show significant wage premia of 1.6 to 1.9 percent for formal jobs and wage penalties of 0.8 to 21.4 percent for informal jobs.

Furthermore, the wage gap between formal and informal jobs is highly heterogeneous. Particularly, the formal-informal sector wage gap is likely to be the smallest among low-education jobs and rises as the required education increases. Hence, workers in formal jobs may be more advantaged in jobs requiring a higher level of education. The formal-informal wage differential is also heterogeneous across jobs of different required work experience and job positions, but there are no clear patterns.

1.2. Overview of the Vietnamese labour market

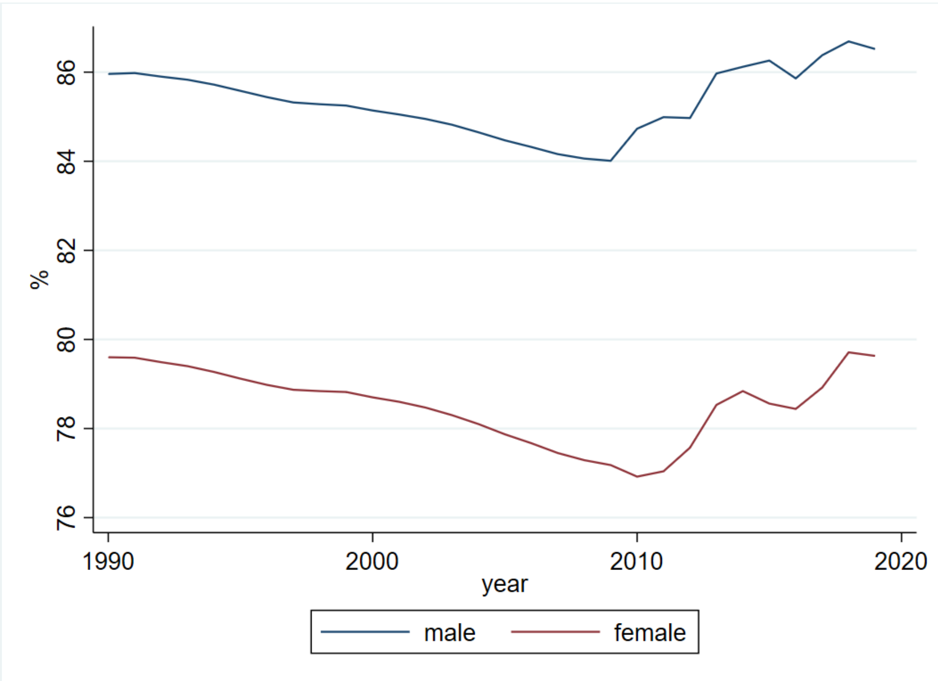
In the following section, an overview of the Vietnamese labour market will be presented, including the detailed information about the labour force participation rate,

1.2.1. Labour force participation rate

Vietnamese labour market is among the largest labour markets in Southeast Asia, with the total labour force of approximately 56 million people, which is equivalent to a labour force participation rate of 57.20 percent of the total population in 2021. However, the size of the workforce has been shrinking since 2014. The number of employed people is 1.1 million people, which is equivalent to an unemployment rate of 2.58%. The unemployment rate in urban areas is 2.64%, which is higher than in rural areas.

Figure 1.1 reports the labour force participation rate among population ages 15-64 across genders from 1990 to 2019. As we can see, the female participation rate has always been lower than that of their male counterparts and the male-female gap in labour force participation rate remains stable over the last 30 years.

Figure 1.1. Labour force participation rate among population ages 15-64[1]



Notes: The figure illustrates the labour force participation rate among population ages 15-64 for men (blue line) and women (red line).

[1] Source: https://data.worldbank.org/indicator/SL.TLF.ACTI.MA.ZS?locations=VN, accessible 4th November 2022.

1.2.2. Skill levels of the workforce

Vietnam is in the middle of an ongoing economic structural transformation that continues to boost economic growth. The economic growth during the 1990s was mostly driven by the productivity growth in the agriculture sector, coupled with the de-collectivisation of farms and tradable land use rights. Since 2000, a principal driver of economic growth has been the shift of workers out of agriculture into manufacturing and other higher productivity sectors, which led to a growth of the proportion of the population working in salaried jobs. This expansion of salaried workers has taken place together with a large increase in the level of education.

As of 2021, the number of employees in the agriculture, forestry and fishery sector is 13.9 million people, accounting for 27.9% in the total labour force. For industry and construction sector, there are approximately 16.4 million workers (accounting for 32.8%). The number of workers in service sector account for 39.3% of the total workforce (19.6 million employees). The proportion of workers with informal employment is 57.2%, in which, the ratio of workers with informal employment in rural areas is 64.5%, and the proportion of female informal workers is 52.7%.

1.2.3. Salary level

Income and salary levels in Vietnam are diverse across gender, geographical areas and occupations. According to statistics reported by the General Statistics Office of Vietnam, the average monthly income of salaried workers is VND 7 million per month (in 2021)[2]. The salary is ranging from VND 3.07 million per month (minimum salary) to the highest average of VND 76.9 million per month. The median salary in Vietnam is VND 16.2 million per month. This means that half the workers get more than VND 16.2 million per month, while the other half receives less than VND 16.2 million per month. The average income of male employees is 1.2 times higher than that of female employees (i.e., VND 7.5 million vs VND 6.4 million respectively). In addition, the average income of workers in urban areas is VND 8 million, which is 1.3 times higher than that of workers in rural areas (i.e., VND 6.3 million).

With regards to occupations, the top occupations that pay the highest salary are mineral and metallurgy (VND 9.2 million per month), banking and financial services (VND 7.6 million per month) and pharmacy (VND 7 million per month), IT, civil engineering and marketing while

[2] Source: https://www.gso.gov.vn/du-lieu-va-so-lieu-thong-ke/2021/07/thong-cao-bao-chi-tinh-hinh-lao-dong-viec-lam-quy-ii-va-6-thang-dau-nam-2021/, accessible 4th November 2022.

some other occupations in textile and food industry sectors only pay the wage level ranging from VND 2.5 to 3.1 million per month.

Furthermore, the salary level also varies across private and state-owned firms. Most employees in state-owned firms receive very low pay of around one fifth to one third of the wage of the workers in private-owned firms in the same position, especially among some occupations in education and health care industries. The salary of workers in the state-owned firms in other sectors such as banking, petroleum and energy might be higher, but still not as high as that of the private sector workers.

1.2.4. Several labour laws

Compulsory minimum wage

In November 2018, the Vietnamese government sets different minimum wage levels across the four regions. Region I, including the urban area of Hanoi and Ho Chi Minh City, has the highest minimum wage level of nearly VND 4.2 million while region IV has the lowest minimum wage of VND 2.9 million. This minimum wage policy becomes effective since the 1st of January 2019.

In particular, the minimum wage depends on the employee's region as follows:

- Region I (urban areas of Hanoi and Ho Chi Minh City): VND 4.18 million per month.
- Region II (rural areas of Hanoi and Ho Chi Minh, as well as Da Nang, Can Tho, and Haiphong cities): VND 3.71 million per month.
- Region III (Bac Giang, Bac Ninh, Hai Duong and Vinh Phuc provinces): VND 3.25 million per month.
- Region IV (other remaining areas): VND 2.92 million per month.

Additionally, workers that have vocational training must be paid a salary of at least seven percent higher than the minimum wage rate in their region.

Social securities

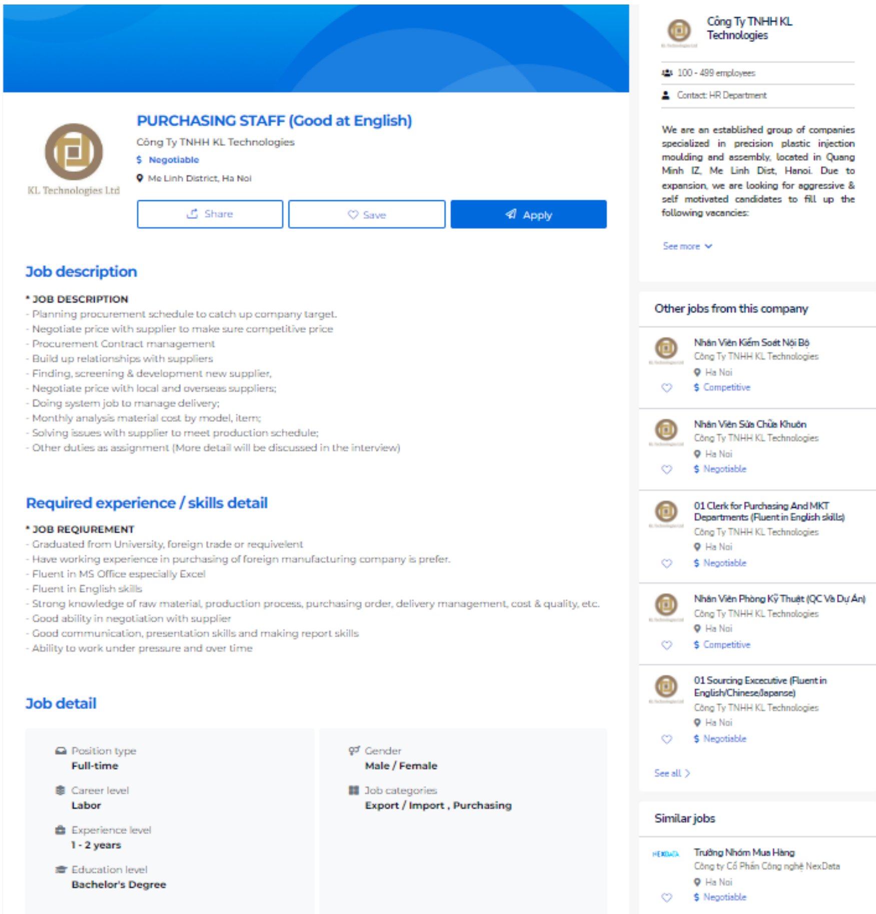
There are three types of mandatory social security that must be covered for the employment, including social insurance, health insurance and unemployment insurance. Employers register and pay insurance contributions monthly on behalf of their employees at the Department of Labour, Invalids and Social Affairs at province level. The total contribution payment is determined based on the workers' monthly salary.

1.3. Online job portal

In the following three thesis chapters, we will use a unique dataset collected from a leading job portal in Vietnam for our empirical analyses. This job board was founded in 2006 and has become one of the top job portals in Vietnam in terms of website traffic and the number of available job vacancies[3]. The data used in the first paper span from February 2019 to July 2020. The dataset used in the next two papers spans from February 2019 to December 2020. This difference in the sampling periods between chapters is because the data continued to be collected after the start of the first paper.

This job site displays information and job vacancies in both English and Vietnamese languages. The structure of a job vacancy is well-organised and enables us to apply textual analysis tools to extract relevant information on the job characteristics and skill requirements. Figure 1.2 represents the screenshot of a job posting taken from the studied job board.

Figure 1.2. Screenshot of a job vacancy



Notes: The figure illustrates an online job posting captured from the job board under study.

[3] https://www.betterteam.com/job-posting-sites-vietnam, accessed 31st July 2022.

1.4. Literature review

This section presents the literature related to the three empirical studies. Each chapter is linked with several related strands of literature.

1.4.1. Study 1

The first empirical study contributes to three main strands of literature. First, this study is related to the controversial literature on the effect of physical appearance on earnings. In a pioneering study using survey data in the US and Canada, Hamermesh and Biddle (1994) reveal that highly attractive individuals receive an earning premium of approximately five percent over less attractive ones. Since then, researchers have documented the existence of positive beauty effects in different countries, i.e., European (e.g., UK - Harper, 2000; Germany - Pfeifer, 2012), the Caucasus (Mavisakalyan, 2018), Australia (Borland and Leigh, 2014) and Asian (e.g., Bangladesh - Islam and Smyth, 2012; China – Peng et al., 2020). Beauty premium is found within a diverse array of occupations (Robins et al., 2011), as well as within specific professions, such as lawyers (Biddle and Hamermesh, 1998), real estate brokers (Salter et al., 2012), restaurant servers (Parrett, 2015) and commercial sex worker (Arunachalam and Shah, 2012). Yet, some other studies show that beauty could have no effect or lead to lower earnings, depending on the worker and the nature of the job. For instance, Deryugina and Shurchkov (2015) suggest that the attractiveness effect is task-dependent: while attractive employees receive a higher wage in jobs involving bargaining tasks, there is no such premium in jobs involving analytical or data entry functions.

We contribute to this body of literature by using a novel measure of beauty. Specifically, we use the textual analysis approach to extract the employers' requirement for good looks stated in online job postings. Prior studies obtain attractiveness measures using self-reported scores (French, 2002) and other people's rating scores based on facial photographs (e.g., Biddle and Hamermesh, 1998; Hamermesh and Parker, 2005; Mobius and Rosenblat, 2006; Scholz and Sicinski, 2015), or face-to-face contacts (French et al., 2009; Mocan and Tekin, 2010; Hamermesh and Abrevaya, 2013; Gehrsitz, 2014; Oreffice and Quintana-Domeque, 2016; Mavisakalyan, 2018). Yet, these measures are usually criticised for being subjective and biased (Bertrand and Mullainathan, 2001).

Alternative measures of physical attractiveness employed by more recent studies include height (e.g., Persico et al., 2004; Rashad, 2008), weight (e.g., Rooth, 2009; Sabia and Rees, 2012), skin tone (e.g., Goldsmith et al., 2006; Kreisman and Rangel, 2015), hair colour (e.g., Johnston, 2010; Guéguen, 2012), waist circumference (e.g., Wada and Tekin, 2010; Bozoyan and Wolbring, 2011), grooming and makeup (Robins et al., 2011; Póvoa et al., 2020). Although these measures are less prone to measurement error, they are indirect and only partially capture the concept of attractiveness (Mavisakalyan, 2018; Deng et al., 2020). Our measure of beauty is distinct from those used in previous papers since it enables us to directly uncover the employers' beauty valuation at the point of hiring.

Second, this paper contributes to the controversial literature on gender differences in the importance of beauty. Some studies show that beauty matters more for women than men. For example, a significant wage premium for attractiveness is found for women but not for men (French, 2002). Likewise, Patacchini et al. (2015) only find a significant advantage for good-looking women in terms of job application call-back rate. Other studies find the opposite results. For instance, Hamermesh and Biddle (1994) reveal that men enjoy a slightly larger looks premium than women. Similarly, Doorley and Sierminska (2015) find that while the attractiveness premium for men presents throughout the earning distribution, the premium for women is lower and only concentrated at the bottom of the earning distribution. The conclusions in those studies are mostly based on survey data, which often lack important labour market dimensions, such as skills used at work and the narrowly-defined occupation classification. The only exception is the paper of Patacchini et al. (2015), which uses experimental data. Yet, it only contains information on two skills (i.e., language and computer) across seven low-skilled occupations (i.e., administrative clerk, bookkeeper, call center operator, receptionist, sales clerk, secretary, shop assistant). We complement this literature by employing online vacancies data, which can provide more detailed job characteristics with broader occupational coverage.

Our paper is also related to the strand of literature that examines discrimination based on looks in online job postings. In one of the earliest studies of this strand, using Chinese online vacancy data, Kuhn and Shen (2013) report that employers' demand for tall and attractive workers is more likely to appear in women-targeted ads. Yet, their work is only limited to providing such statistical analysis without further analysing the offered wage. Ningrum et al. (2020) follow a similar analysis by using an online job vacancy dataset in Indonesia and suggest that the requirement for physical appearance can be found in all job categories, and such requirement occurs more frequently in female-targeted jobs compared to male-targeted or gender-neutral jobs. Why do these studies not conduct further analysis for the offered wage? The reason might be that the data used in these studies contain only a small fraction of job vacancies with salary information (i.e., 16.4% in Kuhn and Shen, 2013), which does not enable them to estimate the effect of looks on the offered wage. Hence, in this paper, we extend this literature by adding estimates of the earning effect for physical attractiveness.

1.4.2. Study 2

The second empirical study contributes to three main strands of literature. Previous studies have examined the impacts of (financial) dollarisation on different labour market dimensions. For example, Borjas and Fisher (2001) develop a theoretical model to study how dollarisation affects the real wage and aggregate employment in Mexico. They find that dollarisation can increase the relative real wage and lower the unemployment rate in Mexico. Galindo et al. (2007) look at the impact of real exchange rate fluctuations on employment growth, conditional on the degree of liability dollarisation. Soto (2009) examines the reasons to explain the sluggishness of employment growth/labour demand following dollarisation in Ecuador. Chidakwa and Munhupedzi (2017) investigate the effect of dollarisation on the unemployment rate in Zimbabwe and show that the unemployment rate is not affected by dollarisation but by other factors. However, those studies either use macro-level data on labour market indices or data on the degree of financial dollarisation from the banking sector. In our paper, we use micro-level labour market data and focus our analysis on payment dollarisation.

Moreover, there are fewer studies that examine the payment dollarisation phenomenon. For instance, De Zamaroczy and Sa (2002) document the true degree of dollarisation in the Cambodian economy and measure the currency substitution by calculating the ratio of foreign currency in circulation to total currency in circulation. Castillo (2006b) studies the effects of price dollarisation in economies featured with both sector-specific shocks and sticky prices. Drenik and Perez (2021) investigate the dollarisation of prices in retail markets of Latin American countries. They conclude that at the micro level, larger retail sellers are more likely to set prices in US dollars, and the price of more tradeable products is more likely to be quoted in US dollars. Kubo et al. (2021) use firm-level survey data to analyse the cost of currency exchange associated with the payment dollarisation and the choice of currency exchange methods by firms (exchange currency via banks versus currency changers). Yet, to the best of our knowledge, the literature has not yet provided any empirical analysis of wage dollarisation. Hence, this study will be the first to examine the payment dollarisation in the labour market (i.e., advertising wage in US dollars).

Furthermore, this paper contributes to the strand of literature on whether there is a substitution or complementarity between wage and non-wage benefits. In one view, there are numerous studies that document the positive wage effect of non-monetary benefits. For example, Johnson and Provan (1995) find that jobs offering family-friendly benefits pay on average 18% higher wages than non-benefit jobs. Gariety and Shaffer (2001) show a wage premium of 6.2% for men and 6.7% for women among jobs that offer flexible time and work arrangements. Similarly, Haynes and Sessions (2013) document that jobs in the public sector offer both pensions and 12% to 13.9% higher salaries. In the same vein, Pailhé and Solaz (2019) present evidence for a wage premium of 51.9% for women among jobs that feature a family-friendly workplace.

In contrast, the alternative view supporting the traditional compensation wage differentials theory shows a negative wage effect for non-pecuniary benefits. Baughman et al. (2003) find that firms offering non-wage benefits such as flexible sick leave and flexible working time will cover part of the cost of providing these benefits by paying 21.9% or 21.59% lower earnings than other firms, respectively. Heywood et al. (2007) also point out that the provision of family-friendly benefits might come at the cost of around 20% lower salary. Eriksson and Kristensen (2011) find that workers might be willing to sacrifice about 11% to 13% of the wage in return for a flexible working schedule. Likewise, employing the Canadian matched employer-employee data, Fakih (2014) show a trade-off between the provision of family-friendly benefits (i.e., childcare, elder care and extended healthcare) and wages. Thus, we complement this strand of literature by providing new empirical evidence on the complementary between wage and non-wage compensations.

In addition, prior studies mainly focus on several kinds of benefits that are often held fixed for a period of time according to the work contract, such as family-friendly benefits, including childcare provisions and maternity leaves (Pailhé and Solaz, 2019), paid holidays, flexible work arrangements (Baughman et al., 2003), assistance for elder care, extended health care and pension benefits (Fakih, 2014). We complement the literature by focusing our analysis on a new type of benefit – exchange rate benefit, which is not fixed and changes with the exchange rate fluctuations. That is, workers might be more beneficial when receiving US dollar-wage payments in times of high exchange rate fluctuations than in the time of low exchange rate volatilities.

Furthermore, our paper is also related to the literature that documents employers' wage setting strategies in online job postings. Prior papers pay particular attention to the differences in skill requirements and job characteristics between explicit wage posting and hidden wage posting. For example, Banfi and Villena-Roldan (2019) show that job ads with explicit wages tend to target unskilled workers (i.e., low education level). In the same vein, Brenčič (2012) shows that employers in three job markets (i.e., US, UK and Slovenia) are less likely to post a wage offer when searching for skilled workers.

Lastly, some studies look at the types of employers/firms that are more likely to adopt such hidden wage posting strategy. For instance, according to Michelacci and Suarez (2006), employers opt to post wages when their bargaining powers are low and the workers' productivities are highly homogeneous. Also, firms that belong to the public sector or are unionised are usually forced to post wages explicitly. Using the US employee survey, Hall and Krueger (2012) find that wage information is usually known among union or government jobs. They also find a negative relationship between education level and precise information about salary. Likewise, Brenzel et al. (2014) point out that wage posting dominates in the public sector, in larger firms and in jobs involving part-time and fixed-term contracts using data from a German vacancy survey. Yet, little is known about the strategy of setting wages in foreign currency. In this paper, we have information on wages quoted in US dollars, and dollarised wage posting can be seen as a type of hidden wage posting since the wage level is not fixed and depends on the exchange rate fluctuations.

1.4.3. Study 3

The third empirical study is related to two main strands of literature. The first strand of literature documents the wage gap between formal and informal jobs. On the one hand, some studies show that the formal sector offers higher wages than the informal sector. For example, employing data from Ecuador, El Badaoui et al. (2010) show that formal sector workers are paid on average 40% higher salary than their informal sector counterparts and explain this result by the firm size wage effect. Lehmann and Zaiceva (2013) provide evidence of a wage penalty in the lower part of the distribution and no statistically significant difference between informal and formal wages in the upper half of the distribution for salaried workers in Russia. Similarly, Bargain and Kwenda (2014) also find significant wage penalties for informal jobs in Brazil, Mexico and South Africa.

On the other hand, other studies find the wage premium for the informal employment sector. For instance, Braithwaite (1995) shows that in Russia, employments in the informal sector offer higher income than those in the official (public) sector. Similarly, after controlling for observable characteristics of formal and informal employees, Staneva and Arabsheibani (2014) still find a wage premium of up to 14.4 percent for informal sector workers in Tajikistan. Evidence in existing studies is mainly drawn from survey data, which might lack many important job dimensions, such as narrowly-defined occupations and the skills used at work. In this study, we aim to provide additional evidence on the wage premium of formal sector jobs and wage penalty for informal sector jobs. We contribute to the literature by employing a more frequent and granular vacancy-level dataset with more detailed job characteristics and narrowly-defined occupations.

Last but not least, it complements the controversial literature on how to define informality in the labour market. Among others, there are studies that rely on firm size to define informality, such as using the threshold of ten or less employees (Khamis, 2012; Lehmann and Zaiceva, 2013) or the threshold of fewer than five employees (Benjamin and Mbaye, 2012; Cuadros-Meñaca, 2020). Some other studies rely on the availability of a labour contract or the payment of social security contributions to identify formal/informal jobs (Henley et al., 2009; Park et al., 2013). Yet, most measures in prior literature are obtained from survey data of employers and employees. In this paper, our contribution is proposing four distinct and complementary definitions for formal and informal jobs among salaried employment segments, based on a textual analysis of the vacancies advertised by firms. These measures enable us to discover new dimensions of job informality, which cannot be easily observed using existing data sources.

Chapter 2. Gender Differences in Returns to Beauty[4]

[4] Some materials and results of this chapter have been submitted as the final assignment of the module Advanced Research Methods (24404). Module leader: Professor Alessandra Guariglia. Module period: 2019-2020 academic year.

2.1. Introduction

Physical appearance has been widely documented to play a significant role in many aspects of life. For instance, in terms of trust and credibility, Wilson and Eckel (2006) point out that beautiful people are viewed as more trustworthy than unattractive ones in trust games, and Chaker et al. (2019) show that physically attractive sales managers are perceived as more credible by salespeople. In terms of career and professional outcomes, beautiful individuals can have higher status in the workplace (Hamermesh, 2006) and a higher probability of employment (Mavisakalyan, 2018), and even in academia, attractive economic researchers receive more citations for their papers (Hale et al., 2021). In other aspects, good-looking people can obtain a higher loan approval rate (Cheng et al., 2020), more success in charitable fundraising (Landry et al., 2006) and higher vote share in political elections (Jones and Price, 2017). Given the importance of physical attractiveness in human life, it is critical to understand how beauty affects individuals' earnings.

Previous studies have found mixed evidence on the effects of physical appearance on wages. While some scholars document the presence of beauty premium among a diverse array of occupations (Robins et al., 2011), some show that beauty could have no effect or lead to lower earnings, depending on the worker and the nature of the job. For instance, Deryugina and Shurchkov (2015) suggest that the attractiveness effect is task-dependent: while attractive employees receive a higher wage in jobs involving bargaining tasks, there is no such premium in jobs involving analytical or data entry functions. In this paper, we present novel evidence of gender-dependent and job level-dependent beauty effects. We contribute to the body of beauty premium literature by utilising a novel measure of good looks obtained through extracting the employers' beauty preferences mentioned in online job postings.

Specifically, prior studies' attractiveness measures include self-reported scores and other people's rating scores based on facial photographs or face-to-face contacts. However, they are usually criticised for being subjective and biased (Bertrand and Mullainathan, 2001). More recent studies rely on measures that are less prone to measurement error, such as height, weight, skin tone, hair colour, waist circumference, grooming and makeup. Yet, these measures are indirect and only partially capture the concept of attractiveness (Mavisakalyan, 2018). Our measure of beauty is distinct from previous papers in enabling us to directly uncover the employers' "beauty pricing" at the point of hiring.

Our empirical exploration relies on job ads from the Vietnamese labour market, which represents a promising case for exploring the gender differentials in returns to physical appearance for two reasons. First, while being uncommon in most developed countries due to stricter legal frameworks, stating a preference for a specific gender or physical appearance requirement in job postings remains persistent in many developing countries (Arceo-Gomez and Campos-Vazquez, 2019; Ningrum et al., 2020). In Vietnam, employers are not penalised for advertising their preferences for physical attractiveness and particular gender in online job vacancies. As a result, during our seventeen-month sampling period, we observe a non-trivial occurrence of physical attractiveness requirements in job ads (i.e., 12.3%) and a high occurrence of gender-biased ads (i.e., nearly one-third). Second, while many prior studies employ vacancies data with a small fraction of job ads explicitly quoting salary (e.g., Brenčič, 2012, Marinescu and Wolthoff, 2020), more than half of our vacancies (i.e., 52.41%) provide information on the offered monthly wage.

Our dataset consists of online job ads, automatically collected from a leading job board in Vietnam, over the period February 2019 - July 2020. More than 38,945 gender-targeted job vacancies cover 97 job titles[5] belonging to both high-skilled and low-skilled job segments. Detailed information on required education level, experience level, job position, location and gender preference are also provided. Hence, this well-structured and rich dataset allows us to match jobs based on a wide range of characteristics and to uncover the heterogeneous beauty effects across different occupation types.

To examine the differentials in beauty effect across genders, we employ propensity score matching (PSM), coarsened exact matching (CEM) and text matching with the Jaccard algorithm. Our PSM and CEM estimations show a significant beauty premium of up to 5.2 percent for women and a significant beauty penalty of up to 4.8 percent for men. Similar patterns are found using the Jaccard text similarity index. The beauty effect is 3.3 percent for women, and no significant beauty effect is found for good-looking. These findings suggest that physical attractiveness plays a more critical role for women than men.

We further show evidence of an earning premium for good-looking women, not only among occupations with intensive interpersonal interactions where beauty is explicitly a productivity-enhancing trait but also among other occupations. In addition, we find that the beauty effect is substantially heterogeneous. Particularly, beautiful women might be more advantaged in jobs

[5] The full list of job titles is reported in Online Appendix Table A2.1.

requiring a low level of education and experience. For employment of higher skill levels, such premium becomes smaller and gradually vanishes.

This paper contributes to the controversial literature on gender differences in the importance of beauty. Several studies point out that attractiveness matters more for women than men. For instance, a significant wage premium for attractiveness is found for women but not for men (French, 2002). Likewise, Patacchini et al. (2015) only find a significant advantage for good-looking women in terms of job application call-back rate. Several studies suggest otherwise. For example, Hamermesh and Biddle (1994) reveal that men enjoy a slightly larger looks premium than women. Similarly, Doorley and Sierminska (2015) find that while the attractiveness premium for men is present throughout the earning distribution, the premium for women is lower and only concentrated at the bottom. Yet, the conclusions in those studies are based on survey data, which often lack important labour market dimensions, such as skills used at work and job types. These sources of wage disparities (Christl and Köppl–Turyna, 2020) can be explored using job vacancy data. We complement this literature by employing a highly frequent and granular micro-level dataset, which allows us to shed new light on different roles of physical attractiveness across genders in the labour market.

Furthermore, we contribute to the broad literature that makes use of publicly available online vacancies to study the labour market. Vacancy data collected from a variety of online job boards in developed countries have been widely used to discover job posting and job search behaviours (e.g., Brenčič and Norris, 2012; Faberman and Kudlyak, 2019; Marinescu and Wolthoff, 2020). Employing aggregated job postings data provided by Burning Glass Technologies (BGT), numerous studies document dynamics of skill demand across firms and occupations, as well as labour market conditions in the US and the UK (e.g., Modestino et al., 2016; Acemoglu et al., 2020; Javorcik et al., 2019; Grinis, 2019). Recently, online vacancies literature has been extended to include more labour market analyses for emerging countries (Nomura et al., 2017; Ahmed, 2018; Matsuda et al., 2019; Hayashi and Matsuda, 2020).

Among emerging countries-focused analyses using online job vacancies, our paper is specifically related to the strand of studies that investigates discrimination based on physical appearance. Employing Chinese online vacancy data, Kuhn and Shen (2013) only conduct statistical analysis and report that employers' demand for tall and attractive workers is more likely to appear in women-targeted ads. Later, using an online job vacancy dataset in Indonesia, Ningrum et al. (2020) suggest that requests for physical appearance can be found in all job categories, and such request occurs more frequently in female-targeted job ads compared to male-targeted or gender-neutral job ads. However, the data used in previous studies contain only a tiny fraction of job vacancies with salary information (i.e., 16.4% in Kuhn and Shen, 2013), which might be a possible reason for the lack of estimations of the wage premium for looks. Hence, we extend this literature by adding estimates of the earning effect for physical attractiveness.

The remainder of the paper is structured as follows. Section 2 describes the dataset employed in our paper. Section 3 presents empirical strategies. Section 4 discusses the findings. Finally, Section 5 concludes and highlights the implications.

2.2. Data

2.2.1. Data collection and process

Over the last decade, online vacancy data have been shown to bring about many advantages for labour economics research. In comparison with traditional channels such as firm and household surveys, collecting web-based data has the advantage of time and cost-effectiveness (Wade and Parent, 2001; Steinmetz et al., 2014; Mang, 2012). Additionally, thanks to both nature of the data (i.e., granular and high frequency) and new textual analytic methods (e.g., Natural Language Processing), the content of online job posts can provide novel and more detailed information than those provided by traditional newspaper sources (Kureková et al., 2013). In this paper, we make use of this wealthy source of labour market data.

Our dataset includes publicly available job vacancies collected from one of the top Vietnamese online job boards over the seventeen-month period from February 2019 to July 2020. We wrote a Python script to automatically scrape vacancies from this job portal every week. A typical job vacancy contains detailed information, including job title, category, job level, job type, work location, job description, preferred gender, education level, experience level, job requirement, offered monthly salary, the firm's name and a number of employees.

The data are then processed with the following steps. Vacancies that do not contain any salary information are excluded from our dataset[6]. Next, we convert the salary amount of those with salaries quoted in US dollars to Vietnam dong using the daily exchange rate[7] at the time when

[6] Instead of stating a specific wage number or a wage range, these vacancies only state that the salary is negotiable or competitive.
[7] USD - VND exchange rate data were downloaded from website https://www.investing.com, accessible on 31st July 2022.

the job ad is first posted. For job ads quoting a range of salaries, the minimum wage level is selected[8]. Furthermore, we restrict our analysis to full-time jobs and those that offer a salary of at least as much as the obligatory minimum wage[9]. Finally, our last step is removing duplications. Since we cannot keep track of whether a vacancy is filled or not, we rely on the ID number assigned to each job posting to identify unique vacancies and drop duplications.

In our sample, 12.3% of the job posts mention physical attractiveness as a requirement. Besides physical appearance, 28.4% of vacancies mention the employer's explicit demand for a specific gender. Among these gender-targeted ads, nearly two-thirds of job ads report a preference for male candidates. Since the focus of our analysis is gender differentials in the beauty effect, we restrict our sample to vacancies that state preference towards a specific gender (i.e., male- or female-targeted ads). In total, we have a sample of 38,945 unique gender-targeted job ads.

We conduct further cleaning steps at the vacancy level. According to Marinescu and Wolthoff (2020), the job title can explain a large proportion of the variation in posted salary, as well as reflect the level of education, level of experience and specialisation of different jobs. Hence, we clean the job title text of each job vacancy and then group them into 97 different narrowly-defined occupations. We also process the job industry text to achieve a set of 43 industries. Additionally, we define firm size based on the number of employees. As shown in Figure 2.1, most jobs are advertised by medium-sized companies of 100 to 499 employees.

[Insert Figure 2.1 here]

For work location, we extract information on city/province where the job locates. As can be seen from Figure 2.2, most jobs are located in Hanoi and Ho Chi Minh city (HCM). This is not surprising, given that those cities are the two most important cities in Vietnam in terms of politics and economics. Other cities that also share a high number of vacancies, such as Da Nang, Hai Phong, Binh Duong or Bac Ninh, are industrialised or port cities. We further classify vacancies into three location groups: Hanoi, HCM and other cities.

[Insert Figure 2.2 here]

Furthermore, we extract skill demand from the job requirement text. More specifically, we use a Python package on Natural Language Processing to extract a set of all unique keywords from

[8] Results for wage equations (reported in Online Appendix Table A2.3 and A2.4) remain robust when we choose the average wage level and maximum wage level instead.
[9] The minimum wage is set at VND 2,920,000 according to the Decree number 157/2018/ND-CP (effective since the 1st of January 2019).

the unstructured text of job requirements. From this set of keywords, those that indicate skills are selected and categorised into twelve different skill groups. We follow Deming and Kahn (2017) in classifying skills as cognitive, social, character, writing, customer service, project management, people management, financial, computer (general) and software (specific). Also, two additional skill groups, including foreign language and art, are added to better capture the skill demand in the Vietnamese labour market. We consider a vacancy as having a particular job skill if it contains one or more keywords or phrases listed in that skill group. Table 2.1 presents the list of example keywords for each skill category.

A similar approach is applied to obtain the physical appearance requirement. Particularly, we classify a job posting as stating beauty preference if it contains one or more of the following keywords: pretty face, nice body, attractive, good looking, etc.

[Insert Table 2.1 here]

2.2.2. Descriptive statistics

Table 2.2 reports bivariate correlations for each pair of job characteristics and requirements, from which we can draw some findings. Firstly, cognitive, writing, people management and project management skills are positively correlated with each other, with education and years of experience and with most of the other nine job skills. This finding implies that they are general skill groups which are considered necessary by firms across a wide range of occupations and industries. Secondly, the correlations between attractiveness and education, experience requirements, as well as job position are mostly negative, which suggest that beauty preference is more likely to appear in low-skilled job segments.

[Insert Table 2.2 here]

Table 2.3 provides summary statistics for subsamples of women- and men-targeted vacancies. On average, jobs preferring women offer 12.7% lower salaries (i.e., VND 7,752,873) than jobs preferring men (i.e., VND 8,882,292). Only a marginal proportion of job vacancies for men (i.e., 0.55%) and for women (i.e., 0.2%) do not mention a specific education level. Among those that specify education levels in both men and women subsamples, nearly one-third of the job ads require an associate degree, and one-fifth require a bachelor’s degree.

Many gender-targeted vacancies are available for early-career candidates and those with zero to little work experience. Particularly, around 40% of job postings ask for less than one year of experience. The number of vacancies decreases as the required year of experience increases.

With regard to job level, the majority of our postings (i.e., 84.72% and 88.73% for men's and women's jobs, respectively) look for candidates to fill in non-manager/employee positions, and only a limited number of jobs advertised for new entry workers, manager to top-manager positions. As mentioned earlier, firms in big cities are more likely to state gender preferences, with two third of vacancies for men and three-quarters of the vacancies for women located in Hanoi and Ho Chi Minh city. These statistics suggest that gender-biased jobs mainly target young and educated candidates in the two largest cities.

[Insert Table 2.3 here]

Regarding detailed statistics on the frequency of requiring beauty in gender-targeted job ads across job types in Table 2.4, beauty preference is more likely to appear in jobs requiring a lower level of education, experience and those of non-managerial positions. For instance, among men-targeted ads, 20% of those looking for intern/entry-level employees require beauty when, while less than 2% of those looking for top-managers mention such requirement. Similarly, among women-targeted ads, more than a quarter of those looking for non-managerial levels request beautiful candidates when nearly none of those looking for top managers have such preference. In terms of location, jobs in the two big cities request beauty as often as those in smaller cities.

Interestingly, statistics in Table 2.4 also reflect the gender gap in job's requirements for physical attractiveness. In general, the share of female-targeted ads looking for beautiful candidates is much higher than that of male-targeted ads across all job types. For example, among the ads requesting one to two years of experience, the occurrence of attractiveness preference in ads targeting women is ten times higher than that in ads targeting men (i.e., 32.54% vs 3.28%).

[Insert Table 2.4 here]

Further, the statistics on the frequency of requiring beauty by job industries are presented in Figure 2.3. Not surprisingly, the frequency of stating beauty requirement for women's jobs are much higher than for men's jobs in the majority of industries, except for the IT sector where the frequency of requiring good looks are substantially higher for men[10]. Overall, these statistics suggest that physical appearance plays a more important role for women than men in most job sectors. This is consistent with Kuhn and Shen (2013) and Ningrum et al. (2020), who

[10] It is noteworthy that the IT job category is mostly made up by customer IT support technician occupation, which is mainly men-targeted and requires good-look very often (since it involves substantial customer interactions).

document that employers are more likely to put beauty requests in female-targeted job ads compared to male-targeted job ads.

[Insert Figure 2.3 here]

2.3. Empirical strategy

2.3.1. Characteristics of jobs requesting a physical appearance

In this section, we investigate specific job requirements/characteristics of a vacancy that prefers beautiful employees by using the following Probit model:

$$P(Attractiveness_{it} = 1|X) = \Phi(\beta_0 + Skill_{it}\beta_1 + Controls_{it}\beta_2 + Quarter\ dummies + Industry\ dummies + Occupation\ dummies\ + Firm\ size\ dummies + \varepsilon_{it}) \quad (2.1)$$

where i and t refer to job posting and time. Φ is the cumulative standard normal distribution function, that is: $\Phi(z)=P(Z\leq z)$ and Z follows a standard normal distribution N (0,1). *Attractiveness* is a dummy variable that equals one if the employer states a preference for physically attractive candidates and zero otherwise. *Skill* is a vector of 12 skill dummy variables (i.e., cognitive, social, character, writing, customer service, project management, people management, financial, computer, software, language and artistic), each of which indicates whether a certain skill appears in the job requirement. Moreover, we include in the model a set of control variables, including dummies for experience level, education level, work location and job level. We also add quarter dummies, job title dummies, industry dummies and firm size dummies into our model. Detailed definitions of our variables are reported in Online Appendix Table A2.2.

2.3.2. Effects of beauty on earnings

This section will document the wage return to physical appearance for women and men. As shown in our summary statistics, jobs with beauty preference and those without beauty preference possess different characteristics and requirements. Hence to examine the beauty premium, we employ matching techniques since beautiful applicants might self-select into the application process, we need to control for any possible selectivity bias. Indeed, the purpose of the matching approach is to balance the unobservable job characteristics by adjusting for differences in observable characteristics among occupations (Staneva and Arabsheibani, 2014).

2.3.2.1. Propensity Score Matching

The first matching approach is Propensity Score Matching (PSM) (Rosenbaum and Rubin, 1983). This matching technique allows us to identify a group of vacancies without beauty preference that exhibit the same characteristics as a group of vacancies with beauty preference based on a single score. Hence, a pair of matched vacancies is considered identical apart from only one key characteristic, which is the demand for attractive candidates.

First, we compute the propensity score (i.e., the probability) that a job with given characteristics states the beauty requirement using the Probit model (2.1). Next, we match each vacancy requesting beauty (i.e., treated unit) with vacancies that do not have such a request (i.e., control unit) of a similar propensity score. We use the nearest neighbour matching algorithm (i.e., one-to-one matching), which matches a control unit that has the closest propensity score to each treated unit. We run PSM with no replacement, which means that a vacancy in the control group cannot be used more than once. We impose common support and use a caliper of 1%. To this end, we obtain a balanced sample of vacancies in terms of education level, experience level, job location, position, industry, occupation, required skills, posted date and firm size. In total, our matched samples contain 4,003 and 1,461 job vacancies for women and men subsamples, respectively.

As the final step, we apply the double-robust approach (i.e., running the wage equation on matched sample to further control for any remaining covariate imbalances after matching) using the following fixed effects model:

$$LogWage_{it} = \beta_0 + \beta_1 Attractiveness_{it} + Controls_{it}\beta_2 + Fixed\ effects + \varepsilon_{it} \quad (2.2)$$

where i and t refer to job ad and time. $LogWage_{it}$ is the logarithm of offered salary for job posting i posted on date t. *Attractiveness* is a dummy variable which equals one if the employer prefers physically attractive candidates and zero otherwise. *Controls* is a vector of covariates, including job characteristics that are likely to affect wage (i.e., 12 skill groups, education and experience level, job position and location). These control variables can be commonly found in wage determination studies (Rosen, 1976; Brown, 1980). *Fixed effects* include time (i.e., the quarter when the job ad was first posted), industry, job title and firm size fixed effects. ε_{it} is an error term.

2.3.2.2. Coarsened exact matching

We further implement a coarsened exact matching (CEM) approach (Iacus et al., 2012) where a control group of job ads without beauty preference is chosen having the same characteristics as job postings with beauty preference. Those characteristics include narrowly-defined job title, required education, experience, job position, posted quarter and firm size. Since all these matching covariates are categorical, our CEM algorithm is similar to exact matching because it matches a treated unit to control units with the same covariate values. This matching method advances PSM in terms of simplicity since it bounds the imbalance ex-ante, hence does not require checking for balance ex-post like PSM does. Particularly, when implementing PSM, we should always check for covariate balance after matching. If the balance has not been achieved, we need to re-specify the matching model and recheck the balance. This process repeats until we obtain an acceptable amount of covariate balances.

In our CEM algorithm, an exact one-to-one matching is used. As a result, after matching, we obtain 2,770 women-targeted and 1,154 men-targeted job vacancies, which are balanced in terms of job characteristics. Finally, we re-run model (2.2) on the matched samples to estimate the beauty effects on earnings for women and men, separately.

2.3.2.3. Text matching

We use an alternative matching approach – text matching to investigate the return to physical attractiveness. Since text matching can outperform other matching methods in performing tasks related to the text (Roberts et al., 2020), it is increasingly employed in recent research papers. For instance, text matching is applied in comparing the technological similarity between patents (Arts et al., 2018), matching job seekers' CVs and vacancies (Chala et al., 2018) or measuring product similarity based on the product's description text (Hoberg and Phillips, 2016).

We conduct text matching by looking for a job not stating beauty preference, which has the most similar job descriptions (i.e., the whole text of a job vacancy) as the job stating it. First, we remove all stop words[11], which do not add additional meaning to a text, then convert all the text to lowercase and remove special characters as well as punctuations. Second, we pre-match job postings based on 112 narrowly-defined occupations. Next, we employ the Python package *TextDistance* to measure the semantic distance between two texts. More specifically, we apply the Jaccard similarity algorithm to compute the text similarity, which can be described as follows:

$$Jaccard(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (2.3)$$

Where $|A \cap B|$ is the intersection of document A and document B and $|A \cup B|$ is the union of document A and document B. Hence, the Jaccard index is the ratio of common words in two

[11] Some examples of stop words are: a, an, the, to, but, how, what, etc.

texts to the sum of words that are present in either of the two texts. The value of this index is between zero and one. If the index equals one, two documents are identical. The lower the score, the more different two documents are.

For each vacancy in the treated group, we compute the similarity indices between itself and each of the vacancies in the control group and then select the job posting in the control group that has the highest Jaccard score. To ensure that our jobs are considered identical, we exclude matched pairs of vacancies with a similarity score of less than 30% from our analysis. This process results in samples of 2,753 and 1,909 text-matched women-targeted ads and men-targeted ads, respectively. Finally, we re-run model (2.2) on these text-matched samples of women's and men's jobs.

2.3.3. Beauty effects across jobs with high and low social and customer interactions

One may argue that the beauty effect (if any) is mainly driven by occupations with high interpersonal interactions since beauty can enhance productivity in jobs involving intensive social interactions and customer service. The productivity-enhancing effect of beauty among those jobs might come from customer/people discrimination - the preference of customers and other people to interact with more physically attractive employees in a company. In this context, beauty represents a form of human capital that is beneficial to the firm (Reingen and Kernan 1993, Pfann et al. 2000), hence the estimated beauty premium might merely reflect the productivity effect of beauty.

In this section, we examine the wage effect of attractiveness across both occupations with high and low interpersonal interactions. Particularly, we estimate the following model on the propensity-score matched samples:

$$LogWage_{it} = \beta_0 + \beta_1 Attractiveness_{it} + \beta_2 Social\_interact_{it} + \beta_3 Attractiveness_{it} \times Social\_interact_{it} + Controls_{it}\beta_4 + Fixed\ effects + \varepsilon_{it} \quad (2.4)$$

where i and t refer to job posting and time. $LogWage_{it}$ is the logarithm of offered salary for job ad i posted on date t. *Attractiveness* is a dummy variable which equals one if the employer prefers physically attractive candidates and zero otherwise. $Controls$ is a vector of other job characteristics and skills. *Fixed effects* include time (i.e., the quarter when the job ad was first posted), industry, job title and firm size fixed effects. ε_{it} is an error term.

We run model (2.4) using two different measures of $Social\_interact$. We use the social skill and customer service skills mentioned in the job ad as two separate proxies for the high level of social and customer interactions. Thus, $Social\_interact$ is a dummy variable that represents a job with a high level of social interactions or customer service. The reason we choose social and customer service skills stated in the vacancy instead of job titles to classify occupations with a high level of social/customer interactions is that the requested skills line up better with the content of the tasks performed on the job than the job titles or narrowly-defined job categories (Deming and Kahn, 2018).

2.3.4. Heterogeneity of return to beauty

It is suggested that physical appearance may be more relevant for some jobs than others (Gilmore et al., 1986; Johnson et al., 2010; Stinebrickner et al., 2019). Hence, in this section, we further explore how the physical appearance effect varies across different job types. First, we document the heterogeneity across education levels by interacting the attractiveness indicator with education categories. Second, we document the heterogeneous beauty effect across experience levels by interacting the attractiveness indicator with experience categories. Third, we add interaction terms between *Attractiveness* variable and job position dummies to examine the heterogeneous beauty effect across positions. Finally, we include interaction terms between *Attractiveness* variable and job locations to investigate the heterogeneous beauty effect across geographical areas.

2.4. Common support and balance diagnostics for PSM and CEM

2.4.1. Common support and balance diagnostics for PSM

In this section, we check the common support assumption for our PSM models. This assumption implies that any job characteristics found in the treatment group can also be found in the control group (Bryson et al., 2002). It is necessary as the treatment effect cannot be estimated outside the area of common support. Following Caliendo and Kopeinig (2008), based on the minima and maxima of the propensity score distributions of the treated and control groups, we discard observations in one group with a propensity score lower than the minimum or higher than the maximum score of the other group. To do so, while running PSM, we impose the common support to discard observations in treated and control groups outside the common support.

In the next step, we examine the matching quality by checking the balance between the treatment and control groups. In other words, we compare the distribution of job characteristics between the treatment and control groups in the matched and unmatched samples. The more similar the covariate distributions of the treatment and control groups are, the better the matching quality. Following Austin (2011), we consider the standardised difference score, which compares the difference in mean of job characteristics between the treated and untreated groups and is not affected by sample size. Normand et al. (2001) suggest that a standardised difference of equal or less than 10% represents a trivial difference in the mean of job characteristics between treated and control groups, meaning the balance has been achieved.

Figure 2.4 reveals that most variables obtain balance after matching, and only a few variables exceed this threshold. In particular, for the women-targeted job sample, only the social skill variable has the standardised difference of just over 10% (i.e., 11.9%), which is considered a moderate imbalance. For the men-targeted job sample, only 1-2 years of experience, Hanoi-located, computer and customer service skill variables are unbalanced with the standardized bias of 12.2%, 16.1%, 10.6% and 19.3%, respectively. However, this might not be a concern since our regression adjustment (i.e., running the wage regression including unbalanced covariates after PSM) can remove residual confounding bias if there are still some covariate imbalances after PSM (Nguyen et al., 2017).

[Insert Figure 2.4 here]

2.4.2. Balance check for CEM

For CEM, Table 2.5 shows that the resulting amount of imbalance after matching is trivial. The overall imbalance is illustrated by the L1 statistic, which represents the percentage of overlap between the density of the two distributions of treated and control groups. The value of zero (i.e., L1 = 0) indicates perfect global balance, higher values indicate a higher imbalance between treatment and control units, and the maximum value of one indicates complete separation. We observe a substantial reduction in the L1 statistic after matching (i.e., from 0.8243 to 0.1596 for the women subsample and from 0.93 to 0.0451 for the men subsample). This means that we obtain the overall increase in covariate balance as a result of our matching solutions.

We also report unidimensional measures of imbalance computed for each variable separately. As can be seen, our matchings for both women and men subsamples achieve perfect balance in the means, marginal and joint distributions of almost all job characteristics. The only exception is the job title, where the imbalance remains at negligible levels.

[Insert Table 2.5 here]

2.5. Result Discussion

2.5.1. Characteristics of job ads requiring beauty

Estimations from the Probit model (2.1) reported in Table 2.6 suggest that jobs at higher education levels are less likely to request beauty. For instance, among female-targeted ads, a job requiring an Associate degree and Bachelor degree have 1.9% and 3.4% lower chances of preferring beautiful candidates than a job requiring a High-school graduate, respectively. Similar results can be found among male-targeted ads with a decline of eight to ten percent in the probability of stating beauty when an educational level higher than High school is required.

Turning to experience level, among women's jobs, jobs requiring 1-2 years of experience exhibit 2.8% higher likelihood of stating beauty preference than those requiring very little work experience (i.e., less than one year). However, when the required experience exceeds five years, the chance of mentioning physical attractiveness reduces by 6.1% compared to jobs for very early career candidates with less than one year of experience. Clearer negative effects of required work experience on the probability of requesting attractiveness can be observed among the male-targeted vacancy sample. That is, compared to jobs requiring zero to little work experience of less than one year, jobs requiring 1-2 years of experience have 2.5% lower likelihood of targeting attractive candidates. The estimated effect is even double when 5-10 years of experience is required, with 5% lower chance of requesting beautiful candidates. In general, these findings suggest that jobs for highly experienced candidates are less likely to have beauty preferences.

For job positions, vacancies at managerial levels are less likely to seek good-looking candidates. Particularly, compared to entry-level jobs, the preference for attractive candidates for vacancies in managerial roles is 13.1 percent and 10.5 percent lower among female- and male-targeted jobs, respectively. With regards to work location, interestingly among women's jobs, those located in the capital - Hanoi have a three percent lower likelihood of targeting beautiful job seekers. In contrast, among men's jobs, those located in Hanoi have a 4.2 percent greater likelihood of preferring good-looking candidates.

With respect to other skills demand in female-biased ads, jobs requesting general work skills (i.e., computer), managerial skills (i.e., project and people management) and soft skills (such as character, social and writing skills) are more likely to mention beauty requirement. In contrast, those requesting hard skills such as software, cognitive and language skills have lower chances of preferring attractive candidates. This indicates a sorting of beauty preference, which appears more often in job tasks/occupations involving more interpersonal interactions. For male-biased ads, it seems that the occurrence of beauty preference is more diverse across job tasks/occupations. Particularly, job postings requiring some hard skills (i.e., financial and cognitive skills) and soft skills (i.e., character and social skills) have a higher propensity to prefer attractive candidates. In comparison, vacancies requiring project management, artistic, customer service and writing skills are less likely to add physical appearance requirements.

[Insert Table 2.6 here]

2.5.2. Beauty effects on earning

Regarding PSM results, as reported in column (1) of Table 2.7, we find that beautiful women are offered 5.2 percent higher salaries in jobs targeting female candidates. In contrast, according to column (4) of the same table, there is a wage penalty of 4.1% for beautiful men in jobs targeting male candidates. We obtain qualitatively similar results when implementing CEM in columns (2) and (4). More specifically, a slightly smaller wage premium of 3.5 percent is found for attractive women and a wage penalty of 4.8 percent for their male counterparts. Results from the text matching technique reported in columns (3) and (6) are quite consistent with those produced by PSM and CEM, except for no statistically significant beauty effect for men. In particular, we found a positive wage return of 3.3% for physically attractive female candidates but no statistically significant earning effect for their male counterparts.

[Insert Table 2.7 here]

In general, the results from our estimations show that women benefit more from physical attractiveness than men. This finding is generally consistent with prior literature (see, e.g. French, 2002; Patacchini et al., 2015; Abueg et al., 2020; Babin et al., 2021). Such gender differential in beauty effect can be explained by two potential theories which are widely documented in the literature. The first theory starts from different traditionally social roles that have been defined for women and men. In particular, while women had been identified with the traditional role of wife, housekeeper and child-bearer, their male counterparts had been expected to provide economic support for their family (Anastasi, 1981). Bar-Tal and Saxe (1976) argue that physical attractiveness plays an important role for women since it is a major criterion for their traditional social roles.

Despite dramatic changes in women's socioeconomic standing, studies show that the women's beauty ideal is still pervasive and continues to occupy a central role in their lives (Baker-Sperry and Grauerholz, 2003). Moreover, Jeffreys (2014) argues that women's growing social and economic status has been accompanied by increasingly beauty practices (i.e., makeup, diet, plastic surgery). As a result, beauty remains an advantageous trait for women in the labour market. While beauty is a highly desirable quality for women, it is considered less important for men and their traditional roles. In contrast to femininity, masculinity is more associated with other characteristics such as strength, power and success (Connell, 1995). Physical attractiveness is not masculine; hence it might not be rewarded in men.

The second theory suggests that the gender differences in workplace powers can also help explain the steeper beauty gradient in earnings among women. Haveman and Beresford (2012) point out that men are more likely to hold high positions in the workplace (i.e., manager, supervisor, director) and thus have the ability to set a salary. In addition, several studies show that men often discriminate in favour of physically attractive women, but not vice versa, there is little effect of beauty on the ratings of men's performance (Kaplan, 1978; Gueguen and Jacob, 2011). Thus, attractive women might be offered higher wages by their male employers, leading to a larger attractiveness premium among women.

The magnitude and sign of estimated coefficients for other control variables are also comparable to those found in previous studies on wage determination, including Autor and Handel (2013) and Deming and Kahn (2018). Particularly, the positive coefficients (although statistically insignificant for some) on education categories imply a higher wage premium for a higher level of education. Similarly, the higher experience categories will lead to higher wage premia. In the same vein, higher job positions are associated with higher wage premia.

Due to differences in infrastructure and economic growth, it is likely that earnings differ across geographical regions. As expected, we observe significant wage premia ranging between 5.3 and 8.4 percent for the two largest cities in comparison with other cities. This finding is in line with a model of wage disparities across local labour markets with workers of different skill levels and imperfect labour mobility (Enrico, 2011). Companies in highly developed urban areas are more likely to face higher demand for skills and be more productive (Glaeser and Mare, 2001). Thus, to attract the most talented workers, firms have to post higher salaries in developed urban areas than in less developed areas.

Results for the wage returns to different skill groups are mixed. Specifically, the strongest wage premium can be observed for language skills, followed by project management skills, among both female- and male-targeted jobs. Some skills seem to be important in women's jobs but not in men's jobs, and vice versa. For instance, while customer skill is significantly rewarded (e.g., 2.7% to 10.3% of wage premia) among women's jobs, it does not exert any wage effect among men's jobs. In contrast, having character skills in a job ad signals lower-paid jobs for women, having such skill requirements does not change the wage level for men. Regarding the other skill groups (e.g., software, artistic and writing skills), we do not observe significant wage effects.

2.5.3. Beauty effects across jobs with high and low social and customer interactions

Table 2.8 presents regression results for beauty effects across jobs requiring different levels of interpersonal interactions. In columns (1) and (2), the positive and significant coefficients of the interaction term between *Attractiveness* and *Social interaction* indicator imply that the beauty effect for women among occupations with a high level of social interactions is indeed stronger than the beauty effect among those with less social interactions. Yet, the estimated coefficients for *Attractiveness* variable remain positive and significant across two different measures of interpersonal interactions. This result suggests that a good look can yield higher pay for women not only in jobs with intensive social interactions where beauty is explicitly essential for job performance but also in those where beauty is not expected to be a productivity-enhancing characteristic.

The estimated coefficients for *Attractiveness* in columns (3) and (4) remain insignificant across two different measures of social interactions. This finding reveals that the beauty effect does not exist among men-targeted jobs with a low level of social and customer interactions. Interestingly, the significant coefficient of the interaction term in column (3) indicates a negative effect on earnings for good-looking men in occupations requiring a high level of customer service. A possible interpretation is that even in the same occupation, men performing job tasks requiring beauty are less appreciated and rewarded than men performing job tasks requiring other skills. This further provides evidence of the negligible role of physical appearance for men in the labour market.

[Insert Table 2.8 here]

2.5.4. Heterogeneity of returns to beauty

As can be seen from column (1) of Table 2.9 for the women subsample, attractiveness can yield a wage premium of 10.1 percent in jobs requiring high school graduates. However, at higher levels of education, such as associate and bachelor's degrees, the beauty premium declines substantially by 3.6% and 15.2%, respectively. In general, our results show that beauty premium is largest in jobs at low-education job segments and decreases and even disappears in higher-education job segments.

[Insert Table 2.9 here]

Column (1) of Table 2.10 presents regression results for the men subsample. Interestingly, we only found a wage penalty of 7.3 percent for beautiful men among job ads requiring high school level, and no beauty effect among jobs requiring other educational levels.

[Insert Table 2.10 here]

With regards to experience level, estimation results from column (2) of Table 2.9 reveal that beautiful women enjoy an earning premium of 5.7% among vacancies requiring less than one year of experience. Such premium declines by 9.8% among those requesting 2-5 years of experience. As in column (2) of Table 2.10, attractive men receive a wage premium of 5.4% among job ads requesting 1-2 years of experience. Additionally, we find 8.3% and 33.4% wage penalties among jobs requiring less than one year of experience and those requiring more than five years of experience, respectively.

The last columns of Table 2.9 and Table 2.10 show estimates for the heterogeneous attractiveness effect on earning across job levels. Particularly, we only find a significant salary premium of 10.5% for attractive women in job ads at entry-level. For medium-managerial positions, beauty can bring a small wage penalty of 2.2% for attractive women. Likewise, among men's jobs, those at entry-level offer a wage premium of 12% for good-looking male candidates. The beauty premium for women vanishes then turns into wage penalties of 4.7% and 7.7% among higher-level jobs (i.e., non-managerial and medium-managerial positions, respectively).

Overall, our estimation results reveal that the beauty effect among women mainly exists in occupations requiring a low education level, low experience level and non-managerial position. This effect gradually lessens and then disappears in higher-tier jobs. This finding is in line with previous studies documenting the "beauty is beastly" effect (Heilman et al., 1979; Braun et al., 2015), which states that attractiveness benefits women in non-managerial positions but disadvantages women in managerial positions. Additionally, it might be the case that for high-skilled jobs that require a high level of education and experience, the salary must be carefully set and free from any discrimination based on physical attractiveness. The beauty effects among men's jobs are somewhat mixed, but in general, there is no beauty effect or beauty penalty among most job types.

The physical attractiveness effect might also be influenced by the differences in the level of competition among local labour markets. Specifically, Table 2.11 reports results for differential beauty effects across job locations (i.e., Hanoi, Ho Chi Minh city and others). Among jobs located in the two big cities, women enjoy a premium of 5.1% to 6.4%, whereas among jobs located in other regions, no significant beauty premium is found. Again, there is no evidence of beauty premium for men in the three location groups. This finding is consistent with Deng et al. (2020), who state that the beauty premium for women in big cities is larger than that for women in small cities. They explain such results by the competitiveness of the job market. In particular, attractive women are more likely to be favoured in highly competitive markets, hence a higher wage premium exists in big cities.

[Insert Table 2.11 here]

2.6. Conclusion

In this paper, we examine the effect of physical attractiveness on earning. We employ a rich dataset of online job ads gathered from one of the largest job portals in Vietnam. To mitigate the selection bias in estimating the beauty effect, we use the propensity score matching method to match vacancies on all observable characteristics, including job title, location, education level, experience level, job position, skills required and firm size. In addition, we rely on coarsened exact matching to match vacancies based on job title, location, education level, experience level, job position and firm size. Furthermore, we use the Jaccard score – a text-matching algorithm to match job postings based on the job ad text.

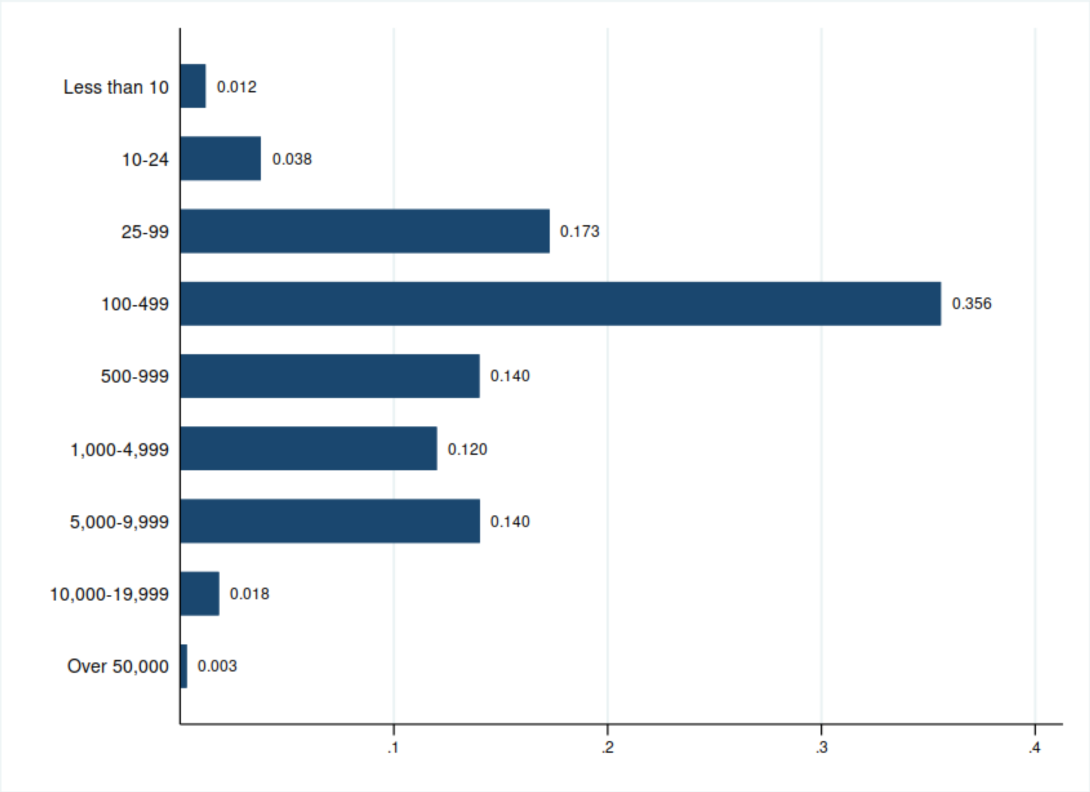
According to our estimation results, beauty matters differently for men and women. Particularly, physically attractive women are offered on average 3.3% to 5.2% higher salaries than plain people. Yet, negative beauty effects of zero to 4.8% were found for good-looking men. These results support the body of literature documenting that physical appearance is more important among women than men. Such differences might stem from the persistent traditional gender roles or different treatments of attractive employees between male and female managers. Besides, we rule out the productivity effect of good look by showing evidence of wage premia for physically attractive women, not only among jobs involving a high level of social/customer interactions where beauty is essential for job performance but also among other occupations where beauty is not a requisite.

Additionally, the wage premium varies across different education, experience levels and job positions. For jobs which require a low level of education and entry-level jobs, beauty can bring an earning premium of around 4% for beautiful women. However, the higher the job level, the smaller the beauty effect is. Furthermore, we document that return to beauty is heterogeneous across work locations. We find evidence of beauty premium for women in big cities and beauty penalty for women in small towns. There is no significant evidence of beauty premium for male workers in big and small towns.

Our findings suggest that beauty premium exists even after controlling for a rich set of job's observable characteristics. This result can be beneficial for policymakers to reduce discrimination and promote equal pay in the labour market. However, estimating the level of discrimination might encounter disadvantages using traditional labour market data sources, which might be biased due to the lack of many labour market dimensions and might not be promptly and regularly updated. In our paper, we can address this potential issue by using online vacancies as a rich and real-time data source for instantaneous labour market dynamics analyses.

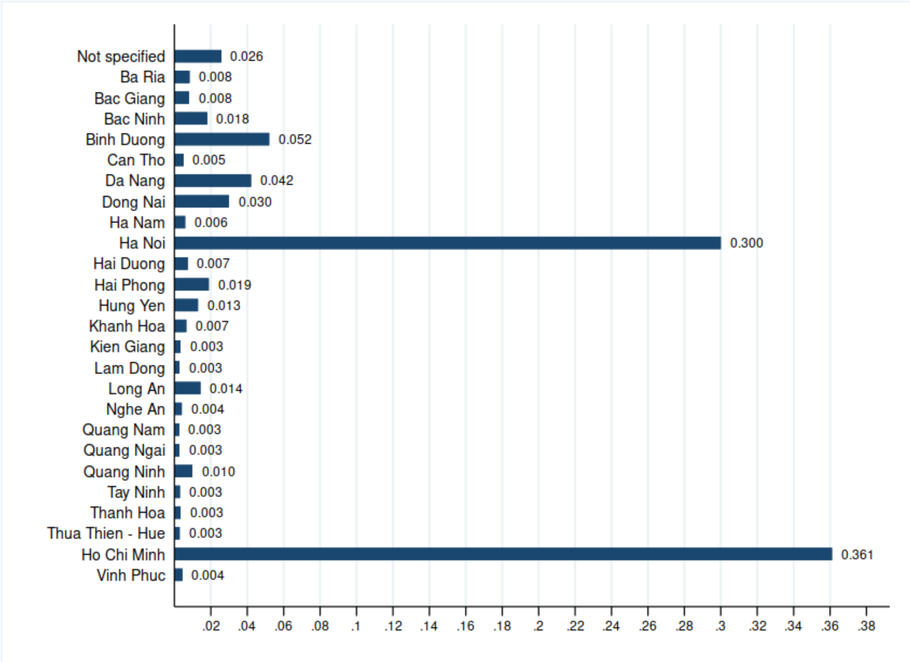
Figures

Figure 2.1. Share of job ads by firm sizes

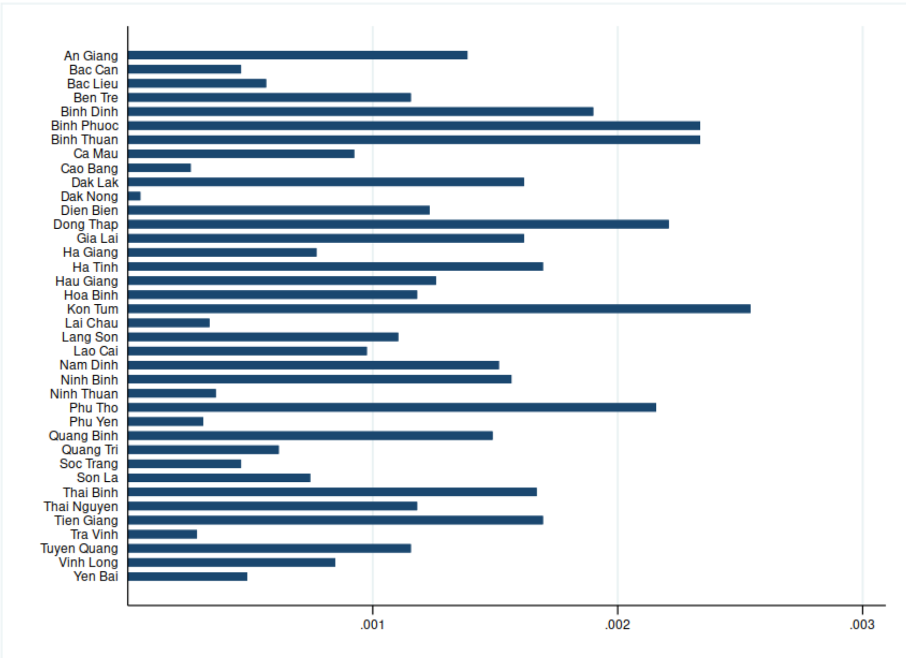


Notes: This figure shows the share of vacancies by firm size.

Figure 2.2. Share of job ads by locations



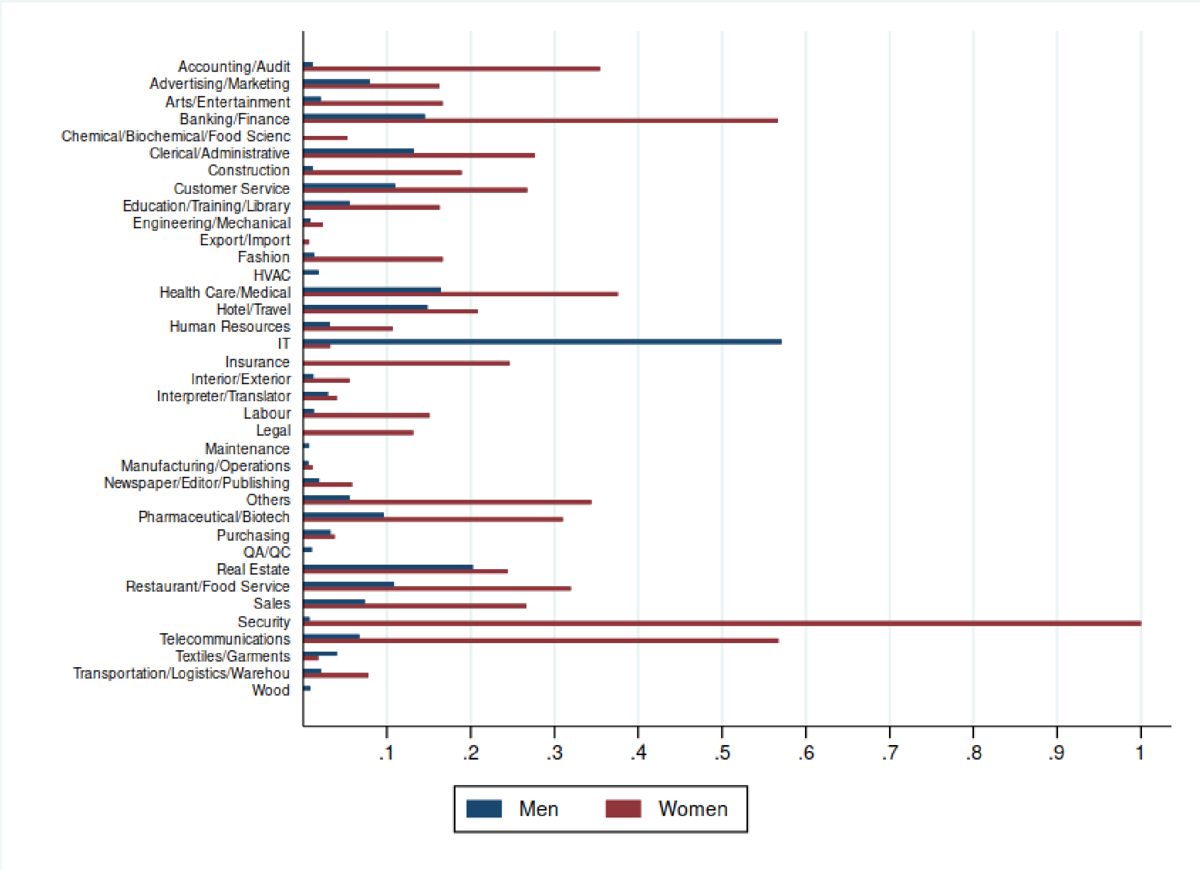
(a)



(b)

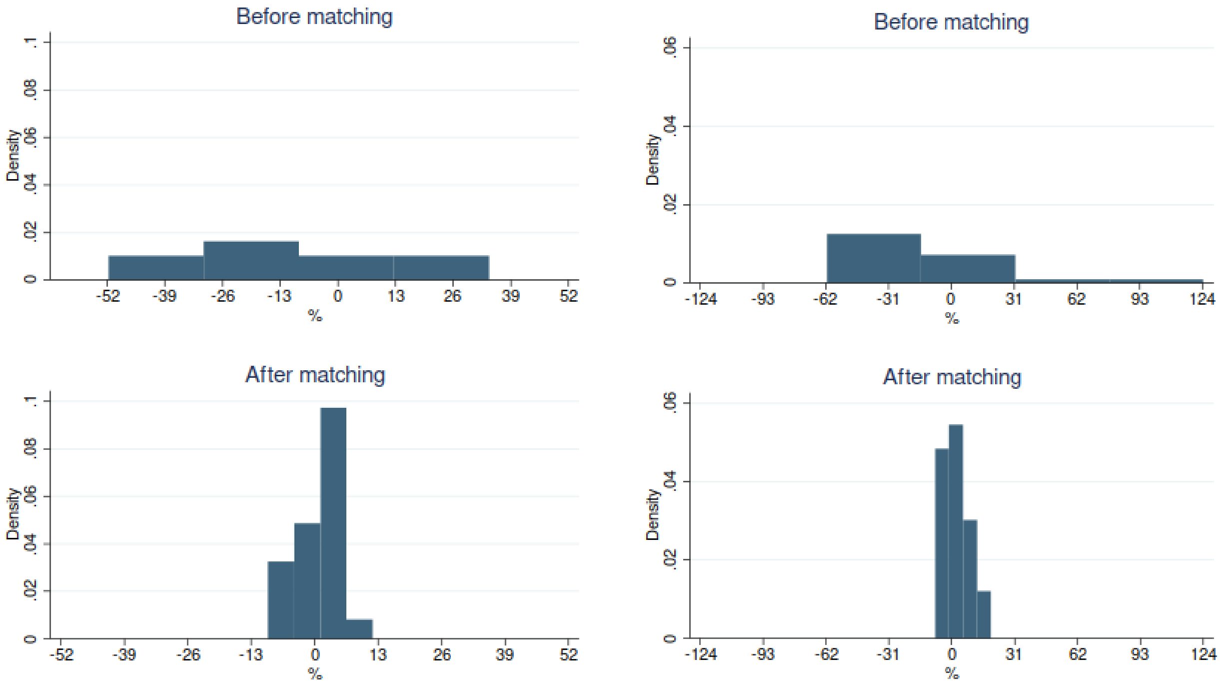
Notes: Figure 2 (a) shows the share of vacancies by job locations having less than 100 job postings. Figure 2 (b) shows the share of vacancies by job locations having at least 100 job postings.

Figure 2.3. Frequency of stating beauty requirements by industries for men's versus women's jobs



Notes: This figure shows the frequency of stating beauty requirements by job industries across male- and female-targeted ads.

Figure 2.4. Standardised bias distribution across covariates before versus after matching



(a) (b)

Notes: This figure shows the standardised bias distribution across covariates before and after matching for the subsample of (a) female-targeted ads and (b) male-targeted ads. The horizontal line presents the standardised bias value (in %), and the vertical line presents the density.

Tables

Table 2.1. List of keywords

Skill	Keywords
Art	Art, artistic
Character	Organised, detail-oriented, multitasking, time management, meeting deadlines, energetic
Cognitive	Problem-solving, research, analytical, critical thinking, math, statistics
Computer	Computer, spreadsheets, common software (e.g., Microsoft Excel, PowerPoint)
Customer service	Customer, sales, client, customer service
Financial	Budgeting, accounting, finance, cost
Language	English, Japanese, Korean, Chinese, foreign language
Project management	Project management
People management	Supervisory, leadership, management (not project), mentoring, staff
Social	Communication, teamwork, collaboration, negotiation, presentation
Software	Programming language or specialized software (e.g., Java, SQL, C++)
Writing	Writing
Attractiveness	Pretty face, attractive face, good looking, pretty, nice face

Notes: The table presents the categorisation of keywords into skill/requirement groups.

Table 2.2. Correlation matrix

	Beauty	Education	Experience	Level	Computer	Software	Language	Financial	People	Project	Art	Character	Cognitive	Customer	Social
Education	-0.2204														
Experience	-0.1357	0.2967													
Level	-0.0777	0.1266	0.4375												
Computer	0.0116	0.0963	0.1484	0.0338											
Software	-0.1031	0.1904	0.2497	0.0221	0.2162										
Language	-0.0867	0.1757	0.2249	0.0786	0.1575	0.1061									
Financial	0.0628	0.0450	0.1255	-0.0276	0.1923	0.1185	-0.0622								
People Mgmt.	-0.0294	0.0598	0.1827	0.2176	0.0484	0.0542	0.0382	-0.0026							
Project Mgmt.	-0.0682	0.0911	0.2547	0.2892	0.1064	0.0926	0.0444	0.0187	0.1871						
Art	-0.0512	0.0454	0.1231	0.0445	0.0164	0.1999	0.0134	-0.0469	0.0337	0.0743					
Character	0.1111	-0.0647	-0.1889	-0.0897	0.0224	-0.0817	-0.1716	0.0088	0.0049	0.0013	0.0013				
Cognitive	-0.0909	0.1456	0.2738	0.1432	0.1725	0.2265	0.1273	0.0335	0.1370	0.2548	0.1915	0.0926			
Customer	-0.0815	0.1611	-0.0704	0.0474	-0.0165	-0.0791	-0.0584	-0.0983	0.0015	0.0479	0.0118	0.1690	0.0686		
Social	0.0459	0.1330	-0.0669	0.0044	-0.0146	-0.0746	0.1116	-0.0824	0.0501	0.0590	-0.0093	0.1641	0.0454	0.3163	
Writing	-0.0285	0.0534	0.0654	-0.0080	0.0683	0.0613	0.0765	-0.0214	0.0395	0.0366	0.0863	0.0100	0.0948	0.0670	0.0190

Notes: The table shows the Spearman correlation coefficients for skill requirement variables.

Table 2.3. Descriptive statistics

	Men (N=23,908)		Women N=(15,037)		
	(1) Mean	**(2) SD**	**(3) Mean**	**(4) SD**	**(5) Diff.**
Wage	8,882,292	5,279,748	7,752,873	4,038,827	1,129,419***
Education					
High school	0.1886	0.3912	0.0902	0.2865	0.0984***
Vocational training	0.0131	0.1137	0.0039	0.0625	0.0092***
Associate Degree	0.5929	0.4913	0.7059	0.4556	-0.1131***
Undergraduate	0.1993	0.3995	0.1979	0.3984	0.0014
Master's degree	0.0006	0.0242	0.0000	0.0000	0.0006***
Other educations	0.0055	0.0741	0.0020	0.0446	0.0035***
Experience					
0-1 year	0.4050	0.4909	0.3957	0.4890	0.0093*
1-2 years	0.3327	0.4712	0.4040	0.4907	-0.0713***
2-5 years	0.2242	0.4171	0.1854	0.3886	0.0388***
5-10 years	0.0347	0.1831	0.0146	0.1201	0.0201***
10+ years	0.0034	0.0581	0.0003	0.0163	0.0031***
Level					
New entry	0.0329	0.1783	0.0423	0.2013	-0.0094***
Employee	0.8472	0.3598	0.8873	0.3162	-0.0402***
Medium-level manager	0.1094	0.3122	0.0628	0.2427	0.0466***
Top manager	0.0105	0.1021	0.0075	0.0864	0.0030***
Location					
Ha Noi	0.2746	0.4463	0.3461	0.4758	-0.0716***
Ho Chi Minh city	0.3572	0.4792	0.3674	0.4821	-0.0102**
Other cities	0.3683	0.4823	0.2865	0.4521	0.0818***
Skill					
Computer	0.2771	0.4476	0.5246	0.4994	-0.2476***
Software	0.1791	0.3835	0.1617	0.3682	0.0174***
Financial	0.0675	0.2508	0.3346	0.4719	-0.2672***
People Management	0.0289	0.1675	0.0348	0.1832	-0.0059***
Project Management	0.1031	0.3042	0.0763	0.2655	0.0269***
Artistic	0.0898	0.2860	0.0416	0.1998	0.0482***
Language	0.1746	0.3796	0.3058	0.4608	-0.1313***
Character	0.7972	0.4021	0.8615	0.3455	-0.0643***
Cognitive	0.2545	0.4356	0.2902	0.4539	-0.0357***
Customer Service	0.1833	0.3869	0.2633	0.4404	-0.0800***
Social	0.4057	0.4910	0.4656	0.4988	-0.0599***
Writing	0.0257	0.1583	0.0283	0.1657	-0.0025
Attractiveness	0.0951	0.2933	0.2661	0.4420	-0.1711***

Notes: The table shows summary statistics of all variables used in the regression. Columns (1) and (3) report the mean of men and women subsamples, respectively. Columns (2) and (4) report the standard deviation of men and women subsamples, respectively. Column (5) reports the difference in means between men and women subsamples. *, **, and *** indicate significance at 10%, 5%, and 1% for a t-test of whether men and women subsamples have equal means, respectively.

Table 2.4. Frequency of stating beauty preference among subsamples

	Men	Women	Diff.
	(1)	(2)	(3)
Education			
High school	0.1346	0.5335	-0.3989***
Vocational training	0.0479	0.1864	-0.1385***
Associate degree	0.1662	0.3872	-0.1667***
Bachelor’s degree	0.0288	0.1230	-0.0942***
Master degree	0.0000	_	_
Other educations	0.0530	0.1000	-0.0470
Experience			
0-1 year	0.1870	0.2882	-0.1012***
1-2 years	0.0328	0.3254	-0.2926***
2-5 years	0.0351	0.1044	-0.0693***
5-10 years	0.0157	0.0864	-0.0707***
10+ years	0.0000	0.0000	_
Level			
New entry/Internship	0.1959	0.2689	-0.0729***
Non-manager	0.1007	0.2782	-0.1775***
Medium manager	0.0287	0.1259	-0.0973***
Top manager	0.0198	0.0000	0.0198
Location			
Hanoi	0.1057	0.2375	-0.1317***
Ho Chi Minh city	0.0858	0.2759	-0.1900***
Other cities	0.0961	0.2883	-0.1922***

Notes: The table shows the frequency of stating beauty requirements and mean wage by different subsamples of education, job levels, experience levels and job locations for men-targeted ads in column (1) and for women-targeted ads in column (2). Column (3) reports the difference in means between men and women subsamples. *, **, and *** indicate significance at 10%, 5%, and 1% for a t-test of whether men and women subsamples have equal means, respectively.

Table 2.5. CEM balance checks

Panel A: Female subsample							
Multivariate L1 distance before matching: 0.8243							
Multivariate L1 distance after matching: 0.1596							
Univariate imbalance							
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
	L1	**Mean**	**Min**	**25%**	**50%**	**75%**	**Max**
Education	0	0	0	0	0	0	0
Experience	0	0	0	0	0	0	0
Level	0	0	0	0	0	0	0
Location	0	0	0	0	0	0	0
Quarter	0	0	0	0	0	0	0
Job title	0.1170	-0.3697	0	0	0	0	-2
Firm size	0	0	0	0	0	0	0
Panel B: Male subsample							
Multivariate L1 distance before matching: 0.9300							
Multivariate L1 distance after matching: 0.0451							
Univariate imbalance							
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
	L1	**Mean**	**Min**	**25%**	**50%**	**75%**	**Max**
Education	0	0	0	0	0	0	0
Experience	0	0	0	0	0	0	0
Level	0	0	0	0	0	0	0
Location	0	0	0	0	0	0	0
Quarter	0	0	0	0	0	0	0
Job title	0.0139	0.1733	1	0	0	0	0
Firm size	0	0	0	0	0	0	0

Notes: The table reports multivariate and univariate imbalance measures for the female subsample in panel A and makes subsample in Panel B. For both panels, column (1) reports the L1 imbalance measure for each variable. Column (2) reports the difference in the density distributions between treated and control groups at the mean. Columns (3) to (7) report the difference in the density distributions between treated and control groups for the 0th (Min), 25th, 50th, 75th and 100th (Max) percentiles, respectively.

Table 2.6. Regression results of Probit models for beauty-requesting ads

	Female		Male	
	(1)	**(2)**	**(3)**	**(4)**
	Coefficient	**Marginal effect**	**Coefficient**	**Marginal effect**
Vocational training	-0.307	-0.060	-1.298***	-0.080***
	(0.198)	(0.037)	(0.230)	(0.013)
Associate degree	-0.094*	-0.019*	-1.257***	-0.078***
	(0.055)	(0.011)	(0.130)	(0.009)
Bachelor's degree	-0.169**	-0.034**	-1.830***	-0.104***
	(0.075)	(0.015)	(0.237)	(0.014)
Other educations	-0.291***	-0.057***	-1.905***	-0.108***
	(0.063)	(0.013)	(0.174)	(0.012)
1-2 years	0.148***	0.028***	-0.549***	-0.025***
	(0.035)	(0.007)	(0.109)	(0.005)
2-5 years	0.011	0.002	-0.102	-0.005
	(0.055)	(0.010)	(0.149)	(0.007)
5-10 years	-0.353*	-0.061**	-1.196***	-0.050***
	(0.185)	(0.030)	(0.459)	(0.017)
Non-manager	-0.048	-0.009	-0.010	-0.000
	(0.073)	(0.014)	(0.194)	(0.009)
Medium manager	-0.774***	-0.131***	0.481	0.023
	(0.165)	(0.026)	(0.317)	(0.016)
Top manager			-3.165***	-0.105***
			(1.005)	(0.027)
Hanoi	-0.159***	-0.030***	0.907***	0.042***
	(0.038)	(0.007)	(0.105)	(0.005)
HCM	-0.033	-0.006	0.119	0.005
	(0.036)	(0.007)	(0.100)	(0.004)
Computer	0.062*	0.012*	-0.282**	-0.013**
	(0.035)	(0.007)	(0.117)	(0.005)
Software	-0.507***	-0.095***	-0.119	-0.005
	(0.055)	(0.010)	(0.174)	(0.008)
Language	-0.207***	-0.039***	-0.250	-0.011
	(0.044)	(0.008)	(0.184)	(0.008)
Financial	-0.065	-0.012	0.919**	0.041**
	(0.050)	(0.009)	(0.375)	(0.017)
People management	0.198**	0.037**	0.071	0.003
	(0.097)	(0.018)	(0.318)	(0.014)
Project management	0.173***	0.033***	-1.110***	-0.049***
	(0.065)	(0.012)	(0.218)	(0.010)
Art	0.079	0.015	-0.385*	-0.017*
	(0.086)	(0.016)	(0.217)	(0.010)
Character	0.341***	0.064***	1.287***	0.057***
	(0.054)	(0.010)	(0.214)	(0.009)
Cognitive	-0.204***	-0.038***	0.434***	0.019***
	(0.040)	(0.007)	(0.123)	(0.005)
Customer service	0.072	0.013	-0.380***	-0.017***
	(0.044)	(0.008)	(0.130)	(0.006)
Social	0.327***	0.061***	1.965***	0.087***
	(0.037)	(0.007)	(0.119)	(0.005)
Writing	0.206**	0.039**	-1.622***	-0.072***
	(0.098)	(0.018)	(0.405)	(0.018)
Obs	14,464		15,465	
Pseudo R2	0.428		0.615	

Notes: The table presents the results of Probit models for job characteristics/requirements of beauty-requesting ads. Columns (1) and (3) show the estimated coefficients for female and male subsamples, respectively. Columns (2) and (4) show the average marginal effect for female and male subsamples, respectively. The dependent variable is a dummy variable that indicates whether a vacancy mention beauty preference. Time, industry, job title, and firm size dummies are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table 2.7. Beauty effects on earnings

	Female			Male		
	(1) PSM	(2) CEM	(3) Text matching	(4) PSM	(5) CEM	(6) Text matching
Attractiveness	0.052***	0.035***	0.033**	-0.041***	-0.048**	-0.013
	(0.008)	(0.011)	(0.015)	(0.014)	(0.019)	(0.029)
Vocational training	-0.032		0.179**	-0.046		-0.105***
	(0.061)		(0.074)	(0.066)		(0.012)
Associate degree	-0.022*	-0.018	-0.020**	0.070***	0.038*	-0.020***
	(0.013)	(0.019)	(0.009)	(0.019)	(0.022)	(0.006)
Bachelor's degree	0.049***	0.085***	0.022	0.145***	0.153***	-0.094
	(0.019)	(0.030)	(0.024)	(0.035)	(0.050)	(0.059)
Other educations	0.082		0.149*	-0.000		-0.055
	(0.083)		(0.081)	(0.149)		(0.104)
1-2 years	0.064***	0.103***	0.034***	0.051***	0.068**	0.115***
	(0.009)	(0.011)	(0.008)	(0.019)	(0.027)	(0.033)
2-5 years	0.297***	0.305***	0.219***	0.176***	0.114***	0.179***
	(0.018)	(0.029)	(0.024)	(0.025)	(0.038)	(0.053)
5-10 years	0.497***		0.561**	0.282***	0.105	0.237**
	(0.089)		(0.247)	(0.068)	(0.073)	(0.104)
10+ years			1.502***			
			(0.248)			
Non-manager	0.037**	0.001	0.013	0.026	-0.036	0.053***
	(0.017)	(0.053)	(0.034)	(0.036)	(0.027)	(0.009)
Medium manager	0.217***	0.277*	0.440***	0.332***	0.155**	0.043
	(0.042)	(0.148)	(0.051)	(0.065)	(0.071)	(0.051)
Top manager				0.671***		
				(0.130)		
Hanoi	-0.003	0.030**	0.005	0.084***	0.031	0.006
	(0.011)	(0.015)	(0.009)	(0.019)	(0.021)	(0.010)
HCM	0.072***	0.093***	0.013	0.053***	-0.039**	0.002
	(0.010)	(0.014)	(0.008)	(0.018)	(0.016)	(0.008)
Computer	-0.025***	0.004	0.025*	-0.050***	0.006	-0.031
	(0.009)	(0.012)	(0.015)	(0.019)	(0.022)	(0.033)
Software	-0.001	0.007	-0.002	0.041	0.000	0.034
	(0.016)	(0.019)	(0.020)	(0.026)	(0.026)	(0.049)
Language	0.066***	0.084***	0.100***	0.214***	0.152***	0.236***
	(0.011)	(0.014)	(0.019)	(0.029)	(0.027)	(0.038)
Financial	-0.001	-0.061***	-0.054***	0.111***	0.072	0.026
	(0.012)	(0.013)	(0.013)	(0.041)	(0.049)	(0.063)
People management	-0.023	-0.057*	-0.027	-0.103*	-0.003	0.081
	(0.024)	(0.034)	(0.032)	(0.056)	(0.054)	(0.091)
Project management	0.082***	0.172***	-0.005	0.097***	0.063*	0.123
	(0.019)	(0.028)	(0.026)	(0.033)	(0.033)	(0.083)
Art	-0.026	-0.035	0.001	0.017	-0.004	-0.027
	(0.020)	(0.028)	(0.051)	(0.028)	(0.034)	(0.047)
Character	-0.086***	-0.072***	-0.113***	0.019	-0.051**	0.035
	(0.013)	(0.016)	(0.020)	(0.033)	(0.025)	(0.069)
Cognitive	0.022**	0.027*	0.003	-0.026	-0.003	-0.103***
	(0.011)	(0.014)	(0.017)	(0.021)	(0.024)	(0.035)
Customer service	0.027**	0.046***	0.103***	-0.000	-0.044*	-0.009
	(0.011)	(0.014)	(0.020)	(0.022)	(0.027)	(0.052)
Social	0.011	0.000	-0.010	0.028*	0.031*	0.036
	(0.009)	(0.012)	(0.016)	(0.015)	(0.019)	(0.026)
Writing	0.038	-0.038	0.012	-0.029	-0.041	-0.109
	(0.024)	(0.032)	(0.036)	(0.055)	(0.067)	(0.079)
Obs	4,034	2,770	2,753	1,442	1,154	1,909
R2	0.526	0.461	0.638	0.640	0.648	0.760

Notes: The table presents the beauty effect for women and men. Columns (1) and (4) show results from PSM for females and males, respectively. Columns (2) and (5) show results from CEM for females and males, respectively. Columns (3) and (6) show results from text matching for females and males, respectively. Time, industry, job title, and firm size fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table 2.8. Regression results for beauty effects across jobs with high versus low levels of social/customer interactions

	Female		Male	
	(1) **Customer service**	**(2)** **Social**	**(3)** **Customer service**	**(4)** **Social**
Attractiveness	0.049***	0.027**	-0.020	-0.016
	(0.009)	(0.011)	(0.019)	(0.026)
Social_interact	0.023*	-0.009	0.035	0.049*
	(0.013)	(0.011)	(0.032)	(0.026)
Attractiveness×Social_interact	0.008	0.041***	-0.066*	-0.037
	(0.017)	(0.016)	(0.038)	(0.033)
Obs	4,034	4,034	1,176	1,176
R2	0.526	0.527	0.613	0.612

Notes: The table presents the beauty effect for women and men across jobs with high and low level of social/customer interactions. Columns (1) and (2) show results from PSM for female-targeted ads. Columns (3) and (4) show results from PSM for male-targeted ads. Social_interact indicator in columns (1) and (3) is measured by Customer service skill. Social_interact indicator in columns (2) and (4) is measured by Social skill. Education levels, experience levels, job positions, locations, skill requirements, time, industry, job title, and firm size fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table 2.9. Regression results for heterogeneous beauty effects for females

	(1) Education	(2) Experience	(3) Position
Attractiveness	0.101***	0.057***	0.105***
	(0.019)	(0.011)	(0.033)
Vocational training	0.037	-0.044	-0.032
	(0.073)	(0.062)	(0.061)
Associate degree	-0.006	-0.023*	-0.020
	(0.016)	(0.013)	(0.013)
Bachelor’s degree	0.124***	0.041**	0.049**
	(0.024)	(0.019)	(0.019)
Other educations	0.155	0.072	0.088
	(0.123)	(0.083)	(0.088)
Attractiveness×Vocational training	-0.111		
	(0.111)		
Attractiveness×Associate degree	-0.036*		
	(0.022)		
Attractiveness×Bachelor’s degree	-0.152***		
	(0.030)		
Attractiveness×Other educations	-0.140		
	(0.163)		
1-2 years	0.064***	0.054***	0.063***
	(0.009)	(0.011)	(0.009)
2-5 years	0.290***	0.346***	0.297***
	(0.018)	(0.022)	(0.018)
5-10 years	0.483***	0.574***	0.505***
	(0.089)	(0.127)	(0.088)
Attractiveness×1-2 years		0.019	
		(0.016)	
Attractiveness×2-5 years		-0.098***	
		(0.029)	
Attractiveness×5-10 years		-0.132	
		(0.178)	
Non-manager	0.038**	0.038**	0.062***
	(0.017)	(0.017)	(0.021)
Medium manager	0.215***	0.208***	0.276***
	(0.042)	(0.042)	(0.048)
Attractiveness×Non-manager			-0.053
			(0.035)
Attractiveness×Medium manager			-0.127**
			(0.053)
Obs	4,034	4,034	4,034
R2	0.530	0.529	0.527

Notes: The table presents the heterogeneous beauty premium for female-targeted ads. Columns (1), (2), and (3) show results for different beauty effects across education level, experience level, and job position, respectively. The base categories are High school, 0-1 year of experience, and Entry level/internship, respectively. Other control variables and fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table 2.10. Regression results for heterogeneous beauty effects for male

	(1) Education	(2) Experience	(3) Position
Attractiveness	-0.073***	-0.083***	0.120*
	(0.026)	(0.019)	(0.062)
Vocational training	-0.146	-0.006	-0.044
	(0.126)	(0.064)	(0.066)
Associate degree	0.042	0.078***	0.075***
	(0.027)	(0.019)	(0.020)
Bachelor's degree	0.147***	0.150***	0.158***
	(0.044)	(0.036)	(0.036)
Other educations	-0.043	-0.016	0.016
	(0.268)	(0.148)	(0.141)
Attractiveness×Vocational training	0.200		
	(0.135)		
Attractiveness×Associate degree	0.053		
	(0.034)		
Attractiveness×Bachelor's degree	-0.011		
	(0.048)		
Attractiveness×Other educations	0.086		
	(0.286)		
1-2 years	0.055***	-0.018	0.050***
	(0.020)	(0.027)	(0.019)
2-5 years	0.181***	0.169***	0.176***
	(0.026)	(0.031)	(0.025)
5-10 years	0.280***	0.378***	0.275***
	(0.068)	(0.082)	(0.068)
Attractiveness×1-2 years		0.137***	
		(0.034)	
Attractiveness×2-5 years		0.023	
		(0.036)	
Attractiveness×5-10 years		-0.251**	
		(0.104)	
Non-manager	0.023	0.027	0.091*
	(0.036)	(0.037)	(0.047)
Medium manager	0.322***	0.327***	0.400***
	(0.065)	(0.067)	(0.073)
Top manager	0.656***	0.703***	0.612***
	(0.130)	(0.138)	(0.176)
Attractiveness×Non-manager			-0.167***
			(0.064)
Attractiveness×Medium manager			-0.197**
			(0.086)
Attractiveness×Top manager			0.078
			(0.179)
Obs	1,442	1,442	1,442
R2	0.641	0.647	0.642

Notes: The table presents the heterogeneous beauty premium for male-targeted ads. Columns (1), (2), and (3) show results for different beauty effects across education level, experience level, and job position, respectively. The base categories are High school, 0-1 year of experience and Entry level/internship, respectively. Other control variables and fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table 2.11. Regression results for gender differentials between small versus big cities

	(1) Female	(2) Male
Attractiveness	0.001	-0.032
	(0.019)	(0.027)
Hanoi	-0.030*	0.038
	(0.017)	(0.026)
HCM	0.041**	0.015
	(0.017)	(0.026)
Attractiveness×Hanoi	0.063***	0.031
	(0.023)	(0.034)
Attractiveness×HCM	0.050**	0.021
	(0.023)	(0.035)
Vocational training	0.066	-0.030
	(0.062)	(0.044)
Associate degree	-0.026**	0.043**
	(0.013)	(0.021)
Bachelor's degree	0.089***	0.163***
	(0.020)	(0.038)
Other educations	0.013	0.093***
	(0.015)	(0.027)
1-2 years	0.083***	0.044***
	(0.010)	(0.016)
2-5 years	0.323***	0.169***
	(0.019)	(0.024)
5-10 years	0.462***	0.303***
	(0.097)	(0.069)
Non-manager	0.017	0.008
	(0.019)	(0.028)
Medium manager	0.191***	0.198***
	(0.040)	(0.062)
Top manager		0.565***
		(0.133)
Obs	4,003	1,461
R2	0.498	0.672

Notes: The table presents the heterogeneous beauty premium across job locations. The base category is other job locations. Columns (1) and (2) show results for the subsamples of female- and male-targeted ads, respectively. Other control variables and fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Online Appendix

Table A2.1. List of job titles/occupations

Accountant	Driver	PR staff
Accounting manager	Editor	Primary school teacher
Actor	Engineer	Procedure controller
Admin	English interpreter	Product advisor
Appraiser	English teacher	Product/content reviewer
Architect	Equipment deployment staff	Production statistics staff
Art teacher	Event organiser	Project development staff
Bartender	Financial advisor	QA staff
Beauty teacher	Gymnastics teacher	Real estate businessman
Beauty technician	HR staff	Receptionist
Bodyguard	Import-export staff	Restaurant manager
Branch manager	Internal auditor	Sale assistant
Broker	IT staff	Sales
Business representative	IT teacher	Sales consultant
Business support staff	Legal experts	Sales manager
Businessman	Librarian	Secretary
Cashier	Livestreamer	Security
Chef	Manager	Shop manager
Chess game instructor	Marketing	Spa advisor
Chinese interpreter	Marketing manager	Stylist
Clearance specialist	Maths teacher	Supply chain staff
Content creator	MC	Support staff
Controller/supervisor	Merchandiser	Swimming pool lifeguard
Cook	Network admin	Teaching assistant
Customer service manager	Nurse	Technician
Customer service staff	Nursery teacher	Tour operator
Deliverman	Office supporter	Training specialist
Deputy	Officer	Visa/study abroad advisor
Deputy manager	Online supporter	Waiter
Designer	Others	Warehouse staff

Developer	Partnership staff	Workman
Director	Personal trainer	
Doctor	Pharmacist	

Notes: The table presents the job titles/occupations in our data.

Table A2.2. Definition of main variables

Variable	Definition
High school	A dummy variable that takes the value of 1 if the job ad requires a high school qualification, and 0 otherwise.
Vocational training	A dummy variable that takes the value of 1 if the job ad requires a vocational training qualification, and 0 otherwise.
Associate degree	A dummy variable that takes the value of 1 if the job ad requires an associate degree qualification, and 0 otherwise.
Bachelor's degree	A dummy variable that takes the value of 1 if the job ad requires a bachelor's degree qualification, and 0 otherwise.
Other educations	A dummy variable that takes the value of 1 if the job ad requires other educational qualifications, and 0 otherwise.
0-1 year	A dummy variable that takes the value of 1 if the job ad requires less than one year of experience, and 0 otherwise.
1-2 years	A dummy variable that takes the value of 1 if the job ad requires 1-2 years of experience, and 0 otherwise.
2-5 years	A dummy variable that takes the value of 1 if the job ad requires 2-5 years of experience, and 0 otherwise.
5-10 years	A dummy variable that takes the value of 1 if the job ad requires 5-10 years of experience, and 0 otherwise.
10+ years	A dummy variable that takes the value of 1 if the job ad requires more than 10 years of experience, and 0 otherwise.
Entry level/Internship	A dummy variable that takes the value of 1 if the job ad is at an entry level/internship position, and 0 otherwise.
Non-manager	A dummy variable that takes the value of 1 if the job ad is at a non-managerial position, and 0 otherwise.
Medium manager	A dummy variable that takes the value of 1 if the job ad is at a medium-manager position, and 0 otherwise.
Top manager	A dummy variable that takes the value of 1 if the job ad is at a top-manager position, and 0 otherwise.
Hanoi	A dummy variable that takes the value of 1 if the job ad locates in Hanoi, and 0 otherwise.
HCM	A dummy variable that takes the value of 1 if the job ad locates in HCM city, and 0 otherwise.
Other locations	A dummy variable that takes the value of 1 if the job ad locates in other cities/provinces, and 0 otherwise.
Computer	A dummy variable that takes the value of 1 if the job ad requires computer skill, and 0 otherwise.
Software	A dummy variable that takes the value of 1 if the job ad requires software skill, and 0 otherwise.
Language	A dummy variable that takes the value of 1 if the job ad requires language skill, and 0 otherwise.
Financial	A dummy variable that takes the value of 1 if the job ad requires financial skill, and 0 otherwise.
People management	A dummy variable that takes the value of 1 if the job ad requires people management skill, and 0 otherwise.
Project management	A dummy variable that takes the value of 1 if the job ad requires project management skill, and 0 otherwise.
Art	A dummy variable that takes the value of 1 if the job ad requires artistic skill, and 0 otherwise.
Character	A dummy variable that takes the value of 1 if the job ad requires character skill, and 0 otherwise.

Cognitive	A dummy variable that takes the value of 1 if the job ad requires cognitive skill, and 0 otherwise.
Customer service	A dummy variable that takes the value of 1 if the job ad requires customer service skill, and 0 otherwise.
Social	A dummy variable that takes the value of 1 if the job ad requires social skill, and 0 otherwise.
Writing	A dummy variable that takes the value of 1 if the job ad requires writing skill, and 0 otherwise.
Attractiveness	A dummy variable that takes the value of 1 if the job ad requires physical attractiveness, and 0 otherwise.

Notes: The table presents definitions of the main variables used in our analysis.

Table A2.3. Robustness test for beauty effect with average salary

	(1) PSM	(2) CEM	(3) Text matching	(4) PSM	(5) CEM	(4) Text matching
Attractiveness	0.062***	0.058***	0.036**	0.002	-0.004	0.040
	(0.008)	(0.012)	(0.016)	(0.015)	(0.024)	(0.038)
Vocational training	-0.017		0.258***	-0.001		-0.165***
	(0.058)		(0.069)	(0.047)		(0.018)
Associate degree	-0.030**	-0.039	-0.009	0.039*	0.007	-0.020**
	(0.014)	(0.024)	(0.011)	(0.020)	(0.031)	(0.009)
Bachelor's degree	0.042**	0.070**	0.057**	0.116***	0.149***	-0.045
	(0.021)	(0.033)	(0.027)	(0.036)	(0.053)	(0.060)
Other educations	0.026		-0.082	0.029		0.027
	(0.082)		(0.103)	(0.085)		(0.071)
1-2 years	0.063***	0.080***	0.044***	0.037	0.041	0.050
	(0.010)	(0.012)	(0.009)	(0.022)	(0.034)	(0.041)
2-5 years	0.282***	0.254***	0.218***	0.167***	0.114***	0.158***
	(0.019)	(0.029)	(0.026)	(0.026)	(0.041)	(0.058)
5-10 years	0.492***		0.610***	0.298***	0.120	0.129
	(0.080)		(0.199)	(0.061)	(0.086)	(0.109)
10+ years			1.364***			
			(0.202)			
Non-manager	0.064***	-0.009	0.032	0.100***	-0.048*	-0.004
	(0.017)	(0.045)	(0.043)	(0.031)	(0.027)	(0.011)
Medium manager	0.254***	0.304**	0.410***	0.366***	0.144*	0.093
	(0.043)	(0.132)	(0.057)	(0.066)	(0.085)	(0.059)
Top manager				0.731***		
				(0.121)		
Hanoi	-0.005	0.029*	-0.018*	0.102***	0.062***	0.017
	(0.012)	(0.017)	(0.011)	(0.020)	(0.022)	(0.013)
HCM	0.072***	0.090***	-0.018*	0.066***	-0.022	-0.021*
	(0.012)	(0.017)	(0.010)	(0.022)	(0.019)	(0.011)
Computer	-0.048***	-0.028**	0.040**	-0.044**	-0.006	-0.011
	(0.010)	(0.012)	(0.016)	(0.022)	(0.026)	(0.041)
Software	-0.002	0.021	-0.015	0.073**	0.034	0.088
	(0.017)	(0.019)	(0.022)	(0.036)	(0.040)	(0.067)
Language	0.062***	0.094***	0.118***	0.194***	0.145***	0.163***
	(0.012)	(0.015)	(0.021)	(0.028)	(0.030)	(0.044)
Financial	0.016	-0.033**	-0.062***	0.131***	0.045	0.070
	(0.013)	(0.014)	(0.014)	(0.037)	(0.043)	(0.058)
People management	-0.003	-0.056*	-0.009	-0.092*	-0.034	0.063
	(0.025)	(0.034)	(0.035)	(0.052)	(0.053)	(0.091)
Project management	0.079***	0.130***	-0.004	0.055*	-0.014	-0.027
	(0.019)	(0.025)	(0.025)	(0.031)	(0.038)	(0.078)
Art	-0.012	0.018	-0.000	-0.023	-0.015	-0.017
	(0.026)	(0.036)	(0.054)	(0.029)	(0.032)	(0.050)
Character	-0.067***	-0.055***	-0.114***	-0.051	-0.058**	-0.118
	(0.014)	(0.017)	(0.022)	(0.049)	(0.028)	(0.092)
Cognitive	0.022*	0.035**	0.026	-0.025	-0.003	-0.053
	(0.011)	(0.015)	(0.019)	(0.021)	(0.026)	(0.034)
Customer service	0.038***	0.044***	0.083***	0.013	0.016	0.050
	(0.012)	(0.015)	(0.022)	(0.022)	(0.028)	(0.045)
Social	0.004	-0.009	0.005	0.036**	0.035	0.016
	(0.010)	(0.013)	(0.016)	(0.016)	(0.023)	(0.035)
Writing	0.044*	-0.007	0.011	-0.097*	-0.126**	0.124
	(0.023)	(0.032)	(0.036)	(0.052)	(0.052)	(0.209)
Obs	4,034	2,770	2,753	1,442	1,154	1,909
R2	0.499	0.412	0.471	0.631	0.613	0.660

Notes: The table presents the beauty effect for women and men. Columns (1) and (4) show results from PSM for females and males, respectively. Columns (2) and (5) show results from CEM for females and males, respectively. Columns (3) and (6) show results from text matching for females and males, respectively. Time, industry, job title, and firm size fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table A2.4. Robustness test for beauty effect with maximum salary

	(1) PSM	(2) CEM	(3) Text matching	(4) PSM	(5) CEM	(4) Text matching
Attractiveness	0.067***	0.072***	0.038**	0.025	0.017	0.050
	(0.009)	(0.013)	(0.018)	(0.017)	(0.027)	(0.048)
Vocational training	0.002		0.322***	0.039		-0.204***
	(0.059)		(0.083)	(0.055)		(0.023)
Associate degree	-0.035**	-0.037	-0.002	0.029	0.005	-0.022*
	(0.016)	(0.028)	(0.012)	(0.025)	(0.037)	(0.012)
Bachelor's degree	0.037	0.086**	0.075**	0.116***	0.153**	-0.019
	(0.023)	(0.039)	(0.031)	(0.041)	(0.061)	(0.068)
Other educations	-0.014		-0.272**	0.054		0.085
	(0.096)		(0.127)	(0.060)		(0.068)
1-2 years	0.061***	0.049***	0.047***	0.022	0.040	0.002
	(0.011)	(0.014)	(0.011)	(0.027)	(0.040)	(0.051)
2-5 years	0.277***	0.217***	0.220***	0.163***	0.169***	0.151**
	(0.021)	(0.032)	(0.029)	(0.031)	(0.052)	(0.068)
5-10 years	0.490***		0.633***	0.308***	0.092	0.086
	(0.080)		(0.172)	(0.068)	(0.107)	(0.130)
10+ years			1.271***			
			(0.177)			
Non-manager	0.083***	0.041	0.048	0.148***	-0.056	-0.037**
	(0.019)	(0.049)	(0.052)	(0.035)	(0.035)	(0.014)
Medium manager	0.280***	0.315**	0.396***	0.379***	0.205*	0.099
	(0.048)	(0.131)	(0.067)	(0.078)	(0.122)	(0.072)
Top manager				0.747***		
				(0.141)		
Hanoi	-0.004	0.054***	-0.033**	0.117***	0.020	0.027*
	(0.014)	(0.019)	(0.013)	(0.023)	(0.025)	(0.016)
HCM	0.072***	0.107***	-0.035***	0.076***	-0.059**	-0.033**
	(0.014)	(0.020)	(0.012)	(0.026)	(0.023)	(0.014)
Computer	-0.063***	-0.038***	0.047**	-0.048*	-0.047	0.007
	(0.011)	(0.014)	(0.019)	(0.026)	(0.032)	(0.051)
Software	-0.003	0.010	-0.017	0.076*	0.020	0.093
	(0.018)	(0.021)	(0.026)	(0.044)	(0.047)	(0.082)
Language	0.062***	0.090***	0.131***	0.181***	0.104***	0.096*
	(0.013)	(0.017)	(0.024)	(0.030)	(0.037)	(0.052)
Financial	0.030**	-0.028*	-0.068***	0.148***	0.117**	0.131**
	(0.014)	(0.016)	(0.017)	(0.040)	(0.046)	(0.066)
People management	0.012	-0.053	0.005	-0.068	0.002	0.052
	(0.028)	(0.037)	(0.041)	(0.057)	(0.068)	(0.103)
Project management	0.079***	0.106***	0.001	0.033	-0.048	-0.125
	(0.020)	(0.027)	(0.028)	(0.034)	(0.044)	(0.079)
Art	-0.002	0.040	-0.002	-0.035	0.028	0.009
	(0.030)	(0.042)	(0.058)	(0.033)	(0.034)	(0.058)
Character	-0.056***	-0.039*	-0.116***	-0.060	-0.074**	-0.179
	(0.016)	(0.020)	(0.025)	(0.059)	(0.032)	(0.112)
Cognitive	0.024*	0.044**	0.041*	-0.024	0.038	-0.020
	(0.013)	(0.018)	(0.022)	(0.023)	(0.031)	(0.039)
Customer service	0.044***	0.056***	0.067***	0.018	0.008	0.069
	(0.014)	(0.018)	(0.025)	(0.025)	(0.033)	(0.048)
Social	-0.000	-0.019	0.012	0.040**	0.028	0.019
	(0.011)	(0.015)	(0.018)	(0.019)	(0.026)	(0.045)
Writing	0.050**	0.030	0.012	-0.135**	-0.153**	0.212
	(0.025)	(0.032)	(0.038)	(0.058)	(0.066)	(0.316)
Obs	4,034	2,770	2,753	1,442	1,154	1,909
R2	0.455	0.349	0.370	0.581	0.565	0.561

Notes: The table presents the beauty effect for women and men. Columns (1) and (4) show results from PSM for females and males, respectively. Columns (2) and (5) show results from CEM for females and males, respectively. Columns (3) and (6) show results from text matching for females and males, respectively. Time, industry, job title, and firm size fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Chapter 3. Wage Dollarisation: Evidence from Online Job Vacancy Data

3.1. Introduction

A distinctive feature of many highly inflationary economies is the persistent use of foreign currencies by domestic agents, either partially or fully, known as dollarisation. Dollarisation refers to the substitution of the domestic currency for a foreign currency (most commonly the US dollar) for different purposes. Besides financial dollarisation, in which case foreign currencies are used in deposits and credit in the banking system, foreign currencies can also be employed as the unit of account or medium of exchange (i.e., payment dollarisation). One of the main driving forces for the use of foreign currency – a more stable currency – is to avoid the adverse consequences caused by high inflation. However, even when inflation has declined and remains stable for a period, dollarisation still persists in many emerging countries regardless of government efforts to fight against it.

A plausible explanation for such hysteresis effect is that economic agents also rely on foreign currency for other purposes, i.e., hedging against exchange rate fluctuations/domestic currency devaluation, which is under-examined and needs further research. Thus, in this paper, we conduct an empirical analysis to shed light on the use of foreign currency as a non-pecuniary benefit offered to workers to hedge against domestic currency depreciation - exchange rate benefit. More specifically, we examine the relationship between wage and exchange rate benefits as well as job/skill requirements associated with such benefit provision.

Previous studies have examined the impacts of (financial) dollarisation on different labour market dimensions, namely the aggregate labour supply, employment growth and unemployment rate (see for instance, Borjas and Fisher, 2001; Soto, 2009; Chidakwa and Munhupedzi, 2017). However, those studies either use macro-level data on labour market indices or data on the degree of financial dollarisation from the banking sector. In our paper, we use micro-level labour market data and focus our analysis on payment dollarisation.

Additionally, there are a few studies that examine payment dollarisation (i.e., De Zamaroczy and Sa, 2002; Drenik and Perez, 2020; Kubo et al., 2021). Yet, to the best of our knowledge, the literature has not yet provided any empirical analysis of wage dollarisation. Thus, our study is the first to study payment dollarisation in the labour market (i.e., setting wages in US dollars). This examination will have meaningful implications for the fight against dollarisation in emerging economies as it helps understand which labour market segments that have a high degree of dollarisation as well as the wage setting behaviours and skill demand by the employers.

To document the wage dollarisation, we employ an online job vacancies dataset that was collected weekly from one of the top three largest job boards in Vietnam over the period between February 2019 and December 2020. In total, there are more than 170,000 job postings covering more than 170 occupations. Detailed information for each job, including offered salary, required education qualification, experience level, job position, location and skill requirements, is provided. Thus, this wealthy dataset allows us to uncover the heterogeneous wage effects across different job types.

The paper starts off with a Mincer wage model to investigate the link between wage and non-wage compensations. Our results show that there exists a complementarity between wage and exchange rate benefits. That is, dollarised jobs can offer both a hedge against exchange rate risks and higher wages (around 40% to 46%) in comparison with normal jobs. Our results are robust to different matching models (i.e., propensity score matching, coarsened exact matching and text matching). After controlling for the potential presence of foreign firm wage effect, our results remain unchanged. The finding can be explained by the efficiency wage theory, which asserts that employers are willing to pay higher wages and higher benefits to increase workers' productivity.

We then conduct an analysis on heterogeneities in the wage effect of non-pecuniary benefits. We further show that while there are substantial heterogeneities across education levels, experience levels and job positions exist. Particularly, the complementarity between wages and benefits is most prevalent among less experienced workers and lower job levels.

We further explore the requirements for potential workers in offering exchange rate benefits. We find that setting wage in US dollars is similar to hiding wages in online job postings to the extent that they both target high-skill job segments. Specifically, they are more likely to require higher education qualifications and highly experienced workers and advertise for high job positions. Furthermore, we document that the wage dollarisation concentrates in the most developed cities. In addition, employers who set wages in US dollars seek workers with foreign language, writing and artistic skills.

Our paper contributes to three additional strands of literature. First, there are two competing views of whether there is a trade-off between wage and non-wage benefits. In one view, studies supporting the traditional compensation wage differentials theory widely document the positive wage effect of non-monetary benefits (see, e.g., Heywood, 2007; Eriksson and Kristensen, 2011; Fakih, 2014). In contrast, the alternative view shows a negative wage effect associated with higher non-pecuniary benefits (see, e.g., Dale-Olsen, 2006; Pailhé and Solaz, 2019). Hence, we complement this strand of literature by providing new empirical evidence on the complementary between wage and non-wage compensations.

Moreover, existing studies are mainly interested in various forms of benefits which are often held fixed for a period of time according to the work contract, such as family-friendly benefits (e.g., childcare provisions and maternity leaves), paid holidays, flexible work arrangements, assistance for elder care, extended health care and pension benefits. Unlike previous papers, our analysis focuses on a new type of benefit – exchange rate benefit, which varies with the exchange rate fluctuations. More specifically, receiving wage in US dollars in a time of high exchange rate fluctuations becomes relatively more beneficial for workers than in a time of low exchange rate volatilities.

Furthermore, our paper is also related to the literature that documents employers' wage setting strategies (see, e.g., Hall and Krueger, 2012; Brenzel et al., 2014). Particular attention has been devoted to the differences in skill requirements and job characteristics between explicit wage posting and hidden wage posting. For example, Banfi and Villena-Roldan (2019) show that job ads with explicit wages tend to target unskilled workers (i.e., low education level). Employers in three job markets (US, UK and Slovenia) are less likely to post a wage offer when searching for skilled workers (Brenčič, 2012). According to Michelacci and Suarez (2006), employers opt to post wages when their bargaining power is low and the workers' productivities are highly homogeneous. Also, firms that belong to the public sector or are unionised are usually forced to post wages explicitly. Yet, little is known about the strategy of setting wages in foreign currency. In this paper, we have information on wages quoted in US dollars and dollarised wage posting can be seen as a type of hidden wage posting since the wage level is not fixed and depends on the exchange rate fluctuations.

Finally, we contribute to the growing literature that makes use of publicly available online vacancies to study labour market dynamics. One line of work looks at trends in the labour market (i.e., automation, AI skills, flexible work arrangements) (Hershbein and Kahn, 2017; Turrell et al., 2018). Another line of research focuses on discrimination in the labour market (Woodhams et al., 2009; Kuhn and Shen, 2013; Chowdhury et al., 2018; Ningrum et al., 2020). Some other papers discover the job search behaviour of job seekers (Marinescu and Rathelot, 2018) or the skill premium (Deming and Kahn, 2018; Ziegler, 2020). Lastly, some studies concentrate the analysis on the aggregate labour market indices (e.g., labour supply and demand mismatch, labour market concentration, tightness, wage inflation) (Adrjan and Lydon, 2019; Azar et al., 2020). We complement this literature by linking the labour market dynamics with a macroeconomic condition, which is, in our case, dollarisation.

The remainder of the paper proceeds as follows. Section 3.2 provides information on the background and context of the paper. Section 3.3 describes the data used in this study and provides descriptive statistics. Section 3.4 outlines our empirical strategies. Section 3.5 discusses the findings. Section 3.6 set out additional analyses and robustness checking. Finally, section 3.7 concludes.

3.2. Background

The widespread use of US dollars in the Vietnamese economy has been noticed since 1988 when domestic banks were allowed to accept dollar-denominated deposits. Coupled with the Asian financial crisis in 1997 and the Vietnamese dong (VND) being a weak medium to store value, since then, Vietnam has always been recognised as being among the top highly dollarised countries in Asia and worldwide with its foreign currency deposits accounting for 20% to 40% of total bank deposits. Figure 3.1 plots the rate of foreign currency-denominated money deposited in banks. As we can see, although the ratio of foreign currency money to the total money decreases by more than half over time, it still remains high at 8.09% in 2019. This degree of dollarisation is comparable to economies such as Russia, Eastern European (e.g., Turkey, Greece) and Latin American countries (e.g., Brazil, Argentina).

[Insert Figure 3.1 here]

One might believe that since dollarisation is caused by high inflation, a monetary stabilisation should result in de-dollarisation. Yet, this is not necessarily the case. For example, like many Latin American countries, Vietnam has experienced dollarisation hysteresis, which is the situation where dollarisation becomes persistent and irreversible despite lower and stable inflation rates (Calvo and Gramont, 1992).

The Vietnamese government has implemented many policies to restrict the use of foreign currencies in the economy and fight against dollarisation over the last three decades. For example, some of the most recent de-dollarisation policies include the ordinance on foreign exchange control in 2013, which states that: "All transactions, payments, listings and advertisements of residents and non-residents must not be executed in foreign currency…". In 2015, the government set the USD deposit interest rate at 0% per annum. Most recently in 2019, banks are prohibited from making loans in foreign currency to firms.

Some research has identified causes for dollarisation hysteresis. The first explanation is based on the network externalities theory, which states that if economic agents can choose among several currencies for transactions, they will prefer the currency that is already widely used in the economy. The second explanation relies on the money illusion theory, where the US dollar might have a higher perceived value than the local currency. In particular, the price or wage quotation in US dollars can raise the perceived valuation of a product or a job (Raghubir and Srivastava, 2002).

Lastly, the persistence of dollarisation might be explained by the expected depreciation of the local currency. This means that the less risky currency – USD – will be preferred because it can preserve value better than the depreciating domestic currency (Valev, 2010). The more the exchange rate volatiles, the higher risk of a future depreciation of the domestic currency. This might be the case for the dollarisation in Vietnam. The VND-USD exchange rate from 2010 to 2019 is plotted in Figure 3.2. As we can see from the figure, the Vietnamese Dong has depreciated against the US dollar at a considerably high rate over time. For instance, in January 2010, one US dollar was worth more than VND 18,000, then reached more than VND 23,000 by December 2020.

[Insert Figure 3.2 here]

3.3. Data

3.3.1. Processing data

We examine the dollarisation in the online labour market during the 22-month period from February 2019 to December 2020. To conduct our analysis, we download publicly available job vacancies from one of the leading online job boards in Vietnam. A Python script was written to weekly scrape all vacancies posted on this job portal. In general, a job posting provides detailed information on the job title, industry, job level, job type, work location, job description, education level, experience level, job requirements, monthly salary (including both salary level and currency) and the recruiting company's name and a number of employees.

Our data are processed with the following steps. First, we removed job ads that do not advertise any salary information (i.e., an employer only states that salary is negotiable or competitive). Next, we used the daily exchange rate statistics to convert the salary amount stated in dollarised job posts to Vietnam dong[12]. For job posts advertising a range of salaries, the minimum wage

[12] USD - VND exchange rate data were downloaded from website https://www.investing.com.

level is selected[13]. Furthermore, part-time jobs and jobs that offer salaries lower than the obligatory minimum wage[14] were also excluded from our analysis.

Moreover, the job board has a separate job category for foreigners, we excluded this category from our analysis to only focus on jobs recruiting Vietnamese workers. Additionally, we removed oversea jobs to keep only jobs located in Vietnam. Finally, the last step was to de-duplicate observations in the sample. The limitation of online job vacancy data is that we do not have information on whether a job posting is filled or not, it usually stays on the site for a certain period of time and might be re-scraped every week. Hence, we rely on the job site's identification number to identify unique vacancies. After the screening process, our final sample consists of 173,106 job postings, of which 4,948 vacancies advertise wages in US dollars.

For each job posting, the full text of the job title is cleaned, and similar job titles are classified into one of the 176 groups. Marinescu and Wolthoff (2020) emphasized that relative to the standard occupational classifications, the job title can better capture the specialisation of different jobs, the level of the required experience and the hierarchy in job positions. In addition, the job title can account for more than 90% of the variance in the wage posted by the employers. Moreover, based on the information of the recruiting firm's number of employees, we categorise firms into different firm sizes. As can be seen in Figure 3.3, medium-sized companies of 100 to 499 workers dominate job postings in this job board, followed by smaller companies of 25 to 100 workers and larger firms of 1000 to 4999 workers.

[Insert Figure 3.3 here]

For job location, we use the information on the city/province where the job locates. As can be seen from Figure 3.4, most jobs have locations in Hanoi and Ho Chi Minh city (HCM). This is expected since these two cities are the most important cities in Vietnam in terms of politics and economics. Other important industrialised or port cities such as Da Nang, Hai Phong, Binh Duong or Bac Ninh also account for a high number of job postings on the job board. Moreover, we clean the job category text to obtain a set of job industries. We then group them into three location groups, which are Hanoi, HCM and others.

[13] Results for wage equations (reported in Online Appendix Table A3.3 and A3.4) remain robust when we choose the average wage level and maximum wage level instead.
[14] The minimum wage was set at VND 2,920,000 according to Article 91 of the Vietnam's Labour Code, became effective as of 1/1/2019.

[Insert Figure 3.4 here]

Subsequently, the next step is to extract skill requirements from the job requirement section. First, we use a Python package in Natural Language Processing to extract a set of unique keywords from the unstructured text of the job requirement. Second, we filter out irrelevant keywords and categorise skill-related keywords into different skill groups. Ten skill groups (i.e., cognitive, social, character, writing, customer service, project management, people management, financial, computer (general) and software (specific)) are identified following the work of Deming and Kahn (2018). Also, two additional skill groups (i.e., foreign language and artistic) are added to better capture the skill demand in the Vietnamese labour market. Hence, there are twelve distinctive skill categories in total. We consider a vacancy as requiring a particular skill if it mentions at least one of the keywords or phrases in that skill group (we do not take into account the frequency of mentioning the skill). Table 3.1 present the list of example keywords for each skill category.

[Insert Table 3.1 here]

3.3.2. Descriptive statistics

The summary statistics for the job posting characteristics by salary currency (i.e., VND vs USD) is presented in Table 3.2. Overall, firms tend to target the high-skill job segment when offering exchange rate benefits. Specifically, domestic currency jobs are more likely to offer lower wages than dollarised jobs: VND 9 million per month as opposed to VND 21 million per month. Moreover, employers appear to require a higher level of education in dollarised jobs than in normal jobs. Specifically, only a marginal fraction (14.1%) of dollarised ads specify a level of education below an Associate degree. While two third of normal jobs (i.e., 66.2%) require an Associate degree, only one-third of dollarised jobs (i.e., 34.9%) require such a degree. Nearly to third of dollarised jobs specifies a Bachelor’s degree and above, as opposed to one-fifth in normal jobs.

[Insert Table 3.2 here]

Regarding experience level, whereas nearly half of normal jobs require less than one year of experience, a half of dollarised jobs expect two to five years of work experience from the job candidates. As for job level, the share of dollarised jobs advertises manager positions is higher than that of normal jobs (24% vs 12%). Most jobs concentrate in the three largest cities, with the share of dollarised jobs slightly higher than that of normal jobs (86.6% vs 77.9%). We also find important differences in the characteristics of skill requirements. Jobs providing exchange rate benefits tend to specify software, language, artistic, cognitive, writing, people and project management skills. In contrast, jobs that offer wages in VND tend to require computer, customer service, character, social and financial skills.

The extent of dollarisation in the online labour market over the studied period is illustrated in Figure 3.5. Overall, we find that the share of dollarised ads ranges from more than 1% to nearly 7% during our sampling period and normally fluctuates around the 3% level.

[Insert Figure 3.5 here]

Table 3.3 panel A reveals the top ten occupations with the highest fraction of jobs that offer exchange rate benefits. The frequency of offering exchange rate benefits dominates among customer service managers, with more than half of ads offering a salary in US dollars. Following closely are developer and English interpreter occupations. Panel B reveals the top ten occupations with the highest numbers of jobs that offer exchange rate benefits. Among others, developer, manager and sales are three leading occupations with the largest numbers of dollarised job postings.These statistics suggest that wage dollarisation presents in a wide range of occupations.

[Insert Table 3.3 here]

Table 3.4 presents bivariate correlations for each pair of job characteristics and requirements. We can draw some main findings from this table. Firstly, cognitive skill and social skill requirements are positively correlated with each other, with years of education and experience required, and with most of the other eight job skills. This implies that they are general skills considered as necessary by firms across a wide range of occupations and industries. Secondly, the pairwise correlations between education/experience requirements and twelve skill groups are mainly positive but relatively low. This suggests that across job postings, the skill requirements can still vary significantly even after accounting for the required education and work experience.

[Insert Table 3.4 here]

3.4. Empirical strategy

3.4.1. The link between monetary and non-monetary benefits

Compensating wage differentials vs. efficiency wage theories

The theory of compensating wage differentials (CWD), initially conceived by Adam Smith in his work The Wealth of Nations (1776). In particular, the theory suggests that in a perfectly competitive labour market, jobs with disagreeable characteristics, poor working conditions or other adverse job amenities are expected to be compensated by higher wages, other things equal. Indeed, many studies found a substantial wage premium for the risk of work-related fatal injuries or illnesses (Kniesner et al.; Pouliakas and Theodossiou, 2013). For instance, using the US' Current Population Survey and Bureau of Labor Statistics data from 2012 to 2016, Mridha (2021) show evidence to support the compensating wage differentials for work-related fatal injuries, controlling for the impact of local unemployment on the risk premium.

In line with the predictions of the compensating wage differentials theory, numerous research shows a negative wage effect for non-pecuniary benefits. For example, Eriksson and Kristensen (2011) find that workers might be willing to sacrifice about 11% to 13% of the wage in return for a flexible working schedule. Likewise, employing the Canadian matched employer-employee data, Fakih (2014) show a trade-off between the provision of family-friendly benefits (i.e., childcare, elder care and extended healthcare) and wages.

However, many other studies find a negative or inconclusive link between wages and work-related risks (Hersch, 1998; Black and Kniesner, 2003). One of the reasons for the failure of compensating wage differentials theory is the existence of efficiency wage. Particularly, efficiency wage theory states that firms are willing to pay workers more than their market-clearing wages in exchange for increased effort and productivity (Akerlof, 1982; Yellen, 1984). For example, using data from the Brazilian Labor Monthly Survey in 2006-2007, Menezes and Raposo (2014) explain the wage differentials across firm sizes in Brazil by the efficiency wage theory. They show that large firms offer higher wage premium for its workers to increase the labour effort. With the evidence of positive relationships between wages and labour effort as well as job duration.

Consistent with the predictions of the efficiency wage differentials theory, previous papers show a positive wage effect for non-pecuniary benefits. For instance, Gariety and Shaffer (2001) show a wage premium of 6.2% for men and 6.7% for women among jobs that offer flexible time and work arrangements. In the same vein, Haynes and Sessions (2013) document that jobs in the public sector offer both pensions and 12% to 13.9% higher salaries.

Empirical analysis

Receiving wages in foreign currency is a benefit for workers, which can help them hedge against exchange rate fluctuations. We follow the estimating framework by Schiller and Weiss (1980) to test for the compensating wage differentials vs. efficiency wage theories using the following model:

$$LogWage_{it} = \beta_0 + \beta_1 USD_{it} + Controls_{it}\beta_2 + FixedEffects + \varepsilon_{it} \quad (3.1)$$

Where $LogWage_{it}$ is the logarithm of offered salary for job posting i posted on date t. USD_{it} is a dummy variable which equals one if the vacancy posts salary in US dollar and zero otherwise. $Controls_{it}$ is a vector of covariates including job characteristics that are likely to affect wage (i.e., 12 skill groups, beauty requirement, education and experience level, job position and location). These control variables can be commonly found in wage determination studies (Brown, 1980; Rosen, 1974). *FixedEffects* includes quarter, industry, job title and firm size fixed effects and ε_{it} is an error term. If the sign of β_1 is positive, a complementary relationship between wage and non-wage compensations exists. And vice versa, if the sign of β_1 is negative, there is a trade off between wage and non-wage compensations.

Before estimating model (3.1) using the fixed effect estimator, we employ three different matching approaches, namely Propensity score matching, Coarsened exact matching and text matching. The reason for using matching is that because higher-skilled applicants may self-select into applying job offering exchange rate benefits, we need to control for any possible selectivity bias. Particularly, there might exist job characteristics that can affect both treatment assignment (posting wage in USD) and the outcome (i.e., wage), which are called confounders. And matching simply represents a method of preprocessing data to control for these potential confounders by balancing the distribution of vacancy characteristics between the treated and control groups.

3.4.1.1. Propensity score matching

The first adopted approach is Propensity score matching. Using this method, we can match jobs based on a single dimension – the propensity score, which is defined as the likelihood of receiving the treatment (i.e., posting wage in USD) given the observed job characteristics (Rosenbaum and Rubin, 1983b). We start by estimating the propensity score using the following Probit model:

$$E(USD_{it}|X) = P(USD_{it} = 1|X) = \Phi(\beta_0 + Skill_{it}\beta_1 + Controls_{it}\beta_2 +$$
$$Quarter\ dummies + Industry\ dummies + Occupation\ dummies + \varepsilon_{it}) \quad (3.2)$$

Where Φ is the cumulative standard normal distribution function, that is: $\Phi(z)=P(Z\leq z)$ and Z follows a standard normal distribution N (0,1); i and t refer to job posting and posted date, respectively. USD_{it} is a dummy variable which equals one if the vacancy posts wage in USD and zero otherwise. $Skill$ is a vector of 12 skill dummy variables, which indicates whether a certain skill is mentioned in the job posting. Moreover, we include in the model a set of control variables, including dummies for beauty requirement, experience level, education level, work location, and job position. We also add quarter, job titles and job industry dummies and an error term into our model.

We then rely solely on the estimated propensity score to match vacancies in treated and controls groups. We use the nearest neighbour matching algorithm (on a one-to-one basis), which match job ads that are closest in terms of their propensity scores. We match without replacement, which means that a vacancy in the control group cannot be used more than once. We also impose common support and use a caliper of 0.1%. Using the 0.1% caliper is equivalent to choosing a job post from the control group as a match for a job post in the treated group that lies within 0.1% of the propensity range and is closest in terms of the propensity score. As a result, our matched sample contains 8,419 matched vacancies.

3.4.1.2. Coarsened exact matching

Our second approach is Coarsened exact matching – CEM (Iacus et al., 2011). This method simply matches job vacancies by a set of job characteristics that have been "coarsened". Coarsening a covariate refers to splitting its value into different bins. Then, CEM will perform an exact match with the coarsened data, ensuring that treated and control jobs, which have characteristics of the same bins will be matched. This process can help reduce the number of matching values for a given job characteristic to increase the number of final matches. One of the advantages of CEM is that it does not require a balancing check. Due to the exact matching process for coarsened data, our post-matched sample size is reduced to 4,182 job postings.

3.4.1.3. Text-based matching

In the last approach – text matching, we pre-match vacancies based on the job titles first. To be specific, we split vacancies into 176 subsamples for 176 job titles. For each subsample, we use text matching technique to find a job in the control group, which has the most similar job descriptions text as the job in the treated group. To this end, we remove all stop words[15] that do not add additional meaning to a text, remove special characters and punctuations as well as

[15] Some examples of stop word are: a, an, the, to, but, how, what etc.

convert all the text to lowercase. Then we employ the Python package *TextDistance* to compute the distance between two text documents. This Python package employs the Jaccard similarity algorithm to calculate the text similarity as follows:

$$Jaccard(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \text{ (3.3)}$$

Where $|A \cap B|$ is the number of common words shared by both document A and document B and $|A \cup B|$ is the number of unique words which appears in either document A and B or both. Hence, the Jaccard score is simply the share of common words in two texts to the sum of words that are present in either of the two texts. The value of this index ranges from zero to one. If the score equals one, two documents are identical. The lower the score is, the more different the two documents are.

For each vacancy in the treated group, we compute the similarity indices between itself and each of the vacancies in the control group and then select the job posting in the control group which has the highest Jaccard score. For this matching technique, we allow for replacement, which means that a vacancy in the control group can be employed more than once. We keep only matched pairs with a similarity index of at least 60%, which results in a sample of 3,433 text-matched vacancies.

3.4.2. Heterogeneities in the link between wage and non-wage compensations

In this section, our attention is focused on the extent to which the link between monetary and non-monetary benefits is subjected to various types of heterogeneity. To examine whether such heterogeneities exist, first we investigate whether the wage premium/penalty for non-pecuniary benefit is different across jobs that require different levels of education. Second, we aim to capture the effects of job experience heterogeneity by incorporating the interaction term between USD job indicator and experience categories into the model (3.2). Third, we examine the link between monetary and non-monetary compensations across different job positions (i.e., entry-level, employee, medium manager and top manager).

3.4.3. Characteristics of dollarised job postings

To examine the worker skills/job requirements that employers target in trade-off with offering wages in foreign currency, we follow Brenčič (2012) and Banfi and Villena-Roldan (2019) in estimating the following model:

$$P(USD_{it} = 1|X) = \Phi(\beta_0 + \beta_1 SKILL_{it} + Controls_{it}\beta_2 + \varepsilon_{it}) \text{ (3.4)}$$

Where the dependent variable, USD_{it}, is a dummy variable indicating a dollarised job. $SKILL_{it}$ is a set of skill dummies. $Controls_{it}$ includes job characteristics such as beauty requirement, education and experience level, job position, job location, quarter, industry, job title and firm size. ε_{it} is an error term. We estimate the equation (3.4) using the Logit model.

Moreover, since the proportion of dollarised vacancies in the total vacancies in our sample is modest (i.e., approximately 4%), we employ the rare event Logit as an additional estimation model for the equation (3.4). As suggested by King and Zeng (2001), if one uses normal logistic regression to estimate the probability of rare events, one is likely to underestimate the probability of rare events. To address such concern, we apply the Firth correction for logistic regression as a robustness check.

3.5. Common support and balance diagnostics

3.5.1. PSM

In this subsection, we need to check the area of common support and the degree of overlap in the propensity score between jobs in treatment and control groups. Generally, common support means that any set of job characteristics that are found in the treated group can also be found in the control group (Bryson et al., 2002). This is necessary as the treatment effect cannot be estimated outside the area of common support. Following Caliendo and Kopeinig (2008), we compare the minima and maxima of the propensity score distribution in the treatment and control groups. That is, observations with a propensity score lower than the minimum or higher than the maximum score of the opposite group will be discarded. As a result, the common support condition is satisfied for 4,205 jobs in the treated group, and the remaining 508 jobs are discarded.

In the next step, following Austin (2011), we examine the matching quality by considering the standardised bias score, which compares the difference in mean of job characteristics between the treated and untreated groups. As suggested by Normand et al. (2001), a standardised difference of equal or less than 10% represents a trivial difference in the mean of job characteristics between treated and untreated groups. The results plotted in Figure 3.6 reveal that all job characteristics across two groups obtain balance after matching, with the highest standardised bias being only 5%.

[Insert Figure 3.6 here]

3.5.2. CEM

Without matching, there might be large differences in the job characteristics (i.e., covariate imbalance) between jobs offering exchange rate benefits and those that do not. The overall covariate imbalance is reported by the L1 statistic, which represents the percentage of overlap between the density of the two distributions of treated and control groups. The minimum value of zero (i.e., L1 = 0) implies perfect global balance, high values of L1 indicate a high imbalance between jobs in treated and control groups. The results reported in Table 3.5 reveal that there is a significant overall covariate imbalance in the unmatched sample. After CEM matching, we observe a substantial decrease in the L1 statistic from 0.9633 to 0.1856. Hence, we obtain the overall improvement in covariate balance after matching.

The unidimensional measures of imbalance computed for each job characteristic are also reported in Table 3.5. In general, our matching algorithm has achieved good balances in the means, marginal and joint distributions of most job characteristics. The only exception is the job title and job industry, where there are still some remaining imbalances but at trivial extent.

[Insert Table 3.5 here]

3.6. Result discussion

3.6.1. Relationship between monetary and non-monetary benefits

Table 3.6 reports the regression results from estimating equation (3.2) on different matched samples. Columns (1), (2) and (3) present the estimation results on the PSM, CEM and text-matched sample, respectively. In column (1), the coefficient estimated on the non-pecuniary benefit indicator (i.e., exchange rate benefits) is positive and significant at the one percent level at 0.437. Similar results are reported in columns (2) and (3). That is, the estimated coefficients on non-pecuniary benefit are also positive and significant at the one percent level (i.e., 0.435 and 0.296). Taken as a whole, these findings suggest that there exists a complementarity between monetary and non-monetary compensations. Particularly, dollarised jobs can offer both the hedge against exchange rate risks and higher wages (around 29.6% to 43.5%) in comparison with other jobs.

[Insert Table 3.6 here]

This finding is in line with previous studies that also document the positive wage effect of non-pecuniary benefits and good working conditions. More specifically, using a statewide dataset of individual employees in the US, Johnson and Provan (1995) find that the use of family-friendly practices (i.e., childcare) is associated with an increase of 18% in earnings. Additionally, flextime is associated with 52% increase in earnings among professional women. A positive link between wage and family-friendly practices is also reported by Gariety and Shaffer (2001). They employ a nationally representative sample from the US Current Population Survey. After controlling for human capital, occupation, industry, and reasons for which employees desire family-friendly practices, their analysis reveals that flextime is associated with nearly 7% higher wage.

Likewise, using French matched employee–employer data, Pailhé and Solaz (2019) find a significant wage premium of 51.9% for women working in a family-friendly workplace, where employers provide childcare provisions – in-cash or in-kind child-related benefits and flexible work arrangements. Employing the US Panel Study of Income Dynamics, Altonji and Usui (2007) also find a positive relationship between wages and the provision of paid holidays. That is, a one-week increase in paid holiday leave is associated with a 5.4% rise in salary for both salaried and hourly-paid workers. Exploring data from the British Household Panel Survey, Haynes and Sessions (2013) find a wage premium of 3.9% and 5.5% associated with pension benefits for all workers and public sector workers, respectively.

There are several potential explanations for the positive link between wage and non-wage compensations. The predominant theoretical perspective for explaining the negative effect of job benefits on workers' wages is efficiency wage theory (Stiglitz, 1976). According to this view, employers are willing to pay higher wage and higher benefits to increase the productivity of workers. In other words, employees who receive non-monetary benefits might have higher incentives to work more productively.

In addition, it might be the case that there exist some unobserved confounding factors – un-stated worker's characteristics that employers target when offering higher wages and currency benefits. Those characteristics can only be observed in a later stage of the recruitment process (i.e., interview), which we cannot control for using our dataset.

The magnitude and sign of coefficients for the other control variables are also comparable to those found in prior studies on wage determination, such as Autor and Handel (2013) and Deming and Kahn (2018). Particularly, the positive coefficients (although statistically insignificant for some) on education categories imply a higher wage premium for a higher level of education. Similarly, the higher experience categories will lead to higher wage premia. In the same vein, higher job positions are associated with higher wage premia.

In addition, it is likely that earnings differ across geographical areas. In fact, we observe significant wage premia (in the range between 7.2% and 7.8%, depending on cities and econometric models) for the two largest cities (i.e., Hanoi and HCM city) relative to other cities. This finding is in line with a model of wage disparities across local labour markets with workers of different skill levels and imperfect labour mobility (Enrico, 2011). Companies in highly developed urban areas are more likely to face higher demand for skills and be more productive (Glaeser and Mare, 2001). Thus, to attract the most talented workers, firms have to post higher salaries in developed urban areas than in less developed areas.

Results for the wage returns to different skill groups are mixed. More specifically, a positive wage premium can be observed for language, financial, software, artistic and people management skills. On the contrary, character, basic computer and writing skills are associated with lower earnings. Listing these groups of skills may represent a signal of lower-paying works where obedience, basic office or managing paperwork skills are most often requested, even if these skills themselves can bring added value to all occupations (Deming and Kahn, 2018).

3.6.2. Heterogeneity of the wage effect of exchange rate benefits

The first column of Table 3.7 shows that there are significant differences in wage and exchange rate benefit relation across education categories, including High school, Associate degree and Bachelor's degree. In addition, we show differences in wage and non-wage benefits relationship across experience levels and job positions. The second column of Table 3.7 reports a substantial variation in the wage and non-wage benefits that can be attributed to experience levels. The average earning premium is at its highest among those with little to zero previous work experience and at its lowest among those with at least a university degree. In other words, the complementarity between wages and benefits is most prevalent among less experienced workers.

Column (3) of Table 3.7 shows considerable heterogeneity across job positions. At a beginner level, workers can benefit more in dollarised jobs, which provide much higher wage and exchange rate benefits than normal jobs. In managerial positions, the wage differentials between jobs with exchange rate benefits provision and normal jobs are not as large as in lower-level jobs.

[Insert Table 3.7 here]

3.6.3. Characteristics of dollarised job postings

Estimated coefficients for education requirements from logit and rare event logit models reported in Table 3.8 suggest that online vacancies quoting salary in US dollars tend to require higher education qualifications than those quoting salary in local currency. More specifically, compared to the high school level, jobs setting wages in US dollar are 0.4 to 0.8 percent more likely to require an associate degree; and 1.8 to 2.6 percent more likely to request a Bachelor's degree.

Regarding work experience, job postings that offer exchange rate benefits are 0.6% to 0.9% less likely to request one to two years of experience. Likewise, a dollarised ad shows a higher chance to request more work experience (i.e., 2-5 years, 5-10 years and more than ten years) than zero to little work experience.

Turning to the job position, we can find similar patterns for the required education and work experience. Particularly, jobs offering exchange rate benefits are 0.8 to 0.5 percent more likely to be at a medium-level managerial position than a new entry-level. They are also 2.4 to 1.9 percent more likely to be at the top managerial position than new entry-level.

These results imply that job ads quoting wages in US dollars have similar characteristics as those that use hidden wage strategy (i.e., not posting salary number(s) explicitly) in the sense that they all target higher-skilled workers (Brenčič, 2012; Banfi and Villena-Roldan, 2019). In our case, quoting wages in foreign currency serves as a signal of jobs in the high-skilled segment.

With regards to skill requirements, employers who advertise salaries in foreign currency are more likely to specify foreign language and soft skills such as writing and artistic skills in job requirements. This finding is in line with previous literature which documents the importance of foreign language competence in the labour market, especially for emerging economies (Ginsburg and Prieto-Rodriguez, 2011; Di Paolo and Tansel, 2015; Lewinski, 2019). In contrast, the likelihood that employers list customer service, character, general computer and project management skills decline when they offer exchange rate benefits.

Jobs advertising salaries in US dollars are more likely to locate in the two biggest cities. This finding is in line with Duma (2014), Siregar and Chan (2014), Aiba and Tha (2017), and Odajima (2017) who document the predominance of dollarisation in metropolitan areas in comparison to rural areas. One of the possible explanations for the higher usage of US dollars in these two cities is the difference in industrial distributions across regions. That is, there is a higher degree of dollarisation in locations with a high concentration of foreign-related industries such as exporting, garment, tourism, or foreign direct investment-intensified sectors (Odajima and Aiba, 2017).

[Insert Table 3.8 here]

3.7. Additional analysis and robustness tests

3.7.1. Foreign firm wage effect

It is plausible that higher wages may be driven by the foreign firm wage effect. That is, jobs paying salary in domestic currency might belong to domestic firms and jobs paying salary in foreign currency might belong to foreign firms. Additionally, it has been documented in many previous studies that foreign-owned firms pay higher wages than local firms, given the same worker' characteristics (Feliciano and Lipsey, 2006; Hijzen et al., 2013; Malchow-Møller et al., 2013).

Hence to test this alternative explanation, we re-run model (3.2) for the sub-samples of domestically-owned and foreign-owned firms separately. We manually collect data on firm ownership status from the website http://www.thongtindoanhnghiep.co – a website that provides information on Vietnamese enterprises – by using the full company name in the job advertisement. The results reported in Table 3.9 show that the point estimates for the wage effect of exchange rate benefit remain positive and significant among both domestically-owned and foreign-owned firms (i.e., 44.9% and 29.1%, respectively). In brief, after accounting for the ownership status of recruiting firms, the wage premium for jobs offering exchange rate benefits persists, which supports the positive association between monetary and non-monetary benefits.

[Insert Table 3.9 here]

3.7.2. Further exact matching

We restrict our analysis to a small sample of vacancies with exactly the same job title (i.e., we match vacancies based on exact job title text) posted by the same company. The exact matching process based on job title and company name results in a subset of 102 job postings of 28 distinct job positions[16] advertised by 26 companies, where there is changes in currencies used

[16] Due to the nature of our dataset, we do not have information on whether a job posting is filled or not. Hence we rely on the ID number assigned to each job vacancy to identified unique vacancies. There are two cases when identical job ads (with different posted time) can appear in our dataset. The first case might be that vacancies can be re-advertised by the employers if it has not been filled after a period of time, hence they can be assigned a new

to advertise salary. We observe that 64% of jobs (i.e., eighteen job positions) switch from local currency to the US dollar. The average wage premium for dollarised jobs remains positive at around 39.98%.

3.7.3. Why do firms set wages in foreign currency?

One might raise a question is that instead of offering currency benefits, why employers do not just increase the wage level and set it in domestic currency? It may be the case that firms would like to post higher wages but not too high compared to the market's wage level. A possible explanation for such behaviour is that setting high wages might lead to a reduction in the number of job applications. More specifically, Marinescu and Wolthoff (2015) demonstrate that high-wage jobs attract significantly lower job applications than low-wage jobs. In the same vein, using a US survey, Faberman and Menzio (2018) show a negative link between posted wages and the number of applicants after controlling for occupation and industry.

Furthermore, as suggested by Ekman (2013), firms might have a preference for setting wages in foreign currency to match the incomes with expenditures in terms of currency. Firms paying wages in local currency (i.e., agriculture, public sector, small enterprises) are firms that are mainly producing for the domestic market, hence receiving revenues mainly in local currency. In contrast, firms that obtain their revenues in foreign currency (i.e., bigger firms, banks, import/export firms) are more likely to pay wages in foreign currency.

3.8. Conclusion

This study explores the use of foreign currency (i.e., US dollar) to set wages in online job vacancies. Particularly, we look at the complementarity between monetary reward (wage) and non-monetary reward (receiving wage in US dollars to hedge against exchange rate volatilities). Further, we examine various heterogeneities across education levels, experience levels and job positions. Finally, we investigate the skill requirements associated with offering exchange rate benefits.

We contribute to the literature in four ways. First, to the best of our knowledge, our study is the first to analyse payment dollarisation in the labour market (i.e., setting wages in US dollars). Second, we provide new empirical evidence on the complementary between wage and non-wage compensations using a new type of benefits – exchange rate benefits – that fluctuate depending on exchange rate volatilities. Third, we consider wage setting in foreign currency as

ID number. The second case might be that the employers are looking for new employees, hence a new position is created.

a type of hidden wage strategy. Fourth, we contribute to the growing literature that makes use of publicly available online vacancies to study labour market dynamics by linking the labour market dynamics with macroeconomic conditions (i.e., dollarisation).

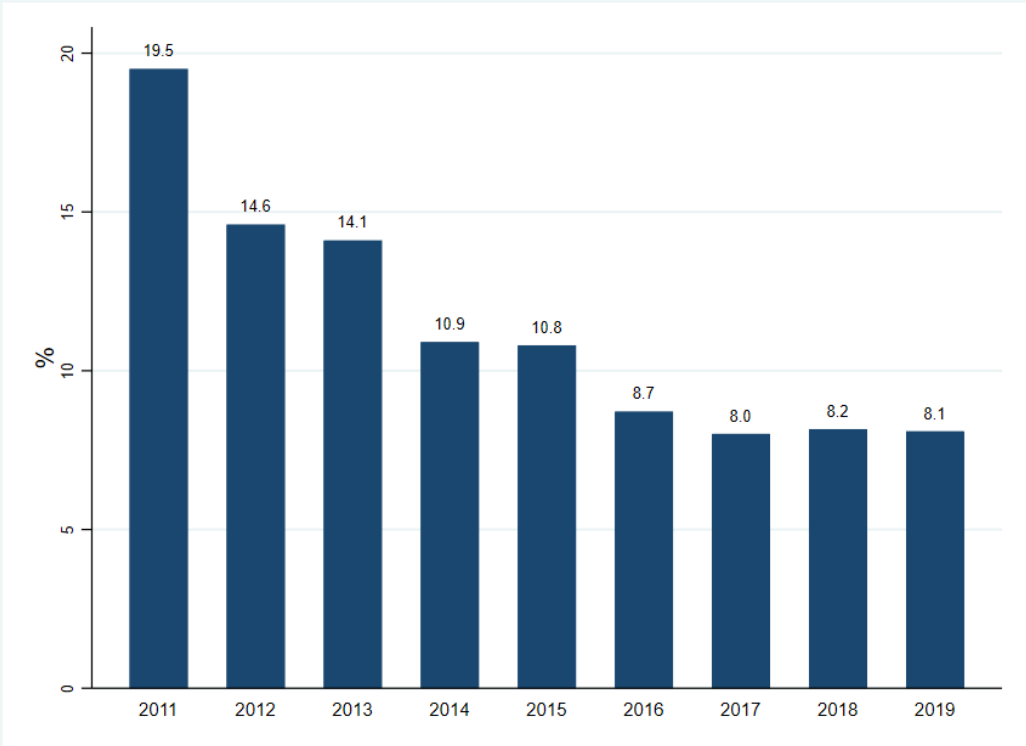
Our empirical analysis yields three main conclusions. First, we find evidence of complementarity between monetary and non-monetary compensations. That is, dollarised jobs can offer both a hedge against exchange rate risks and higher wages (around 40% to 46%) in comparison with normal jobs. Our results are robust to different matching models (i.e., propensity score matching, coarsened exact matching and text matching). After controlling for the potential presence of foreign firm wage effect, our results remain unchanged. The finding can be explained by the efficiency wage theory, which asserts that employers are willing to pay higher wages and higher benefits to increase the productivity of workers.

Second, this complementary relationship varies across different work experience categories and job positions. We show that while there are no significant differences in the link between wage and exchange rate benefits across education levels, substantial heterogeneities across experience levels and job positions exist. Particularly, the complementarity between wages and benefits is most prevalent among less experienced workers and lower job levels.

We further explore the requirements for potential workers in offering exchange rate benefits and find that setting wage in US dollars is similar to hiding wages in online job postings to the extent that they both target high-skill job segments. Specifically, they are more likely to require higher education qualifications and highly experienced workers and advertise for high job positions. Furthermore, we also document that dollarisation concentrate in the most developed cities. In addition, employers who set wages in US dollars emphasize the importance of language, writing and artistic skills.

Figures

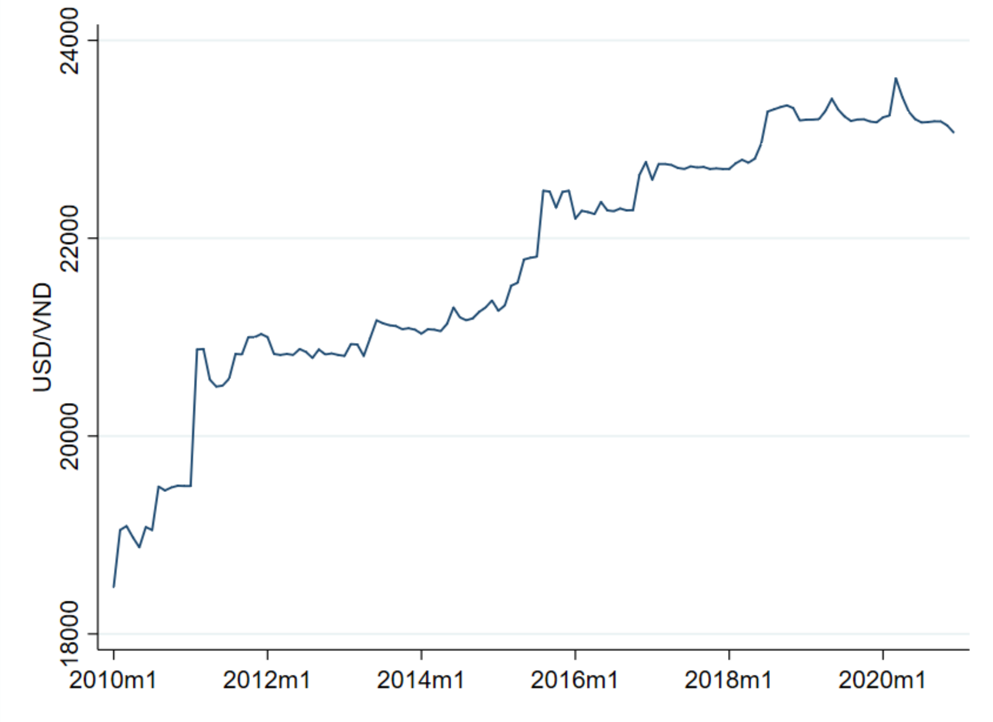
Figure 3.1. The ratio of foreign currency deposits to the money supply in Vietnam from 2011 to 2019



Notes: This figure shows the ratio of foreign currency deposits to the money supply (i.e., M2) over the period from 2011-2019 (Data source: Tran, 2021[17]).

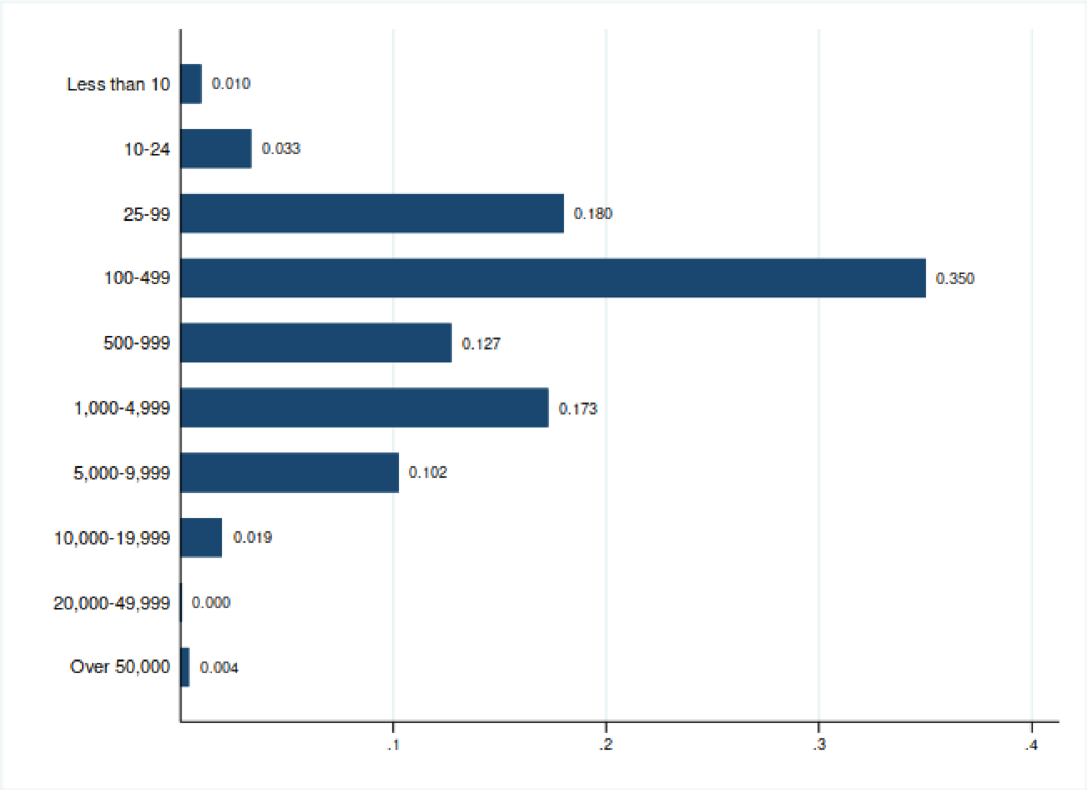
[17] https://kinhtevadubao.vn/hien-tuong-do-la-hoa-o-viet-nam-thuc-trang-va-giai-phap-20647.html, accessed 31st July 2022.

Figure 3.2. USD/VND exchange rate over time



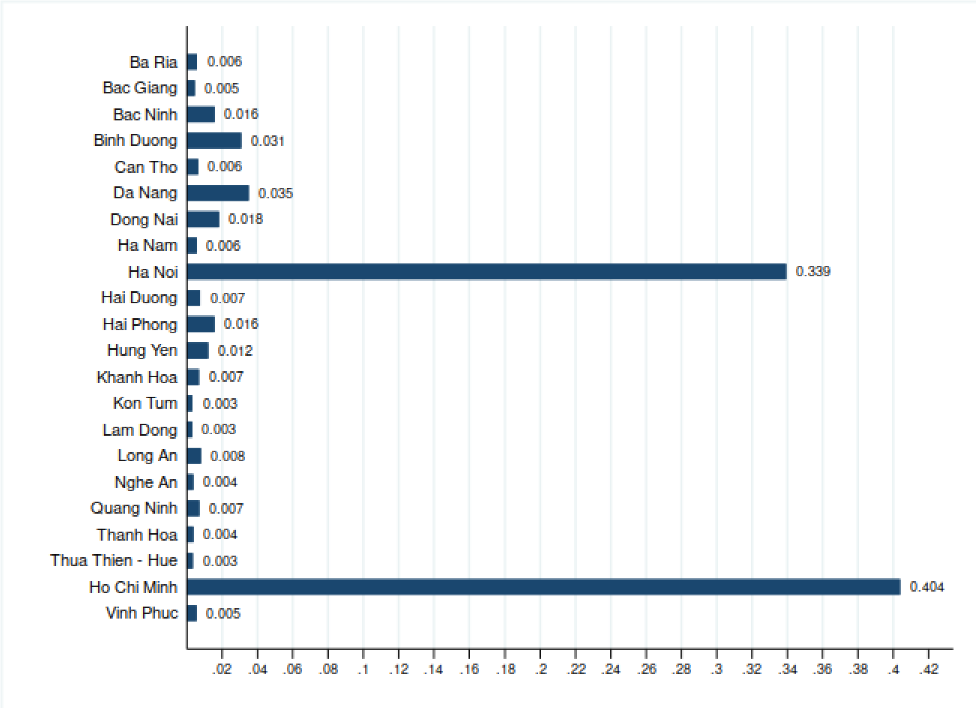
Notes: This figure shows the monthly USD/VND exchange rate over the period from January 2010-December 2020.

Figure 3.3. Share of job ads by firm sizes

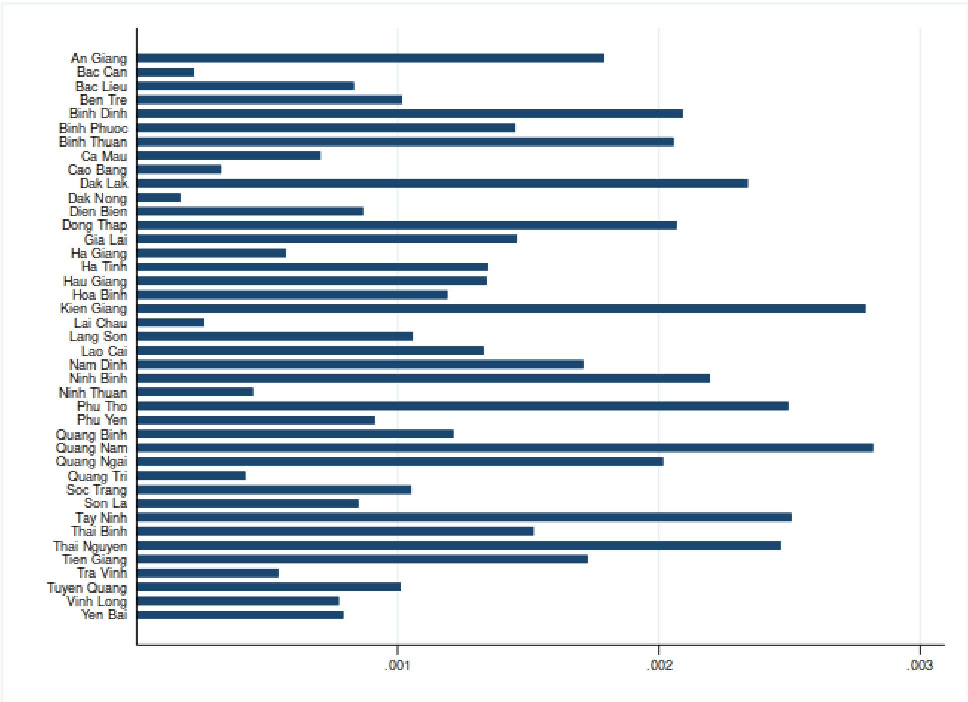


Notes: This figure shows the share of vacancies by firm size.

Figure 3.4. Share of job ads by locations



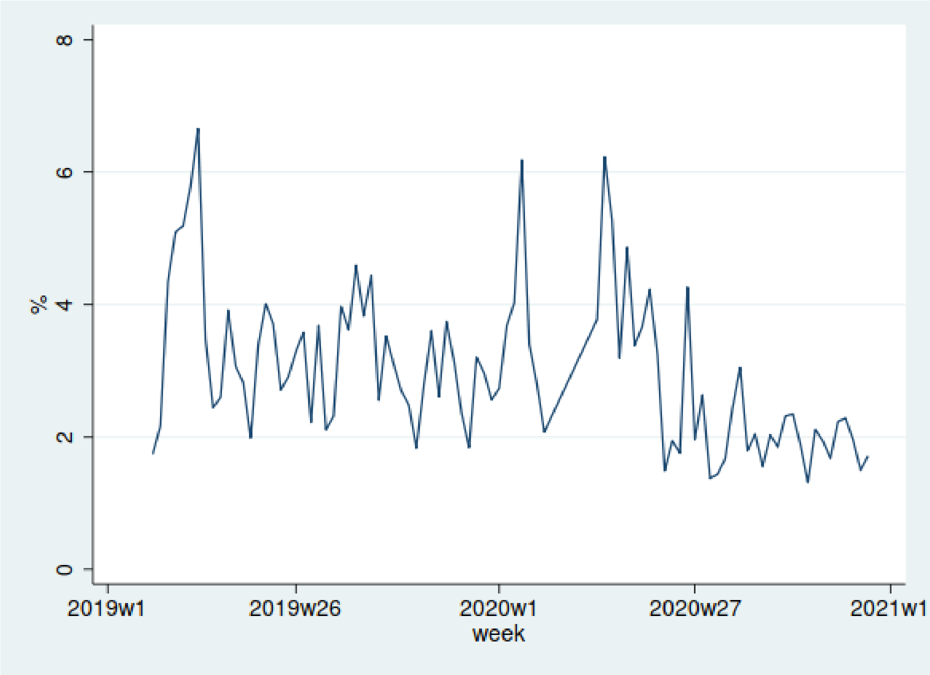
(a)



(b)

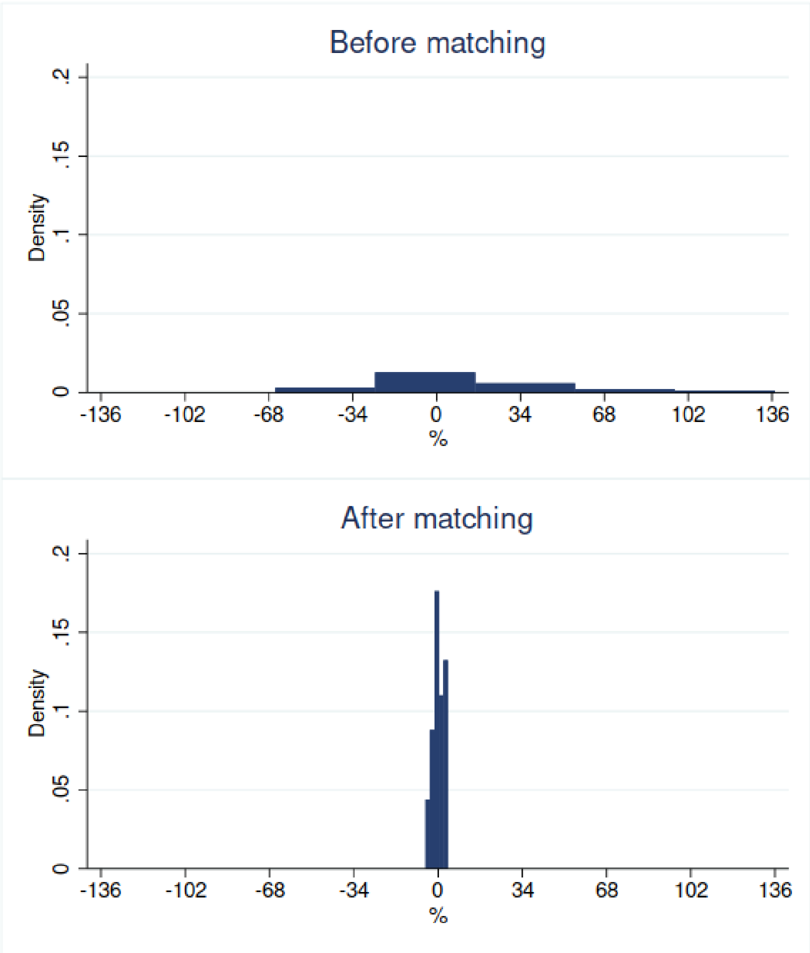
Notes: Figure 2 (a) shows the share of vacancies by job locations having at least 500 job postings. Figure 2 (b) shows the share of vacancies by job locations having less than 500 job postings.

Figure 3.5. Share of dollarised ads over time



Notes: The figure illustrates the proportion of dollarised ads over the sampling period.

Figure 3.6. Standardised bias distribution across covariates before versus after matching



Notes: This figure shows the standardised bias distribution across covariates before and after PSM. The horizontal line presents the standardised bias value (in %), and the vertical line presents the density.

Tables

Table 3.1. List of keywords

Skill	Keywords
Art	Art, artistic
Character	Organized, detail oriented, multitasking, time management, meeting deadlines, energetic
Cognitive	Problem solving, research, analytical, critical thinking, math, statistics
Computer	Computer, spreadsheets, common software (e.g., Microsoft Excel, PowerPoint)
Customer service	Customer, sales, client
Financial	Budgeting, accounting, finance, cost
Language	English, Japanese, Korean, Chinese, foreign language
Project management	Project management
People management	Supervisory, leadership, management (not project), mentoring, staff
Social	Communication, teamwork, collaboration, negotiation, presentation
Software	Programming language or specialized software (e.g., Java, SQL, C++)
Writing	Writing

Notes: The table presents the categorisation of keywords into requirement groups.

Table 3.2. Descriptive statistics

	Wage in VND N=168,158		Wage in USD N=4,948	
	Mean (1)	**SD (2)**	**Mean (3)**	**SD (4)**
Wage (VND)	8,770,899	5,027,481	20,889,000	17,325,000
Education				
High school/Vocational	0.141	0.348	0.017	0.128
Associate degree	0.662	0.473	0.349	0.477
Bachelor's degree	0.193	0.395	0.626	0.484
Master degree	0.001	0.029	0.001	0.035
No education required	0.003	0.051	0.007	0.086
Years of experience				
0-1 year	0.441	0.497	0.123	0.328
1-2 years	0.345	0.475	0.304	0.460
2-5 years	0.190	0.392	0.441	0.497
5-10 years	0.022	0.148	0.126	0.332
10+ years	0.002	0.042	0.007	0.081
Position				
New entry	0.032	0.175	0.012	0.109
Employee	0.840	0.366	0.739	0.439
Medium-level manager	0.119	0.324	0.218	0.413
Top manager	0.009	0.093	0.032	0.175
Location				
Hanoi	0.341	0.474	0.388	0.487
HCM city	0.403	0.491	0.433	0.496
Other locations	0.256	0.598	0.18	0.549
Skill				
Computer	0.324	0.468	0.317	0.465
Software	0.200	0.400	0.492	0.500
Financial	0.142	0.349	0.099	0.298
People Management	0.037	0.190	0.061	0.239
Project Management	0.097	0.296	0.128	0.334
Artistic	0.088	0.283	0.172	0.377
Language	0.196	0.397	0.767	0.423
Character	0.863	0.343	0.625	0.484
Cognitive	0.289	0.453	0.439	0.496
Customer Service	0.371	0.483	0.231	0.421
Social	0.597	0.491	0.536	0.499
Writing	0.023	0.150	0.054	0.226

Notes: The table shows summary statistics of all variables used in the regression. Columns (1) and (3) report the mean of VND and USD jobs subsamples, respectively. Columns (2) and (4) report the standard deviation of VND and USD jobs subsamples, respectively.

Table 3.3. Occupations with the highest level of dollarisation

Panel A: Share of dollarised ads for top 10 job titles	
Job title	Share of dollarised ads (%)
Innovation staff	6.00
Branch manager	6.23
Sales manager	7.93
Network admin	12.68
QA staff	13.39
Stylist	16.67
Marketing manager	20.00
English interpreter	24.22
Developer	29.96
Customer service manager	54.55
Panel B: Top 10 jobs with highest number of dollarised ads	
Job title	Number of dollarised ads
Sales manager	115
Designer	117
Director	120
Accountant	131
Secretary	170
Businessman	188
Engineer	239
Sales	251
Manager	330
Developer	980

Notes: The table shows occupations with the highest share and number of dollarised job postings.

Table 3.4. Correlation matrix

	Education	Experience	Level	Computer	Software	Language	Financial	People	Project	Art	Character	Cognitive	Customer	Social
Experience	0.4649													
Level	0.2337	0.4419												
Computer	-0.0370	0.0216	-0.0219											
Software	0.1521	0.1390	0.0408	-0.0370										
Language	0.2846	0.2730	0.0370	-0.0018	0.2389									
Financial	0.3070	0.1926	0.0752	-0.0387	0.1670	0.1257								
People	0.0623	0.1302	-0.0275	0.0190	0.1872	0.0987	-0.0642							
Project	0.1404	0.1735	0.2166	0.0017	0.0379	0.0581	0.0395	-0.0059						
Art	0.2073	0.2640	0.3036	0.0007	0.1024	0.1149	0.0666	0.0166	0.1792					
Character	0.0899	0.1215	0.0407	0.0314	0.0102	0.2244	0.0176	-0.0499	0.0230	0.0729				
Cognitive	-0.1380	-0.1745	-0.0835	0.0456	0.0134	-0.0999	-0.1319	0.0067	-0.0033	-0.0024	-0.0031			
Customer	0.2577	0.2872	0.1574	-0.0101	0.1650	0.2619	0.1556	0.0224	0.1328	0.2635	0.1919	0.0735		
Social	0.0331	-0.0544	0.0470	0.0389	-0.0077	-0.0829	-0.0441	-0.0763	0.0023	0.0400	-0.0008	0.1649	0.0527	
Writing	-0.0119	-0.0895	0.0424	0.0338	-0.0207	-0.1240	0.0826	-0.0916	0.0386	0.0592	-0.0237	0.2259	0.0561	0.3581

Notes: The table shows the Spearman correlation coefficients for skill requirement variables.

Table 3.5. CEM balance check

Multivariate L1 distance before matching: 0.7761							
Multivariate L1 distance after matching: 0.1572							
Univariate imbalance							
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
	L1	**Mean**	**Min**	**25%**	**50%**	**75%**	**Max**
Education	0	0	0	0	0	0	0
Experience	0	0	0	0	0	0	0
Level	0	0	0	0	0	0	0
Location	0	0	0	0	0	0	0
Quarter	0	0	0	0	0	0	0
Industry	0.0034	0.0134	0	0	0	0	0
Job title	0.0933	-0.2415	0	0	0	0	2
Firm size	0	0	0	0	0	0	0

Notes: The table reports multivariate and univariate imbalance measures for CEM. Column (1) reports the L1 imbalance measure for individual variables. Column (2) reports the difference in the density distributions between treated and control groups at the mean. Columns (3) to (7) report the difference in the density distributions between treated and control groups for the 0^{th} (min), 25^{th}, 50^{th}, 75^{th} and 100^{th} (max) percentiles, respectively.

Table 3.6. Monetary versus non-monetary benefits

	(1) PSM	(2) CEM	(3) Text matching
Exchange rate benefit	0.437***	0.435***	0.296***
	(0.009)	(0.026)	(0.015)
Associate degree	-0.093***	-0.038	0.078**
	(0.033)	(0.117)	(0.035)
Bachelor's degree	-0.006	0.048	0.178***
	(0.033)	(0.120)	(0.038)
Master degree	0.300***		
	(0.088)		
Doctorate			-0.048
			(0.086)
Others	-0.123**		
	(0.056)		
1-2 years	0.175***	0.094***	0.108***
	(0.014)	(0.029)	(0.023)
2-5 years	0.331***	0.231***	0.215***
	(0.014)	(0.031)	(0.025)
5-10 years	0.547***	0.423***	0.326***
	(0.021)	(0.050)	(0.040)
More than 10 years	0.633***	0.557***	0.314***
	(0.065)	(0.095)	(0.081)
Employee	0.512***	1.159***	0.802***
	(0.062)	(0.211)	(0.120)
Medium manager	0.765***	1.522***	1.017***
	(0.064)	(0.214)	(0.124)
Top manager	1.133***	1.688***	1.436***
	(0.078)	(0.250)	(0.155)
Ha Noi	0.078***	0.016	-0.003
	(0.013)	(0.039)	(0.030)
Ho Chi Minh	0.072***	0.053	0.075***
	(0.012)	(0.038)	(0.028)
Computer	-0.051***	-0.028	-0.040**
	(0.010)	(0.019)	(0.019)
Software	0.007	0.071***	0.003
	(0.011)	(0.023)	(0.023)
Language	0.046***	0.031	0.030
	(0.011)	(0.024)	(0.029)
Financial	0.015	-0.016	0.087**
	(0.018)	(0.044)	(0.040)
People management	0.040**	0.046	0.076*
	(0.020)	(0.038)	(0.044)
Project management	-0.004	-0.015	-0.159***
	(0.015)	(0.031)	(0.035)
Art	0.044***	-0.001	0.060*
	(0.014)	(0.024)	(0.031)
Character	-0.043***	-0.047**	-0.027
	(0.009)	(0.020)	(0.019)

Cognitive	-0.004	0.013	-0.027
	(0.010)	(0.019)	(0.022)
Customer service	0.040***	0.005	0.105***
	(0.012)	(0.026)	(0.027)
Social	0.011	0.042**	-0.013
	(0.010)	(0.020)	(0.019)
Writing	-0.059***	-0.108**	-0.044
	(0.019)	(0.043)	(0.059)
Attractiveness	-0.044	-0.066*	0.194***
	(0.031)	(0.037)	(0.063)
Obs	8,419	4,182	3,433
R2	0.547	0.805	0.813

Notes: The table presents the link between wage and exchange rate benefits. Columns (1), (2) and (3) show results for PSM, CEM and text-based matching methods, respectively. Time, industry, job title, and firm size fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table 3.7. Heterogeneous effects

	(1) **Education**	(2) **Experience**	(3) **Position**
Exchange rate benefit (ERB)	0.292***	0.483***	0.787***
	(0.060)	(0.024)	(0.118)
Associate degree	-0.189***	-0.090***	-0.098***
	(0.047)	(0.033)	(0.033)
Bachelor's degree	-0.081*	-0.005	-0.011
	(0.047)	(0.033)	(0.033)
Master degree	0.229*	0.312***	0.296***
	(0.133)	(0.085)	(0.088)
Others	-0.204**	-0.115**	-0.125**
	(0.090)	(0.056)	(0.056)
ERB×Associate degree	0.174***		
	(0.061)		
ERB×Bachelor's degree	0.131**		
	(0.061)		
ERB×Master degree	0.117		
	(0.167)		
ERB×Others	0.144		
	(0.111)		
1-2 years	0.177***	0.180***	0.173***
	(0.014)	(0.019)	(0.014)
2-5 years	0.332***	0.361***	0.329***
	(0.014)	(0.019)	(0.014)
5-10 years	0.548***	0.621***	0.546***
	(0.021)	(0.027)	(0.021)
More than 10 years	0.635***	0.765***	0.632***
	(0.065)	(0.105)	(0.065)
ERB×1-2 years		-0.008	
		(0.029)	
ERB×2-5 years		-0.060**	
		(0.027)	
ERB×5-10 years		-0.146***	
		(0.036)	
ERB×More than 10 years		-0.230*	
		(0.129)	
Employee	0.514***	0.510***	0.662***
	(0.062)	(0.061)	(0.089)
Medium manager	0.766***	0.764***	0.939***
	(0.064)	(0.064)	(0.091)
Top manager	1.134***	1.128***	1.295***
	(0.078)	(0.077)	(0.106)
ERB× Employee			-0.344***
			(0.119)
ERB×Medium manager			-0.392***
			(0.120)
ERB×Top manager			-0.367***
			(0.133)
Obs	8,419	8,419	8,419
R2	0.547	0.548	0.548

Notes: The table presents the heterogeneous wage effect of exchange rate benefits. Columns (1), (2), and (3) show results for different wage effects of exchange rate benefits across education level, experience level, and job position, respectively. The base categories are High school, 0-1 year of experience, Entry level/internship, respectively. Other control variables and fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table 3.8. Skill requirements of dollarised job ads

	(1) Logit Coefficient	(2) Logit Marginal effect	(3) Rare event Logit Coefficient	(4) Rare event Logit Marginal effect
Education				
Associate degree	0.258*	0.004**	0.542***	0.008***
	(0.134)	(0.002)	(0.118)	(0.001)
Bachelor’s degree	0.946***	0.018***	1.255***	0.026***
	(0.137)	(0.002)	(0.119)	(0.002)
Master degree	0.500	0.008	0.267	0.004
	(0.566)	(0.011)	(0.431)	(0.006)
Others	1.075***	0.022***	1.869***	0.051***
	(0.251)	(0.006)	(0.224)	(0.010)
Years of experience				
1-2 years	0.377***	0.006***	0.534***	0.009***
	(0.061)	(0.001)	(0.053)	(0.001)
2-5 years	1.040***	0.020***	1.059***	0.023***
	(0.065)	(0.001)	(0.054)	(0.001)
5-10 years	1.654***	0.040***	1.604***	0.045***
	(0.089)	(0.003)	(0.074)	(0.003)
More than 10 years	0.973***	0.019***	0.734***	0.014***
	(0.225)	(0.006)	(0.206)	(0.005)
Position				
Employee	0.552***	0.010***	0.362***	0.008***
	(0.164)	(0.002)	(0.140)	(0.003)
Medium manager	0.479***	0.008***	0.253*	0.005*
	(0.180)	(0.003)	(0.146)	(0.003)
Top manager	1.133***	0.024***	0.775***	0.019***
	(0.262)	(0.006)	(0.173)	(0.004)
Location				
Ha Noi	0.268***	0.005***	0.372***	0.008***
	(0.053)	(0.001)	(0.045)	(0.001)
Ho Chi Minh	0.372***	0.007***	0.507***	0.011***
	(0.051)	(0.001)	(0.043)	(0.001)
Skill				
Computer	-0.314***	-0.006***	-0.665***	-0.015***
	(0.041)	(0.001)	(0.035)	(0.001)
Software	0.033	0.001	0.752***	0.020***
	(0.047)	(0.001)	(0.034)	(0.001)
Language	1.770***	0.039***	2.149***	0.061***
	(0.043)	(0.001)	(0.037)	(0.001)
Financial	0.116	0.002	-0.390***	-0.008***
	(0.071)	(0.002)	(0.052)	(0.001)
People management	-0.011	-0.000	-0.046	-0.001
	(0.081)	(0.002)	(0.068)	(0.002)
Project management	-0.225***	-0.004***	-0.372***	-0.008***
	(0.058)	(0.001)	(0.049)	(0.001)
Art	0.213***	0.005***	0.290***	0.008***
	(0.057)	(0.001)	(0.045)	(0.001)
Character	-0.576***	-0.013***	-0.597***	-0.016***
	(0.042)	(0.001)	(0.036)	(0.001)
Cognitive	-0.037	-0.001	0.047	0.001
	(0.042)	(0.001)	(0.036)	(0.001)
Customer service	-0.133***	-0.003***	-0.332***	-0.008***
	(0.049)	(0.001)	(0.038)	(0.001)
Social	0.008	0.000	-0.126***	-0.003***
	(0.041)	(0.001)	(0.034)	(0.001)
Writing	0.276***	0.006***	0.176**	0.004**

	(0.086)	(0.002)	(0.072)	(0.002)
Attractiveness	-0.800***	-0.013***	-0.991***	-0.017***
	(0.108)	(0.001)	(0.097)	(0.001)
Obs	163,933	163,933	173,106	173,106
Pseudo R2	0.426			

Notes: The table presents results of Logit and rare event Logit models for job characteristics/requirements of dollarised job ads. Columns (1) and (3) show the estimated coefficients for Logit and rare event Logit models, respectively. Columns (2) and (4) show the average marginal effect for Logit and rare event Logit models, respectively. Dependent variable is a dummy variable that indicates whether a vacancy quotes wages in US dollars. Time, industry, job title, and firm size dummies are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table 3.9. Wage regression for sub-samples of domestically-owned versus foreign-owned firms

	(1) Domestically-owned firms	(2) Foreign-owned firms
Exchange rate benefit	0.449***	0.291***
	(0.013)	(0.025)
Associate degree	0.064	-0.094
	(0.192)	(0.081)
Bachelor's degree	0.130	0.034
	(0.192)	(0.084)
Master degree	0.534**	
	(0.231)	
Others	0.086	0.178
	(0.201)	(0.170)
1-2 years	0.149***	0.222***
	(0.024)	(0.037)
2-5 years	0.320***	0.351***
	(0.025)	(0.038)
5-10 years	0.518***	0.629***
	(0.033)	(0.088)
More than 10 years	0.633***	0.808***
	(0.121)	(0.260)
Employee	0.592***	0.235***
	(0.075)	(0.091)
Medium manager	0.802***	0.502***
	(0.080)	(0.100)
Top manager	1.166***	1.035***
	(0.095)	(0.213)
Ha Noi	0.071***	0.079*
	(0.025)	(0.046)
Ho Chi Minh	0.047*	0.030
	(0.025)	(0.034)
Computer	-0.028**	-0.102***
	(0.014)	(0.029)
Software	-0.006	0.058
	(0.018)	(0.036)
Language	0.054***	0.037
	(0.015)	(0.038)
Financial	0.010	0.127**
	(0.027)	(0.062)
People management	0.048*	0.004
	(0.029)	(0.075)
Project management	-0.037**	-0.061
	(0.018)	(0.085)
Art	0.016	-0.062
	(0.018)	(0.054)
Character	-0.042***	0.011
	(0.016)	(0.030)
Cognitive	0.016	0.036
	(0.015)	(0.042)
Customer service	0.045**	-0.056
	(0.018)	(0.035)
Social	0.001	0.004
	(0.015)	(0.028)
Writing	-0.021	-0.312***
	(0.027)	(0.075)
Attractiveness	-0.118***	-0.105*

	(1)	(2)
	(0.044)	(0.059)
Obs	3,969	939
R2	0.565	0.615

Notes: The table presents the link between wage and exchange rate benefits for subsamples of domestically-owned and foreign-owned firms. Columns (1) and (2) show results for domestically-owned and foreign-owned firms, respectively. Time, industry, job title, and firm size fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Online Appendix

Table A3.1. List of job titles/occupations

Accountant	Database administrator	Marketing	R&D manager
Accounting advisor	Deliveryman	Marketing director	R&D staff
Accounting manager	Deputy	Marketing manager	Real estate businessman
Actor	Deputy manager	Mathematics teacher	Real estate consultant
Admin	Designer	MC	Receptionist
Admin manager	Developer	Mechanics	Restaurant manager
Animator	Digital marketing instructor	Merchandiser	Sale assistant
Appraiser	Director	Network admin	Sales
Architect	Doctor	EHS staff	Sales consultant
Art teacher	Driver	Nurse	Sales manager
AutoCAD senior designer	Editor	Nursery teacher	Sap consultant
Banker	Engineer	Nutritionist	School vice president
Bartender	English interpreter	Office supporter	Scrum master
Beauty teacher	English teacher	Officer	Secretary
Beauty technician	Equipment deployment staff	Online supporter	Security
Bodyguard	Event organiser	Operator	Senior developer
Branch manager	Financial advisor	Others	Senior merchandiser
Bridge system engineer	Financial investment	Partnership manager	Shop manager
Broker	Ga staff	Partnership staff	Spa advisor
Business analyst	Gymnastics teacher	Personal trainer	Stylist
Business development	Health insurance staff	Pharmacist	Supply chain staff
Business representative	Helpdesk	Photographer	Support staff
Business support staff	Housing resource staff	Photoshop	Swimming pool lifeguard
Businessman	HR director	Planning staff	Teaching assistant
CAD/CAM	HR manager	Planning supervisor	Technical consultant
CAD/CAM senior	HR staff	PR	Technician
Cashier	Import-export staff	PR executive	Technology staff
Chef	Information security	Primary school teacher	Tender
Chess game instructor	Innovation staff	Procedure controller	Tester
Chinese interpreter	Internal auditor	Producer	Tour guide

Cleaner	IT staff	Product advisor	Tour leader
Clearance specialist	IT system administrator	Product specialist	Tour operator
Compliance officer	IT teacher	Product/content reviewer	Training specialist
Comtor	Japanese language staff	Production management	Vietnamese teacher
Content creator	Japanese teacher	Production management	Visa/study abroad advisor
Controller/supervisor	Japanese translator	Production statistics staff	Waiter
Cook	Korean interpreter	Project development staff	Warehouse staff
Costing staff	Legal experts	Project manager	Web developer
Customer service director	Librarian	Purchasing staff	Web management
Customer service manager	Live streamer	QA manager	Worker
Customer service staff	Logistics staff	QA staff	Yoga instructor
Customer support	Manager	QC staff	
Data analyst	Market consultant	QC team leader	
Data scientist	Market research analyst	Quality inspector	

Notes: The table presents the job titles/occupations in our data.

Table A3.2. Definitions of main variables

Variable	Definition
High school	A dummy variable that takes the value of 1 if the job ad requires a high school qualification, and 0 otherwise.
Associate degree	A dummy variable that takes the value of 1 if the job ad requires an associate degree qualification, and 0 otherwise.
Bachelor's degree	A dummy variable that takes the value of 1 if the job ad requires a bachelor's degree qualification, and 0 otherwise.
Other educations	A dummy variable that takes the value of 1 if the job ad requires other educational qualifications, and 0 otherwise.
0-1 year	A dummy variable that takes the value of 1 if the job ad requires less than one year of experience, and 0 otherwise.
1-2 years	A dummy variable that takes the value of 1 if the job ad requires 1-2 years of experience, and 0 otherwise.
2-5 years	A dummy variable that takes the value of 1 if the job ad requires 2-5 years of experience, and 0 otherwise.
5-10 years	A dummy variable that takes the value of 1 if the job ad requires 5-10 years of experience, and 0 otherwise.
10+ years	A dummy variable that takes the value of 1 if the job ad requires more than 10 years of experience, and 0 otherwise.
Entry level/Internship	A dummy variable that takes the value of 1 if the job ad is at an entry level/internship position, and 0 otherwise.
Non-manager	A dummy variable that takes the value of 1 if the job ad is at a non-managerial position, and 0 otherwise.
Medium manager	A dummy variable that takes the value of 1 if the job ad is at a medium-manager position, and 0 otherwise.
Top manager	A dummy variable that takes the value of 1 if the job ad is at a top-manager position, and 0 otherwise.
Hanoi	A dummy variable that takes the value of 1 if the job ad locates in Hanoi, and 0 otherwise.
HCM	A dummy variable that takes the value of 1 if the job ad locates in HCM city, and 0 otherwise.
Other locations	A dummy variable that takes the value of 1 if the job ad locates in other cities/provinces, and 0 otherwise.
Computer	A dummy variable that takes the value of 1 if the job ad requires computer skill, and 0 otherwise.
Software	A dummy variable that takes the value of 1 if the job ad requires software skill, and 0 otherwise.
Language	A dummy variable that takes the value of 1 if the job ad requires language skill, and 0 otherwise.
Financial	A dummy variable that takes the value of 1 if the job ad requires financial skill, and 0 otherwise.
People management	A dummy variable that takes the value of 1 if the job ad requires people management skill, and 0 otherwise.
Project management	A dummy variable that takes the value of 1 if the job ad requires project management skill, and 0 otherwise.
Art	A dummy variable that takes the value of 1 if the job ad requires artistic skill, and 0 otherwise.
Character	A dummy variable that takes the value of 1 if the job ad requires character skill, and 0 otherwise.
Cognitive	A dummy variable that takes the value of 1 if the job ad requires cognitive skill, and 0 otherwise.

Customer service	A dummy variable that takes the value of 1 if the job ad requires customer service skill, and 0 otherwise.
Social	A dummy variable that takes the value of 1 if the job ad requires social skill, and 0 otherwise.
Writing	A dummy variable that takes the value of 1 if the job ad requires writing skill, and 0 otherwise.

Notes: The table presents definitions of the main variables used in our analysis.

Table A3.3. Robustness test for main regression using average salary measure

	(1) PSM	(2) CEM	(3) Text matching
Exchange rate benefit	0.452***	0.422***	0.315***
	(0.008)	(0.026)	(0.014)
Associate degree	-0.072**	-0.000	0.084**
	(0.032)	(0.111)	(0.033)
Bachelor's degree	0.028	0.087	0.187***
	(0.033)	(0.113)	(0.036)
Master degree	0.274**		
	(0.111)		
Doctorate			-0.138*
			(0.078)
Others	-0.103*		
	(0.054)		
1-2 years	0.157***	0.097***	0.115***
	(0.018)	(0.025)	(0.021)
2-5 years	0.277***	0.205***	0.226***
	(0.018)	(0.028)	(0.023)
5-10 years	0.498***	0.347***	0.352***
	(0.025)	(0.051)	(0.037)
More than 10 years	0.541***	0.530***	0.408***
	(0.065)	(0.106)	(0.079)
Employee	0.524***	1.088***	0.812***
	(0.062)	(0.195)	(0.121)
Medium manager	0.780***	1.452***	1.027***
	(0.064)	(0.199)	(0.124)
Top manager	1.153***	1.725***	1.398***
	(0.076)	(0.225)	(0.154)
Ha Noi	0.054***	0.019	-0.016
	(0.016)	(0.035)	(0.026)
Ho Chi Minh	0.065***	0.040	0.049*
	(0.014)	(0.035)	(0.026)
Computer	-0.052***	-0.038**	-0.048***
	(0.010)	(0.018)	(0.018)
Software	0.024**	0.083***	0.008
	(0.011)	(0.023)	(0.021)
Language	0.030***	0.038*	0.034
	(0.011)	(0.021)	(0.026)
Financial	0.003	-0.006	0.096***
	(0.017)	(0.042)	(0.035)
People management	0.070***	0.126***	0.076*
	(0.020)	(0.043)	(0.043)
Project management	0.004	0.030	-0.151***
	(0.014)	(0.028)	(0.036)
Art	0.061***	0.010	0.081***
	(0.016)	(0.023)	(0.029)
Character	-0.029***	-0.053***	-0.025
	(0.010)	(0.018)	(0.017)

Cognitive	0.017	0.040**	-0.001
	(0.011)	(0.018)	(0.021)
Customer service	0.026**	0.018	0.085***
	(0.012)	(0.025)	(0.024)
Social	0.004	0.026	-0.008
	(0.009)	(0.019)	(0.017)
Writing	-0.049***	-0.122***	-0.081
	(0.019)	(0.040)	(0.050)
Attractiveness	-0.055*	-0.052	0.182***
	(0.032)	(0.035)	(0.054)
Obs	8,419	4,182	3,433
R2	0.560	0.829	0.810

Notes: The table presents the regression results of the wage equation using the average salary measure. Columns (1), (2) and (3) show results for PSM, CEM and text-matched samples, respectively. Time, industry, job title, and firm size fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table A3.4. Robustness test for main regression using maximum salary measure

	(1) **PSM**	(2) **CEM**	(3) **Text matching**
Exchange rate benefit	0.451***	0.408***	0.327***
	(0.009)	(0.028)	(0.015)
Associate degree	-0.057	0.047	0.085**
	(0.036)	(0.116)	(0.037)
Bachelor's degree	0.049	0.141	0.189***
	(0.036)	(0.118)	(0.040)
Master degree	0.286**		
	(0.126)		
Doctorate			-0.203**
			(0.090)
Others	-0.090		
	(0.060)		
1-2 years	0.158***	0.108***	0.126***
	(0.020)	(0.027)	(0.023)
2-5 years	0.261***	0.199***	0.244***
	(0.020)	(0.030)	(0.025)
5-10 years	0.487***	0.333***	0.376***
	(0.027)	(0.054)	(0.040)
More than 10 years	0.502***	0.535***	0.479***
	(0.069)	(0.120)	(0.086)
Employee	0.525***	1.051***	0.820***
	(0.066)	(0.200)	(0.128)
Medium manager	0.779***	1.411***	1.035***
	(0.067)	(0.204)	(0.131)
Top manager	1.154***	1.675***	1.380***
	(0.079)	(0.238)	(0.160)
Ha Noi	0.051***	0.035	-0.020
	(0.017)	(0.037)	(0.028)
Ho Chi Minh	0.066***	0.043	0.031
	(0.016)	(0.038)	(0.029)
Computer	-0.050***	-0.041**	-0.052***
	(0.010)	(0.019)	(0.018)
Software	0.033***	0.097***	0.011
	(0.012)	(0.025)	(0.023)
Language	0.017	0.046**	0.039
	(0.012)	(0.022)	(0.027)
Financial	-0.002	-0.000	0.102***
	(0.018)	(0.044)	(0.034)
People management	0.084***	0.143***	0.079*
	(0.022)	(0.047)	(0.047)
Project management	0.009	0.041	-0.140***
	(0.015)	(0.029)	(0.039)
Art	0.060***	0.013	0.094***
	(0.017)	(0.025)	(0.031)
Character	-0.024**	-0.052***	-0.022
	(0.011)	(0.019)	(0.019)

Cognitive	0.031***	0.044**	0.010
	(0.012)	(0.019)	(0.022)
Customer service	0.016	0.018	0.077***
	(0.013)	(0.026)	(0.025)
Social	0.006	0.023	-0.002
	(0.010)	(0.020)	(0.018)
Writing	-0.038*	-0.106**	-0.097*
	(0.021)	(0.043)	(0.050)
Attractiveness	-0.054	-0.033	0.181***
	(0.035)	(0.040)	(0.059)
Obs	8,419	4,182	3,433
R2	0.522	0.819	0.797

Notes: The table presents the regression results of the wage equation using the maximum salary measure. Columns (1), (2) and (3) show results for PSM, CEM and text-matched samples, respectively. Time, industry, job title, and firm size fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Chapter 4. Formal versus Informal Employment: Evidence from Online Job Vacancy Data

4.1. Introduction

The informal job sector, also known by other names such as hidden or grey sector, while many definitions exist, is generally defined as a sector of employment that does not meet the requirements of national legislation and regulations. Informal employment accounts for a large proportion of the total labour force, especially in developing and transitional economies (Karabchuk and Zabirova, 2018). The informal sector is traditionally considered as a labour market segment that usually employs disadvantaged workers, including less educated and lower-skilled ones, ethnic minorities or immigrants (Mazumdar, 1983). As a result, informal employment might be a source of low productivity, low economic growth and reduced government tax revenue (Lackó, 2000). Additionally, workers in this sector are less likely to be covered by the social securities, social benefits and other protections provided by the labour contract. It is further shown that informality adversely affects income inequality, job satisfaction and the subjective well-being of the workers (De Graaf-Zijl, 2012; Lukiyanova, 2013; Karabchuk and Soboleva, 2014). Hence, identifying and understanding the informal employment as well as the wage gap between the formal and informal sectors are crucial questions that have long been pursued by economic researchers.

Yet, in the literature, there is mixed evidence on the wage gap between formal and informal jobs. On the one hand, some studies show that the formal sector pays higher wages than the informal sector (Carneiro and Henley, 2001; Lehmann and Zaiceva, 2013). For example, Lehmann and Zaiceva (2013) show a wage penalty in the lower part of the distribution and no statistically significant difference between informal and formal wages in the upper half of the distribution for salaried workers in Russia. On the other hand, other studies provide evidence of the wage premium for the informal employment sector (Marcouiller et al., 1997, Staneva and Arabsheibani, 2014). For instance, after controlling for observable characteristics of formal and informal employees, Staneva and Arabsheibani (2014) still find a wage premium of up to 14.4 percent in favour of informal job workers in Tajikistan.

Evidence in existing studies mainly relies on survey data, which might lack detailed and important job dimensions, such as narrowly-defined occupations and the skills used at work. In this study, we add new evidence to the debate of the formal and informal sector wage differentials. Our contribution to this literature is shedding new light on the wage gap between formal and informal employment by using a more frequent and granular vacancy-level dataset, which can provide more detailed job characteristics and narrowly-defined occupations.

Our data contain online job postings, which are scraped automatically from one of the largest job boards in Vietnam from February 2019 to December 2020. There are more than 170,000 job ads across over 170 different job titles of both high-skilled and low-skilled job segments. Each vacancy provides detailed job characteristics and requirements such as the job title, required education level, experience level, job position, job location, industry and advertised date. Thus, this wealthy and well-structured dataset allows us to match vacancies based on various job dimensions and to explore the wage gaps across different job types.

To examine the wage gap between formal and informal employments, we rely on various matching techniques, including propensity score matching (PSM), coarsened exact matching (CEM) and text-based matching with the Jaccard similarity score. Our estimations show significant wage premia of 1.6 to 1.9 percent for formal jobs and wage penalties of 0.8 to 21.4 percent for informal ones. These findings suggest that informal jobs – regardless of the definition/measure used, pay on an average lower wage than formal jobs.

In addition, we find that the wage gap between formal and informal jobs is substantially heterogeneous. Particularly, workers in formal jobs can be even more advantaged in jobs requiring a higher level of education. That is, the formal-informal employment wage gap tends to be the smallest among low-education jobs and is widened as the required education increases. Although the wage gap is considerably heterogeneous across jobs of different required work experience and job positions, there are no clear trends for such heterogeneities.

This paper contributes to two additional strands of literature. First, it complements the controversial literature on how to define informality in the labour market. More specifically, extant literature uses three approaches to identify informal jobs, including the enterprise-based, jobs-based or activity-based definitions (see, e.g., Henley et al., 2009; Khamis, 2012; Adair, 2020). Each measure has its own pros and cons and mostly relies on information obtained from survey data of firms or workers. In this paper, our contribution is proposing four distinct and complementary definitions for formal and informal jobs among salaried employment segments, based on a textual analysis of the vacancies advertised by firms. In particular, we classify a vacancy as formal if it explicitly advertises social benefits; and as informal if it contains discriminatory content against the candidate’s (i) marital status, (ii) gender and (iii) disability. These measures enable us to discover new dimensions of job informality, which cannot be easily observed using previous data.

Furthermore, we contribute to the growing literature that employs online job vacancy data to investigate the labour market dynamics. In particular, online vacancy data have been extensively used to examine the dynamics of skill demand, such as AI and social skills (see, e.g., Deming and Kahn, 2018; Babina et al., 2021; Acemoglu et al., 2022). Other studies employ online job postings data to investigate the wage-setting behaviours such as hidden wage posting or the spillover of changes in wage-setting at major employers to other employers (see e.g., Brenčič and Norris, 2012; Faberman and Kudlyak, 2019; Marinescu and Wolthoff, 2020). In addition, other usages of this data source include measuring labour market indicators such as the labour market tightness or labour market concentration (see e.g., Modestino et al., 2016; Acemoglu et al., 2019; Javorcik et al., 2019; Grinis et al., 2019); exploration of regional/occupational labour markets such as the Beveridge curve or Phillips curve (see e.g., Morales et al., 2021; Faryna et al., 2022). To the best of our knowledge, we are the first to use online job vacancy data to examine informal employment in the labour market.

The rest of this paper is organised as follows. In the next section, Section 4.2 introduces definitions of informal employment and informality in the context of Vietnam. In section 4.3, we describe the dataset employed in our paper. Section 4.4 presents our empirical strategies. Section 4.5 discusses empirical results. Section 4.6 gives the covariate balance checks for our matching methods. Finally, Section 4.7 concludes and discusses implications.

4.2. Defining informality

4.2.1. Definitions

Although the terminology has been extensively used, there is no consensus on how to define and measure "informality" (Staneva and Arabsheibani, 2014). For a few decades, the concept and definition of informal jobs have been proposed and developed from various viewpoints since this concept was first introduced in the pioneering study of Hart (1973) in urban areas of Southern Ghana nearly five decades ago. In general, there are three different approaches to define informal employment, including the enterprise-based, jobs-based or activity-based definitions.

Starting with the firm-based approach, the 15th International Conference of Labour Statisticians in January 1993 adopted a Resolution that define informality based on the type of firm and its legal status. In particular, the Resolution defined informal firms on the basis of some criteria, such as private unincorporated enterprises, enterprises that involve with the production of domestic or personal services by employing domestic employees, enterprises that have a number of employees below a certain threshold (i.e., usually fewer than five to ten workers depending on the studied context). Among those, one of the most commonly used measures to identify an informal firm is firm size. For instance, Khamis (2012) and Lehmann and Zaiceva (2013) employ a threshold of ten or fewer employees for the US and Russia, respectively; Henley et al. (2009), Benjamin and Mbaye (2012) and Cuadros-Meñaca (2020) use the threshold of fewer than five employees for an analysis of Brazil, three West African countries (i.e., Benin, Burkina Faso and Senegal) and Colombia, respectively.

Since the existing firm-based definition may not capture all cases of informal employment (e.g., informal employment in formal enterprises or private households), the 17th International Conference of Labour Statisticians 2003 complemented the firm-focused definition of informal jobs with a job-based concept to further include employment outside informal firms. In particular, according to the job-focused approach, informal jobs refer to those that are not subject to labour regulations, social security regulations, income taxation, or those that lack basic social protections, legal protections or employment benefits (e.g., overtime compensation, paid annual, sick leave, severance pay or payment in case of dismissal).

This job-focused approach takes into account the case where formal firms choose to employ workers informally in order to avoid regulations, taxation, payment of payroll taxes, employer's contributions to social security or pensions. This definition is extensively adopted in the literature. For instance, some of the measures used in Henley et al. (2009) to identify informal employment in Brazil include the availability of a labour card and the payment of social security contributions. Gallagher et al. (2013) identify informal jobs based on the availability of a labour contract and the payment of social security contributions in the context of China.

While the above-mentioned firm-based and job-based definitions have been largely used in the literature on informal employment in developing countries, an activity-based definition has dominated in the literature on developed and post-Soviet economies, where informal employment refers to the underground, cash-in-hand or undeclared employment (Williams and Windebank, 1998). According to this approach, informal employment covers a broader context of non-observed economic activities, including any paid activities that are legal but not declared to the authority to evade tax payment (e.g., income tax, value added tax), payment of social security contributions or to avoid legal law and standards (e.g., the minimum wage, maximum working hours or safety and health standards). This definition has been extensively adopted in previous studies, such as Adair (2021) in Poland, Williams and Kayaoglu (2020) in 27 European Union nations and the UK, and Williams and Round (2008) in Ukraine.

4.2.2. Informal employment in Vietnam

In the present study, adopting the job-based approach, we use one definition to identify formal jobs and three different definitions for informal ones. More specifically, following previous literature (see, e.g., Henley et al., 2009; Khamis, 2012; Gallagher et al., 2013), we identify formal and informal jobs based on compliance with the legal framework and the labour law. With online job vacancy data, we can observe the compliance with the two major labour laws: one on employees' benefit and the other on discrimination at work. The reason we choose several proxies for informal jobs in our analysis is that each approach on its own has statistical and conceptual limitations as a definition/measure for informal employment, but if considered together, they can provide a robust approximation. However, our definitions of informal jobs have a limitation. Indeed, basing on the compliance of the company with labour law via online job posting, our definitions of informal jobs only indirectly and partially reflect the informality.

First, according to the 2015 Law on social insurance (number 58/2014/QH13)[18], it is required that both employers and employees participate in compulsory social insurance. The social insurance covers cases such as the employee's sickness, maternity, labour accident, occupational disease, retirement or death on the basis of his/her contributions to the social insurance fund. Thus, we identify a job as formal if the employer explicitly mentions their payment and contribution to compulsory social insurance in the job description text.

Second, according to the Prohibited act 1 in Article 8 of the 2012 Vietnam's Labour code, employers are prohibited from performing discrimination at work based on the workers' disabilities, gender, marital status, race, skin colour, social class, religion/belief, HIV status or for the reason of establishing, joining the trade union and taking part in trade union's activities by the workers. We can observe the discriminatory content based on workers' gender, marital status and disabilities in our sample of online job vacancies. Hence, we define a job as informal if it shows discrimination against the candidate's (i) marital status (i.e., either married or single), (ii) gender and (iii) disability.

Figure 4.1 reports the proportion of vacancies classified as formal or informal under the four definitions. The statistics indicate that the percentage of job postings classified as formal under

[18] https://www.ilo.org/dyn/natlex/natlex4.detail?p_lang=en&p_isn=99775&p_country=VNM&p_count=3, accessed on 31st July 2022.

the first definition based on social insurance coverage is 84.98%, thus, 15.02% of job postings are classified as informal. Around 0.24% of the job ads are identified as informal under the second definition on the basis of having discriminatory content towards workers' marital status. Based on the third definition concerning discrimination towards workers' gender, 27.34% of the job vacancies are considered as informal jobs. Finally, under the fourth definition concerning discrimination toward workers' disability, the proportion of job postings classified as informal is 2.37%.

[Insert Figure 4.1 here]

Overall, the second definition based on marital status and the third definition based on gender discriminatory content provides the smallest (i.e., 0.24%) and largest percentage (i.e., 27.34%) of informal jobs, respectively. Lehman and Zaiceva (2013) find that the incidence of informal employment varies widely across definitions. They show that about 25% of the workforce can be classified as informal by all three measures, and around 40% of workers are classified as informal on the basis of having no signed labour contract and no social security coverage. Kapeliushnikov (2012) also points out that depending on different definitions, the proportion of informal employment can vary between 10% to 25% and that the characteristics of informal jobs may differ according to different definitions.

Overall, these statistics reveal that using four measures of formal and informal jobs, our analysis reports a slightly smaller percentage of informal jobs among the salaried job segment than what has been documented in previous studies. For instance, the recent study by Vu and Rammohan (2022) reveals that the percentage of informal non-farm wage employment accounts for 28.78% of total households in Vietnam in 2014. Likewise, according to the report of GSO and ILO (2018), 47.4% of salaried workers are classified as informally employed in 2016. The reason is that our dataset, as discussed in the next section, represents an above medium-skilled salaried jobs in urban areas with a slightly higher average wage level and higher education and experience level. As a result, the percentage of informal employment in our study is lower than that in previous research.

4.3. Data

4.3.1. Data collection and process

The data on which our analysis is based contain online job ads collected weekly from a top job board in Vietnam over the 22-month period from February 2019 to December 2020. For each job posting, we can extract detailed job characteristics and information, namely the job title, category, job level, job type, work location, job description, preferred gender, education level, experience level, job requirement, offered monthly salary, the firm's number of employees.

For data processing, we perform the following steps. First, we discard job postings without salary information. If the wage is quoted in US dollars, we convert it to the equivalent value in local currency – VND using the daily exchange rate on the day when the vacancy is first posted in the job site[19]. For vacancies that quote a salary range instead of a specific number, the lower bound is chosen[20]. Additionally, our analysis only considers full-time jobs and those with a salary higher or equal to the obligatory minimum wage[21]. Finally, we remove duplicated vacancies using the identification number assigned to each job ad, which is explicitly stated in the job ad text.

We take additional steps to extract job requirements and characteristics from the vacancy text. First, we clean the job title text of each vacancy and then classify them into 117 different narrowly-defined job titles. We also process the job industry text to achieve a set of 43 industries. Additionally, we define firm size based on the number of employees. With regards to job location, we extract information on the city/province where the job locates and then group them into three locations, namely Hanoi, HCM city and other cities/provinces.

In the next step, we extract requested skills and other job characteristics based on the words written in the job requirement text. In particular, using Natural Language Processing techniques, we extract all unique keywords from the unstructured text of job requirements. From this set of keywords, those that signal skills are selected and classified into twelve different skill groups. We use a similar method as Deming and Kahn (2017) in classifying skills as cognitive, social, character, writing, customer service, project management, people management, financial, computer (general) and software (specific) skills. We further add foreign language and artistic skills as those two skills are common and unique to the Vietnamese labour market. Basically, a job vacancy is classified as requiring a skill group if it contains at least one keyword listed in such skill group. Examples of keywords for twelve skill groups are illustrated in Table 4.1.

[19] USD - VND exchange rate data were downloaded from the website https://www.investing.com, accessible on 8th June 2022.

[20] Results for wage equations (reported in Appendix Table A4.3 to A4.8) remain robust when we choose the average wage level and maximum wage level instead.

[21] The minimum wage is set at VND 2,920,000 according to the Decree number 157/2018/ND-CP (effective since the 1st of January 2019).

The same text scanning approach is used to obtain the formal and informal job indicators. More specifically, for the formality indicator, we classify a job posting as formal if the job description text contains one or more of the following keywords: social insurance, maternity benefit, insurance coverage, etc. For the first informality measure, a vacancy is classified as informal if the skill requirement text contains one or more keywords such as married, single, not married, etc. For the third informality measure, a job ad is identified as informal if it contains one or more keywords such as not disabled, not deaf, no disabilities, etc. For the second informal job indicator, if the employer specifies a specific preferred gender (either male or female) in the gender section, the job posting is classified as informal. Table 4.1 contains the list of example keywords for each formal/informal job measure or skill category.

[Insert Table 4.1 here]

<u>4.3.2. Descriptive statistics</u>

Figure 4.2 illustrates the top ten job titles with the highest percentage of formal and informal vacancies for each measure. Regarding the first formal definition, the frequency of formal vacancies dominates among partnership managers, gymnastics teachers, and health insurance staff, with almost all vacancies can be classified as formal.

Regarding the first informal definition, bodyguard dominates in terms of the frequency of having informal vacancies with nearly 25% of job vacancies. This is followed by personal trainer and sales assistant. With regards to the second informality definition, bodyguard also dominates in terms of the frequency of having informal vacancies together with equipment deployment staff, producer, gymnastics teacher, IT teacher and chess game instructor. Finally, as for the third informal definition, chess game instructor comes top with the highest percentage of informal employment, followed by supply chain staff and personal trainer.

In general, this suggests that some occupations are more likely to have a higher proportion of informal employment regardless of the informality definition. Overall, informal employment exists among a diversity of occupations of both skilled and low-skilled job segments.

[Insert Figure 4.2 here]

Figure 4.3 plots the top ten industries with the highest proportion of formal and informal vacancies for each formality/informality measure. Overall, we can see that informal employment exists in various job industries. In particular, with respect to the first formality definition, industries with the highest percentage of formal vacancies is Interpreter/Translator with the percentage of 35.1% and followed by the Legal industry with 22.1%.

As regards the first informality measure, the top industry is Sales, with 0.7% of informal vacancies. Next came Accounting/Audit and Security industries at 0.3%. Regarding the second informality definition, Security and Maintenance dominate other industries in terms of having informal vacancies with 89% and 75.7%, respectively. As for the last informality measure, the industry with the highest percentage of informal job postings is Telecommunications at 13.3%, followed by Customer Service at 6.4%.

[Insert Figure 4.3 here]

Figure 4.4 shows the percentage of formal and informal vacancies for each formality/informality measure by different firm sizes. With regards to the first formality definition, formal vacancies are most likely to be posted by large firms of 10,000 to 19,999 employees at 23.7%. Next came medium firms of 500 to 999 and 100 to 499 employees at 18.3% and 17.7%, respectively.

As for the first informality measure, the top firm size is 500 to 999 with 0.5% of informal vacancies, followed closely by smaller firms of 100 to 499 employees at 4%. With respect to the second informality definition, very large firms of 5,000 to 9,999 workers dominate other firm sizes in terms of the proportion of informal vacancies at a significant percentage of 38.5%. Interestingly, using this informality measure, micro firms with ten or fewer employees also have a high proportion of informal employment at 31.5%. As regards the third informality measure, surprisingly largest firms with more than 50,000 employees have the highest fraction of informal employment at 8.2%, followed by large firms with 1,000 to 4,999 and 5,000 to 9,999 employees at 4.9% and 4.8%, respectively. Generally, this suggests that informal jobs appear not only among micro to small firms but also among large firms to very large firms of more than 1,000 employees.

[Insert Figure 4.4 here]

The correlations between measures of formal/informal employment and skill requirements are reported in Table 4.2. As can be seen from the first four columns and three rows of the table, there are very weak correlations between the four measures. This suggests that different indicators of informality are capturing different dimensions of informal jobs, making our analysis based on multiple informality measures more comprehensive.

The correlations between formal/informal measures and skill groups show a mixed picture. For instance, the first informal measure positively correlates with soft skills such as general computer, characteristics and social skills and negatively correlates with other hard skills. Whereas the second informal measure positively correlates with the hard skill financial skill and the third measure positively correlates with the hard skill cognitive. These statistics again point out that informal employment does not necessarily present only in the low-skilled job segment but also in the skilled job segment.

With respect to correlations between skill requirements, the skill groups such as cognitive, writing, people management and project management skills are positively correlated with each other, with education and years of experience and with most of other skills. These skills are general skill groups which are highly frequently requested by a wide range of occupations.

[Insert Table 4.2 here]

Table 4.3 shows the summary statistics of the main variables used in our analysis. First, the average wage in our sample is slightly above VND 9 million, which is 15.83% higher than the average salary of VND 7,77 million in 2019 reported by the 2020 firm survey[22]. In terms of required education level, most job postings require an Associate degree at 65.29% and just above one-fifth of the job ads require a Bachelor's degree. As for requested work experience, nearly half of the job ads (i.e., 43.20%) accept early careers with less than one year of experience. And the more years of experience required, the fewer the job postings.

Regarding job position, the majority of vacancies (i.e., 83.75%) is for employee level, and only a tiny fraction (i.e., 12.21%) is for medium-level manager position such as senior positions, team leaders, etc. Among skill requirements, character, social, customer service and general computer skills have the highest mean values, suggesting that these skills are the general skills that most occupations look for. In contrast, specific skills such as writing, people management, artistic and project management are the least needed skills in our vacancies sample. Overall, we draw a conclusion that our vacancies dataset, to some extent, tends to represent an above medium-skilled salaried jobs in urban areas with a slightly higher average wage level and higher education and experience level.

[Insert Table 4.3 here]

[22] https://laodong.vn/cong-doan/tien-luong-binh-quan-cua-nguoi-lao-dong-trong-nam-qua-997894.ldo, accessed 31st July 2022.

4.4. Empirical strategy

4.4.1. Characteristics of formal and informal jobs

To determine the characteristics and skill requirements featured in formal and informal vacancies, we estimate the following Probit model:

$$P(Formal/Informal_{it} = 1|X) = \Phi(\beta_0 + Skill_{it}\beta_1 + Controls_{it}\beta_2 + Quarter\ dummies + Industry\ dummies + Occupation\ dummies\ + Firm\ size\ dummies + \varepsilon_{it})\ (4.1)$$

where i and t refer to vacancy i posted at time t. Φ is the cumulative standard normal distribution function (i.e., Φ(z)=P(Z≤z)) and Z follows a standard normal distribution N (0,1). *Formal or Informal* is a dummy variable that equals one if the vacancy is classified as formal or informal according to our formality/informality measures and zero otherwise. *Skill* is a vector containing 12 skill dummy variables for cognitive, social, character, writing, customer service, project management, people management, financial, computer, software, language and artistic skills. Furthermore, we include a set of dummy variables to control for the required experience level, education level, work location and job position. Finally, quarter dummies, occupation dummies, industry dummies and firm size dummies are also added to our model. Detailed definitions of the variables used in the analysis are reported in Online Appendix Table A4.2.

4.4.2. The formal-informal wage gaps

In this section, we estimate the wage gaps between these two sectors of employment. Having shown that formal and informal jobs are different in terms of various job characteristics and skill requirements, we now rely on matching techniques to achieve a more balanced and homogeneous job vacancy sample. The reason for using matching techniques is that higher-skilled applicants may self-select into formal jobs and vice versa, lower-skilled applicants may self-select into informal ones, we need to control for any possible selectivity bias.

4.4.2.1. Propensity Score Matching

First, we use the commonly used method – Propensity Score Matching (PSM) (Rosenbaum and Rubin, 1983), to match formal/informal jobs (i.e., treated units) with informal/formal jobs (i.e., control units) of similar characteristics. In the first step, the probability of a job ad being classified as formal/informal given observed job characteristics is estimated by the Probit model (1)[23]. This probability, referred to as the propensity score, is then used to match formal vacancies with informal vacancies. More specifically, since we perform the nearest neighbour

[23] To estimate the propensity score, one can use either Logit or Probit models. However, the choice of model is not too critical since these two models tend to yield similar results (Caliendo and Kopeinig, 2008).

matching method (i.e., one-to-one matching), a formal job is paired with an informal job, which is closest in terms of the estimated propensity score.

The PSM algorithm is run with replacement. This means that a vacancy in the control group is considered only once. A caliper of 0.5% is chosen to improve the matching quality and to avoid bad matches. We also implement the common support condition by discarding all observations whose propensity score is greater than the maximum and smaller than the minimum in the opposite group. This ensures that any combination of skill requirements and characteristics observed in the treated jobs can also be observed among the control jobs (Bryson et al., 2002).

As a result of our matching algorithm, we obtain a sample of vacancies that are balanced in terms of education level, experience level, job location, job position, industry, occupation, required skills, posted date and firm size. In total, the matched sample has 50,604; 699; 93,299 and 7,680 matched vacancies for each of our four formality/informality definitions, respectively.

Having matched job postings, we now run the wage equation on the matched sample to further control for any residual imbalances in job characteristics ex-post matching using the following specification:

$$LogWage_{it} = \beta_0 + \beta_1 Formal/Informal_{it} + Controls_{it}\beta_2 + Fixed\ effects + \varepsilon_{it} \quad (4.2)$$

where i and t refer to vacancy i and posted time t. $LogWage_{it}$ is the logarithm of offered salary for job posting i posted on date t. *Formal/Informal* is a dummy variable which equals one if the vacancy is identified as formal/informal and zero otherwise. $Controls$ is a vector of covariates including 12 skill groups, education and experience level, job position and location. These variables are included following previous studies on wage determination (Brown, 1980; Rosen, 1974). *Fixed effects* include time (i.e., the quarter when the job ad was first posted), industry, job title and firm size fixed effects. ε_{it} is an error term. We are interested in the sign and magnitude of the estimated coefficient β_1, which can reveal the wage gap between formal and informal job sectors.

4.4.2.2. Coarsened exact matching

Furthermore, to identify formal jobs that display the same observable characteristics as informal ones, we use an alternative matching technique, i.e., coarsened exact matching (CEM) by Iacus et al. (2012). The matching job characteristics are job title, required education, experience, job position, industry, posted quarter and firm size. Since all these variables are categorical, the CEM algorithm is equivalent to exact matching in the sense that it matches a treated unit to control units with the same covariate values.

In particular, we perform the one-to-one exact matching algorithm and obtain 30,208; 336; 40,458 and 4,180 job postings, which are balanced in terms of skill requirements and job characteristics for each of our four formality/informality definitions, respectively. Next, we re-estimate model (2) on the CEM-matched samples to estimate the formal-informal employment wage gap.

4.4.2.3. Text matching

The final matching approach is text matching. Text matching has become popular in recent research as it is shown to surpass other approaches in performing textual analysis tasks (Roberts et al., 2020). For instance, Arts et al. (2018) use text matching to compare the technological similarity between patents. Chala et al. (2018) employ text matching to match job candidates' CVs with job vacancies posted on the same job site. Hoberg and Phillips (2016) use this technique to measure product resemblance based on the product's description text. Kammoun and Power (2022) use text matching to match firm names from different financial information dataset.

To start with, we first match formal and informal job postings based on the narrowly-defined occupation. Next, we remove stop words, special characters as well as punctuations and convert the remaining text to lowercase.

Moreover, we match formal jobs with informal jobs that have the closest job ad text. To do so, we apply the Jaccard similarity algorithm to compute the text similarity using the Python package *TextDistance*. The Jaccard similarity score can be computed as follows:

$$Jaccard(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A|+|B|-|A \cap B|} \quad (4.3)$$

Where A and B are the two job ad texts. $|A \cap B|$ represents the number of common words in two job ad texts and $|A \cup B|$ is the sum of words present in either of the two texts. The values of this score range from zero to one. If this equals zero, the two documents are totally different. The higher the score, the more similar the two vacancies are. The basic idea of this approach is that two job postings are considered similar if they share a significantly large number of common words.

For each job ad in the treated group, we calculate the similarity scores between itself and each of job ads in the control group and then keep the job in the control group that has the largest Jaccard score. To increase matching quality and prevent bad matches, we remove from our analysis matched vacancies that have a Jaccard score of less than 50%. This results in a final text-matched sample of 7,598; 373; 12,038 and 1,800 job postings for each of our four formality/informality definitions, respectively. Finally, we re-run model (4.2) on this text-matched sample to estimate the wage gap between formal and informal jobs.

4.4.3. Heterogeneity of the formal and informal wage gap

In this section, we examine the extent to which the formal and informal wage gap changes across job positions and jobs requiring different education and experience levels. First, to capture the effects of education level heterogeneity, we re-estimate the model (4.2) by adding the interaction terms between formal/informal job indicator with education categories. Second, to capture the heterogeneous formal-informal wage gap across experience levels, we interact the formality/informality indicator with experience categories. Third, we add interaction terms between formal/informal job indicator and job position dummies to investigate the heterogeneous wage gap across job positions. Finally, we include interaction terms between formal/informal variable and job locations to examine the heterogeneous wage gap across geographical locations.

4.5. Common support and balance diagnostics for PSM and CEM

4.5.1. PSM

In this section, we check the common support assumption for our PSM estimations. The common support condition implies that any combination of job characteristics observed in the treated group can also be observed in the control group (Bryson et al., 2002). To impose common support, following Caliendo and Kopeinig (2008), we consider the minima and maxima of the propensity score distribution across the treated and control groups and any observations with the propensity score lower than the minimum or higher than the maximum level of the opposite group will be discarded. This condition is satisfied since we implement the common support option to discard observations in treated and control groups that are outside the common support zone.

Another task we need to perform is checking the covariate balance between treated and control groups. In particular, we need to compare the distribution of job characteristics between the treated and control groups after matching. The more similar the two distributions, the better the matching quality. One way to check for covariate balance is to look at the standardised bias score, which compares the difference in means of job characteristics between the treated and control groups and is not affected by sample size (Austin, 2011). As suggested by Normand et al. (2001), a standardised difference of equal or less than 10% can be considered trivial, which implies that we achieve a good balance among covariates (i.e., job characteristics).

Figure 4.5 shows that the majority of job characteristics obtain balance after matching and there are only a few covariates with standardised bias exceeding this threshold. More specifically, for the first formality measure, only language and customer service skills have the standardised difference of just over 10% (i.e., 11.2% and 13.8%, respectively), which is considered as a moderate imbalance. For the first informality measure, Vocational training, computer and social skills are not adequately balanced with the standardised bias of 21.7%, 20.4% and 16.4%, respectively. As for the second informality measure, all covariates are adequately balanced after matching with the maximum standardised bias of 7.8%. Finally, for the third informality measure, 1-2 years of experience, Ho Chi Minh city, language, character and cognitive skills are unbalanced with the standardised bias of 10.4%, 10.3%, 13.4%, 10.8% and 25.7%, respectively.

Yet, the concern about remaining covariate unbalances can be addressed by regression adjustment (i.e., running the wage regression by including unbalanced covariates as independent variables after PSM), which can remove residual confounding bias if there are still some covariate imbalances after matching (Nguyen et al., 2017).

[Insert Figure 4.5 here]

4.5.2. CEM

Table 4.4 shows that the degree of covariate imbalance after CEM matching is negligible. The overall imbalance is reflected by the L1 statistic, which is the percentage of overlap between the density of the two distributions of treated and control groups. While the value of zero (i.e., L1 = 0) indicates perfect overall balance, higher values indicate a lower balance between treated and control units, and the value of one indicates complete imbalance. We observe substantial declines in the L1 statistics after matching for all four measures (e.g., from 0.7761 to 0.1572 for the first formality measure). This suggests that our matching algorithms achieve significant improvements in balancing covariate.

We also report unidimensional measures of imbalance for each job dimension/characteristic separately. In general, our matching achieves perfect balance in the means, marginal and joint distributions of almost all job characteristics for all four formality/informality measures. The only exception is the job title and job industry, where the imbalance still remains negligible.

[Insert Table 4.4 here]

4.6. Result discussion

4.6.1. Characteristics of formal and informal jobs

The results obtained from the examination of the requirement and characteristics of formal and informal jobs are shown in Table 4.5. As for the required education, for the first formality measure, the probability of being a formal job, compared to High school level, declines if Vocational training or Associate degree are required or no specific educational qualification is stated in the vacancy. In other words, formal jobs are more likely to request High school and Bachelor's degree than other qualifications. For the three informality measures, we can observe a clearer pattern that informal jobs tend to target High school graduates and have a lower chance of requesting higher education levels. This finding is in line with findings from previous research on the characteristics of informal jobs and informal workers (Angel-Urdinola and Tanabe, 2012; McCaig and Pavcnik, 2015; O'Higgins and Viegelahn, 2021). For example, McCaig and Pavcnik (2015) find that formal workers are more likely to be better educated than informal workers in the context of Vietnam from 1999 to 2009.

Now turn to the required work experience, a less clear-cut trend to the required education can be observed. That is, formal jobs, as defined by our first formality measure, are more likely to target candidates with very little work experience or highly experienced candidates. Particularly, compared to less than one year of experience, there is a 15.4% higher probability of a formal job requiring more than ten years of experience, but 0.9% to 6.3% lower probability of a formal job requiring one to ten years of experience. Informal jobs, as defined by the first and third measures, have a higher chance of requesting less than one year of experience and a lower chance of requesting more work experience. In contrast, informal jobs, as classified by the third definition, tend to target experienced workers with more than one to ten years of experience.

With regard to the job position, results in the first two columns show that the higher the job position is, the higher probability the vacancy is a formal job. The opposite findings are found for informal jobs according to the second and third definitions. That is, informal jobs have up to 6.4% lower probability of being at high positions (i.e., employee, medium manager or top manager) than at entry level. The results for the second informal measure are slightly different:

informal jobs have just a marginally higher chance (i.e., 0.4%) of being at an employee position than at entry level.

Turning now to the job location, the results are mixed depending on the chosen measure. For example, when using the first formal indicator, formal jobs are more likely to be located in Hanoi and less likely to be in Ho Chi Minh city than in all other smaller cities. Yet, when the third and third informal measures are used, informal jobs are also more likely to be located in Hanoi and less likely to be located in HCM city.

In the case of skill requirements, some skills are more likely to be required by formal jobs, and some other skills are more likely to be required by informal jobs. Particularly, formal jobs have a higher probability of requiring hard skills such as software, language and cognitive skills than informal jobs. For instance, formal jobs, as defined by the first measure, are 2% more likely to state cognitive skills in the job ad text than informal jobs. Likewise, informal jobs as defined by the first and second informal measures, are 0.2% and 0.8% less likely to request cognitive skills than formal jobs, respectively. In contrast, formal jobs display a lower propensity to require soft skills such as writing and beauty requirements than informal ones.

[Insert Table 4.5 here]

4.6.2. The formal-informal wage gaps

4.6.2.1. PSM-matched sample

The results from regressing the wage equation on the PSM-matched sample are reported in Table 4.6. From this table, we can see that there is a positive wage gap between formal jobs and informal jobs. In particular, according to the first formality definition, on average, formal jobs can offer 1.7 percent higher wages than informal ones. The results are consistent throughout the three informality definitions. That is, informal jobs may offer 6.1%, 0.8% and 6% lower wages than formal jobs, according to the first, second and third informality measures, respectively.

[Insert Table 4.6 here]

This finding is in line with previous studies that also document the wage premium for the formal employment sector and wage penalty for the informal employment sector (see, e.g., Marcouiller et al., 1997; Lehmann and Zaiceva, 2013; Bargain and Kwenda, 2014; Tansel et al., 2020). Possible explanations for this result can be provided by the segmented labour market theory and compensating differentials theory.

First, the segmented labour market theory sees informal employment as a survival alternative to avoid involuntary unemployment for low-productivity or disadvantaged workers who only work in the informal sector until they find a better employment opportunity (Dickens and Lang, 1985). As a consequence, there exists is a wage dualism for workers of similar characteristics depending on the labour market segment in which they work. Particularly, while for the formal job segment, there is a limited labour supply that induces higher wages, the informal segment has no institutional or efficiency-wage basis for regulating wages. In addition, low entry barriers and an abundant supply of unskilled workers lead to low wages. Thus, wages do not depend on the workers' skills but on the sector in which they are employed (Uribe et al., 2007).

Second, according to the compensating differentials theory, informal sector workers may accept lower wages in exchange for flexibility in working hours and choice of work location (Maloney, 1999). Alternatively, it might be the case that wages of the informally hired workers are not subject to taxation and other social security contribution payments, wages offered by informal jobs can be higher compared to post-tax wages offered by formal jobs (Henley et al., 2009).

Next, the estimation results for the effects of education qualifications on offered wage show that the higher the education qualification is required, the higher the wage is offered by the employers. For example, on average, the vacancy requiring a Bachelor's degree offers 41.7% and 44.3% higher wage than the vacancy requiring only High school graduate, according to estimations using the first formality measure and the second informality measure, respectively.

A similar finding can be observed for the required work experience. That is, the more work experience is required, the higher the wage is offered by the employers. For instance, on average, when a job requires two to five years of work experience instead of less than one year of experience, the employer will advertise from one-fourth to nearly one-half higher salaries in the job posting, depending on the chosen formal/informal measure. When the required work experience reaches more than ten years, the offered wage can increase substantially by more than one-half.

As expected, the results for job positions also follow similar patterns in the sense that the higher the job position is, the higher the offered wage is. For example, non-manager level jobs can offer from 14.1% to 55.4% higher wages than jobs of new entry. This positive wage effect is even more noticeable for medium-managerial positions with a wage premium of 36.4% to 87.9%, or top-managerial positions with a wage premium of 81.5% to 83.8%.

We further find a significant urban wage premium, which is in line with much of the existing literature (Glaeser and Maré, 2001; Heuermann et al., 2010, Roca and Puga 2017; Hirsch et al., 2022). More specifically, if the job is located in Hanoi or Ho Chi Minh city, which are two major cities in Vietnam, the offered wage can be 5.6 to 7 percent higher than that offered by jobs located in smaller towns. It is possible that the urban wage premia stem from agglomeration economies that result in higher worker productivity in large and dense labour markets (Puga, 2010; Moretti, 2014). It might also be explained by the higher level of competition in urban labour markets, where employers possess less wage-setting power over their employees, hence they need to offer a higher wage than in less competitive labour markets (Combes and Gobillon, 2015; Hirsch et al., 2022).

Among twelve skill groups, there are skills that are positively rewarded regardless of the formality/informality measure used. Those skills include soft skills such as artistic, customer service and social skills, as well as hard skills such as specific software, language, project management and cognitive skills. A general and common skill that presents in a large proportion of job ads, such as good character skills tends to signal lower-paying jobs. That is, on average, jobs specifying character skills pay 2.2 to 5.8 percent lower wages than jobs without this skill requirement. For the remaining skill groups, the wage effects are mixed depending on the chosen formality/informality measure.

4.6.2.2. CEM-matched sample

The estimation results obtained from the CEM-matched sample reported in Table 4.7 are consistent with those from the PSM-matched sample in showing a wage premium for formal jobs (i.e., 1.9%) and wage penalties for informal jobs (i.e., 1.1% to 7.6%).

As for other job characteristics, we also find higher wages among jobs requiring higher education levels than jobs requiring High school graduates. A similar finding can be obtained for the required work experience: jobs requiring more work experience would offer a higher salary than jobs requiring less than one year of work experience. In the same vein, higher wages are associated with higher managerial job positions. We also find significant wage premia for jobs located in the two major cities, in line with the results obtained from the PSM-matched sample.

[Insert Table 4.7 here]

4.6.2.3. Text-matched sample

According to results reported in Table 4.8, the estimation on the text-matched sample also provides consistent results with those from that on the PSM-matched sample in revealing a wage premium for formal jobs (i.e., 1.6%) and wage penalties for informal jobs (i.e., 2% to 21.4%).

Moreover, we find lower wages among jobs requiring low education levels and higher wages among those requiring higher education qualifications. Similar results can be observed for the required work experience: jobs requiring zero to little work experience tend to offer a lower salary than jobs requiring more work experience. Likewise, managerial jobs tend to offer higher wages than entry-level jobs. We also find significantly higher salaries among jobs located in the two major cities, in line with the results obtained from the PSM-matched and CEM-matched samples.

[Insert Table 4.8 here]

4.6.3. Heterogeneity of the formal-informal wage gap

The results reported in Table 4.9 suggest that the formal-informal employment wage gap tends to be smallest among jobs in low-education job segments and then is widened as the required education increases. For instance, as can be seen from column (1) for the first formality measure, there is no significant wage gap between formal and informal jobs among jobs requiring High school graduates. Yet, among jobs requesting Vocational training, formal jobs can offer 6.2% higher wages than informal ones. The wage premium for the formal sector is even getting larger to reach 27% among jobs requiring Bachelor's degree. The same pattern can be observed from the estimation results provided by the second and third informality measures in columns (3) and (4): the wage gap between informal and formal jobs is worsened among jobs at high education levels. This finding is in line with that found in Gong and Van Soest (2002), which shows a significant wage premium for formal jobs for highly educated male workers but not for male workers with a lower level of education in urban Mexico.

[Insert Table 4.9 here]

We draw mixed conclusions from the results in Table 4.10 on the heterogeneity of the formal-informal employment wage gap across required work experience. That is, according to the results from the first formality measure in column (1), the formal-informal employment wage gap is positive among jobs at low-experience job segments, and then the gap is narrowed down as the required work experience reaches 2-5 years. For jobs requesting more work experience, the formal-informal employment wage gap even turns negative, suggesting that for highly skilled workers, informal jobs may offer a higher wage than formal ones.

However, if we turn to three informality measures, the estimation results in columns (3), (4) and (5) suggest otherwise. More specifically, we find that the informal-formal employment wage gap tends to be smallest among jobs with low experience and then is broadened as the required work experience rises. For example, according to the results obtained from the third informality measure, among jobs requesting less than one year of experience, informal jobs will offer 7.5% lower wages than formal ones. The wage penalty for the informal sector is getting worse to reach 12.1%, 38% and 66% among jobs requiring 1-2 years, 2-5 years, 5-10 years of experience, respectively.

[Insert Table 4.10 here]

Table 4.11 shows estimates for the heterogeneous formal-informal wage gap across job positions. Similar to the previous section, we also find mixed results about the heterogeneous wage gap. Particularly, according to the results from the first formality measure and the first informality measure, the formal-informal employment wage gap is positive among jobs at the entry level, and then the gap is narrowed down among jobs at non-manager positions. For managerial positions, the wage gap turns negative, revealing that for managerial jobs, informal jobs may offer a higher wage than formal ones.

Nevertheless, if we turn to the second and third informality measures, the estimation results in columns (3) and (4) suggest otherwise. In particular, we find that the informal-formal employment wage gap tends to be widened as we go up the scale to higher-position jobs. For example, according to the results obtained from the third informality measure, among entry-level jobs, there is no significant difference in the offered wages between two sectors. Then there are wage penalties of 15.6% and 42.5% for informal jobs at medium-manager and top-manager levels, respectively.

[Insert Table 4.11 here]

4.7. Conclusion

In this paper, we examine the wage differentials between formal and informal jobs. We employ a rich dataset of online job ads gathered from one of the largest job portals in Vietnam. To make our estimations valid, we use different matching methods (i.e., PSM, CEM and text-matching) to match vacancies on all observable characteristics, including job title, location, education level, experience level, job position, skills required and firm size.

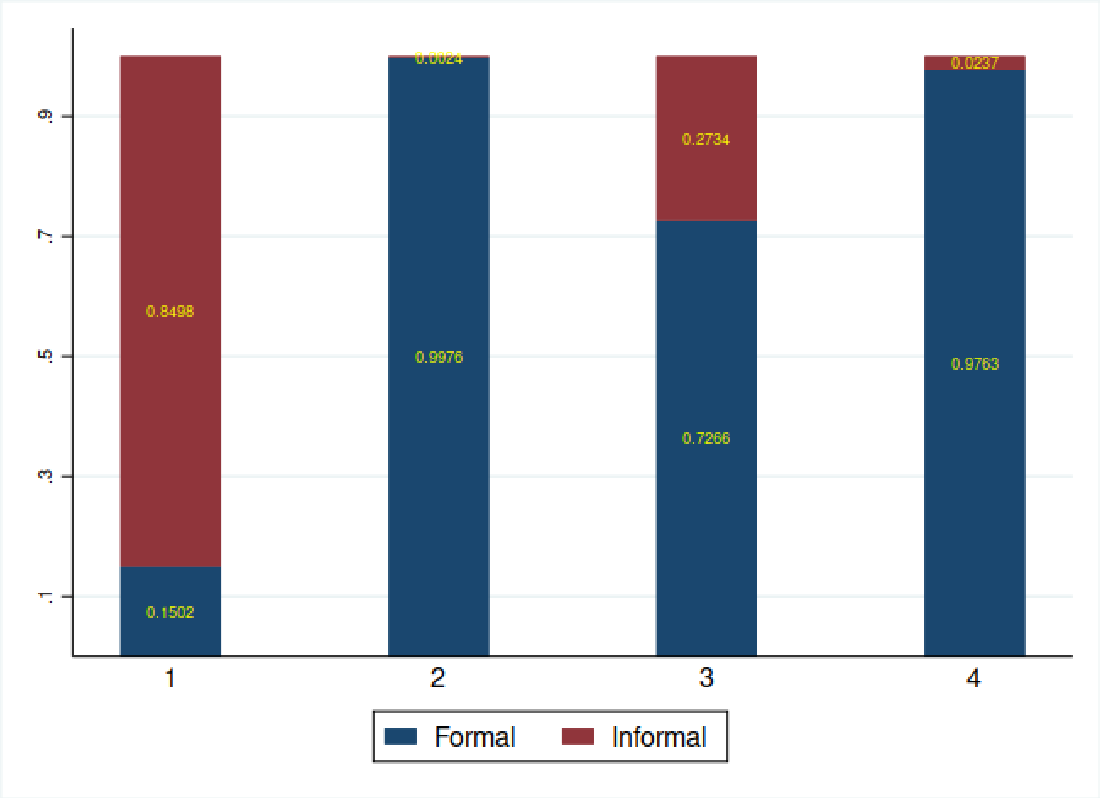
According to our estimation results, there is a significant wage gap between formal and informal employment. Particularly, we find significant wage premia of 1.6 to 1.9 percent for formal jobs and wage penalties of 0.8 to 21.4 percent for informal ones. These findings suggest that informal jobs – regardless of the definition/measure used, offer lower wages than formal jobs. The result can be explained on the ground of the segmented labour market theory, which considers informal jobs as a survival mechanism for workers who cannot enter into the formal job sector. Another explanation relies on the compensating differentials theory, according to which informal sector workers earn lower wages in exchange for flexibility in working hours, place of work or tax and other social security contributions avoidance.

Moreover, the formal-informal jobs wage differentials vary across different education, experience levels and job positions. More specifically, workers in formal jobs can be even more advantaged in jobs requiring a higher level of education. That is, the formal-informal employment wage gap tends to be the smallest among low-education jobs and is widened as the required education increases. Although the wage gap is considerably heterogeneous across jobs of different required work experience and job positions, there are no clear trends for such heterogeneities.

Although there are labour laws, there is no information regarding the extent of their enforcement in reality. Hence our paper aims to inform policymakers about the degree of informality and the wage gap between formal and informal jobs. It is necessary to have a stricter legal framework and impose a higher level of regulations enforcement to reduce informal sector employment and to protect the employees in this sector. However, estimating the wage gap might encounter disadvantages using traditional labour market data sources, which might be biased due to the lack of many labour market dimensions and might not be promptly and regularly updated. In our paper, we can address this potential issue by making use of online vacancies as a rich and real-time data source for instantaneous labour market dynamics analyses.

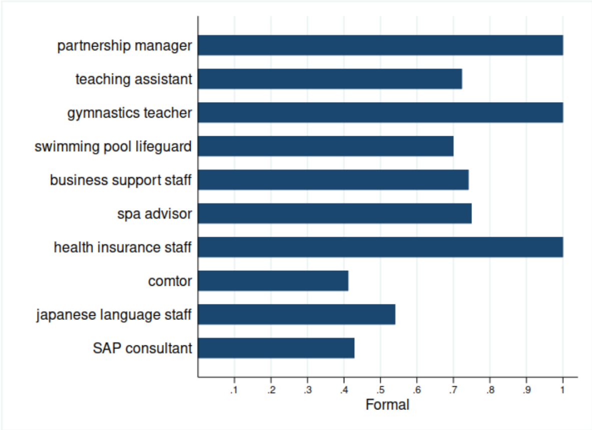
Figures

Figure 4.1. Share of formal and informal vacancies

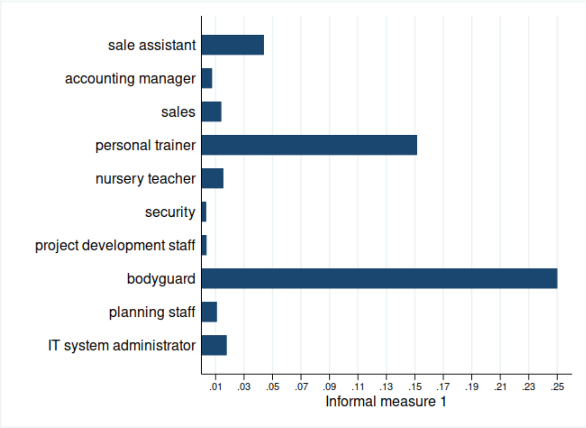


Notes: The figure illustrates the fraction of formal and informal vacancies by different formal/informal measures: (1) first formality measure, (2) first informality measure, (3) second informality measure and (4) third informality measure.

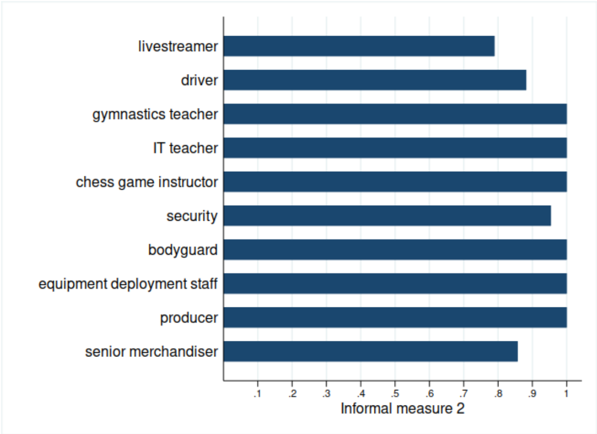
Figure 4.2. Top ten occupations with highest fraction of formal/informal jobs



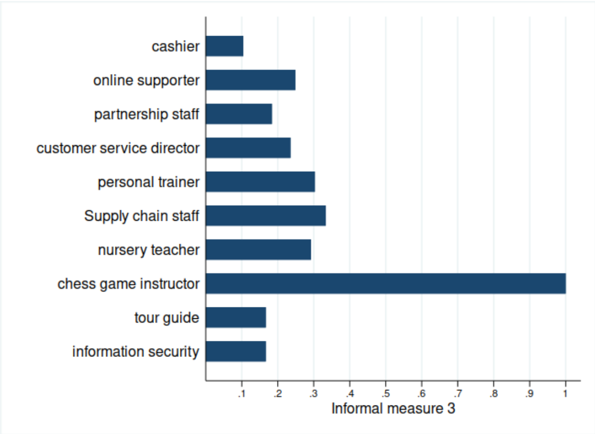
(a)



(b)



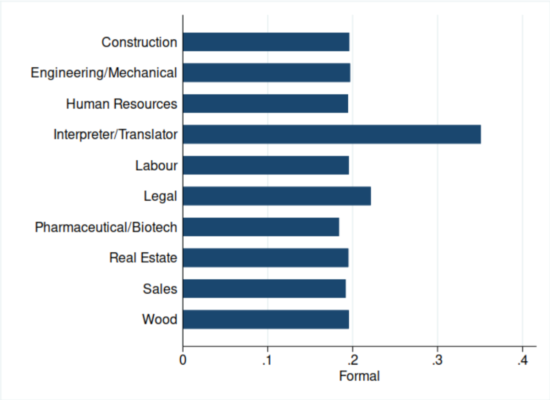
(c)

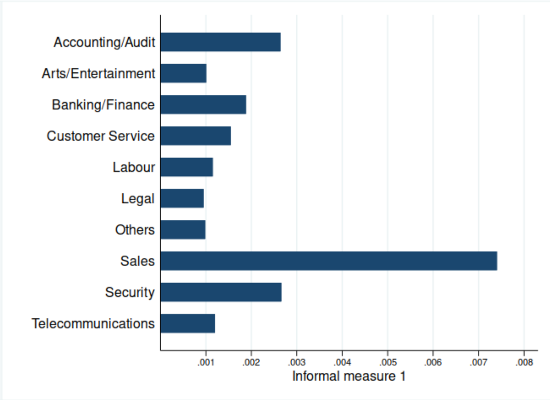


(d)

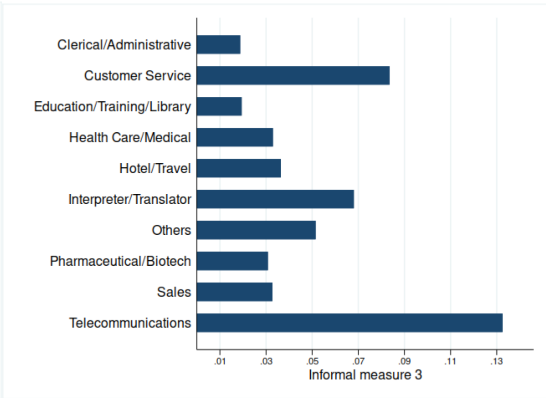
Notes: The figure shows the top ten job titles with the highest percentage of formal and informal vacancies for the (a) first formality measure, (b) first informality measure, (c) second informality measure and (d) third informality measure.

Figure 4.3. Top ten industries with highest fraction of formal/informal jobs





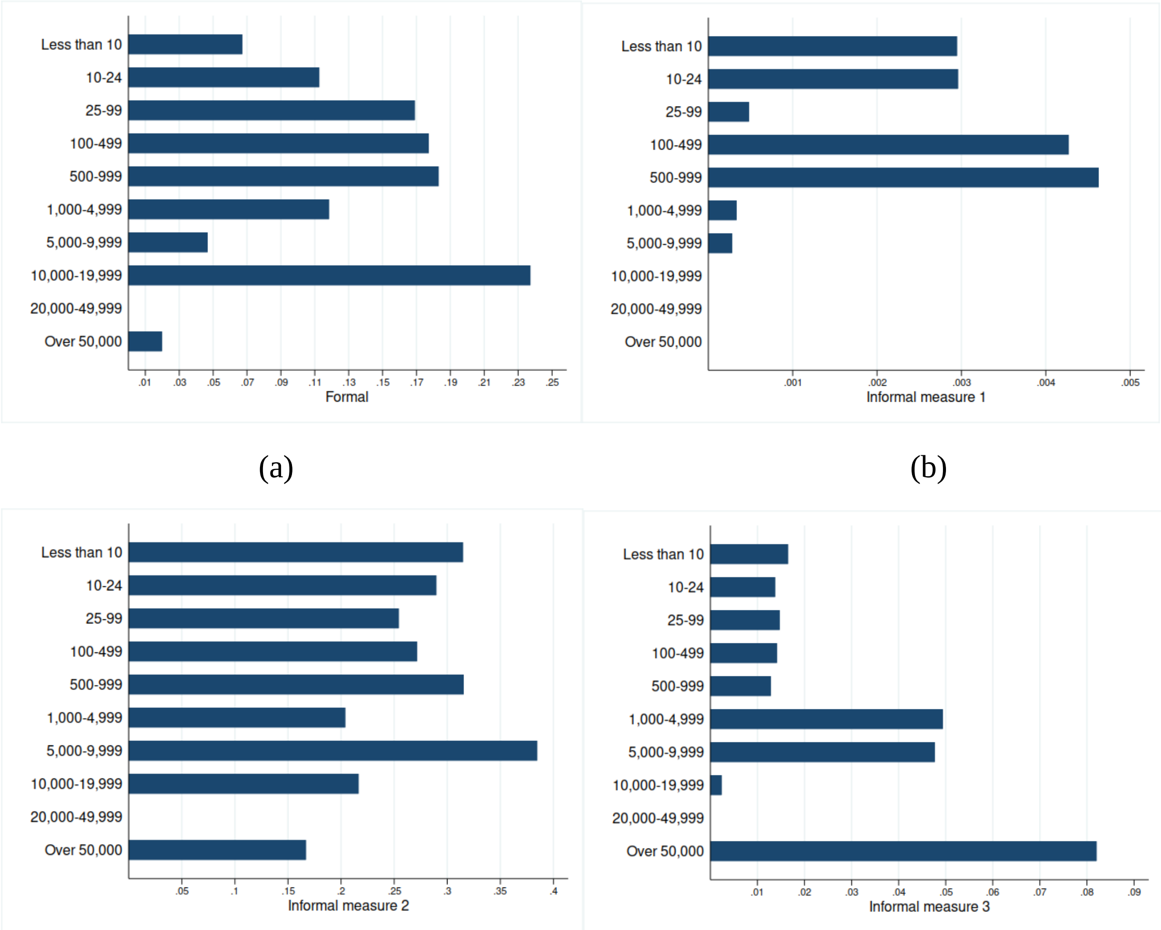
(a) (b)



(c) (d)

Notes: The figure shows the top ten industries with the highest percentage of formal and informal vacancies for the (a) first formality measure, (b) first informality measure, (c) second informality measure and (d) third informality measure.

Figure 4.4. Fraction of formal/informal jobs by firm sizes

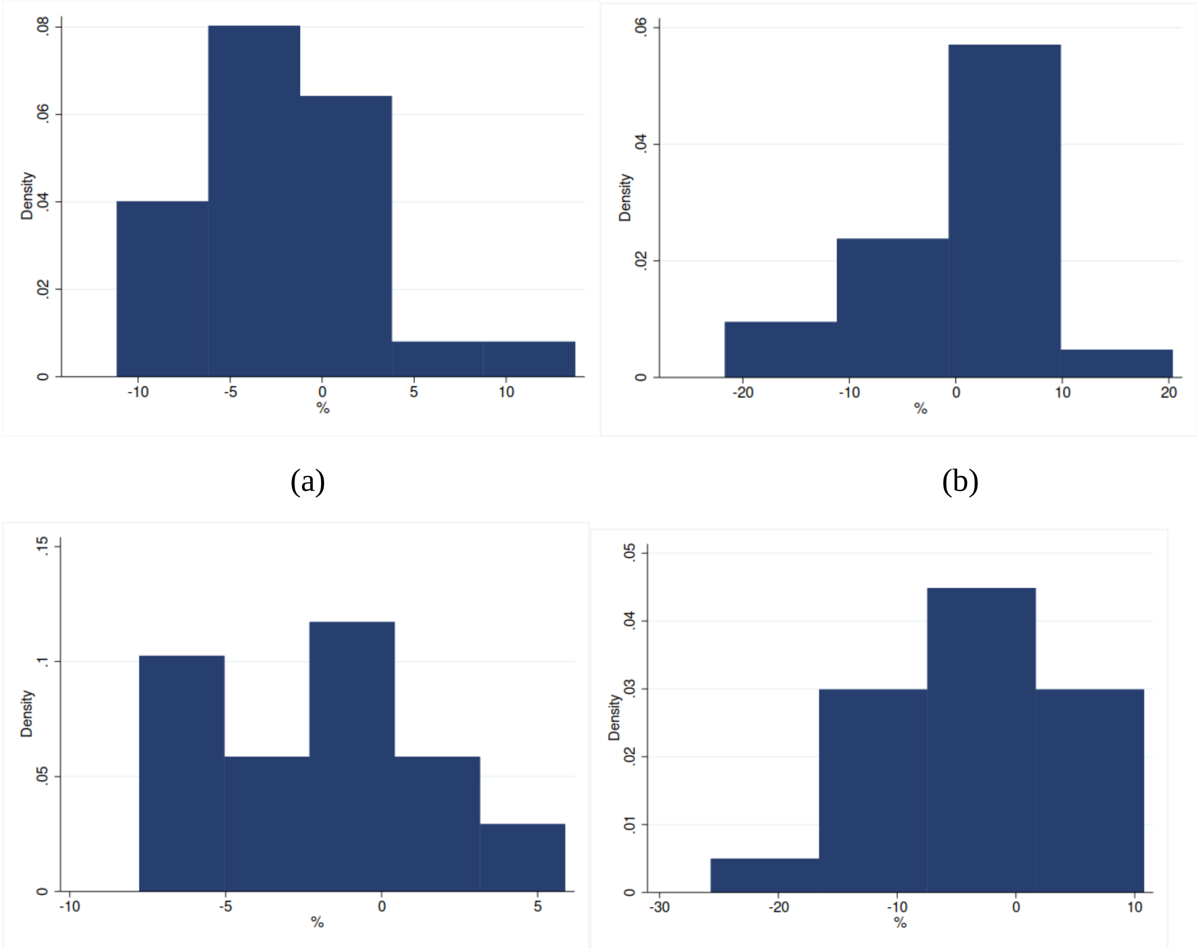


(a) (b)

(c) (d)

Notes: The figure shows the percentage of formal and informal vacancies across firm sizes for the (a) first formality measure, (b) first informality measure, (c) second informality measure and (d) third informality measure.

Figure 4.5. Standardised bias distribution across covariates after PSM



(a) (b)

(c) (d)

Notes: This figure shows the standardised bias distribution across covariates after propensity score matching for the (a) first formality measure, (b) first informality measure, (c) second informality measure and (d) third informality measure. The horizontal line presents the standardised bias value in % and the vertical line presents the density.

Tables

Table 4.1. List of keywords

Skill	Keywords
Art	Art, artistic
Character	Organised, detail oriented, multitasking, time management, meeting deadlines, energetic
Cognitive	Problem solving, research, analytical, critical thinking, math, statistics
Computer	Computer, spreadsheets, common software (e.g., Microsoft Excel, PowerPoint)
Customer service	Customer, sales, client, customer service
Financial	Budgeting, accounting, finance, cost
Language	English, Japanese, Korean, Chinese, foreign language
Project management	Project management
People management	Supervisory, leadership, management (not project), mentoring, staff
Social	Communication, teamwork, collaboration, negotiation, presentation
Software	Programming language or specialized software (e.g., Java, SQL, C++)
Writing	Writing
Attractiveness	Pretty face, attractive face, good looking, pretty, nice face
Formal	Social security, social insurance, maternity leave, maternity benefit
Informal_1	Married, single
Informal_3	No disability, not disable, no hearing problems

Notes: The table presents the classification of keywords extracted from job ad text into skill groups.

Table 4.2. Correlation matrix

	Formal	Informal_1	Informal_2	Informal_3	Education	Exper.	Level	Computer	Software	Language	Financial	People.	Project.	Art	Character	Cognitive	Customer	Social
Informal_1	0.0845																	
Informal_2	-0.0059	0.0128																
Informal_3	0.0606	-0.0061	-0.0054															
Education	-0.0065	-0.0617	-0.0430	-0.0470														
Experience	-0.0149	-0.0357	0.0138	-0.0717	0.4620													
Level	-0.0015	-0.0117	-0.0581	-0.0224	0.2306	0.4419												
Computer	0.0085	0.0005	0.0678	-0.0241	0.1480	0.1390	0.0408											
Software	0.0114	-0.0196	-0.0194	-0.0486	0.2828	0.2730	0.0370	0.2389										
Language	0.0501	-0.0184	-0.0120	0.0437	0.3079	0.1926	0.0752	0.1670	0.1257									
Financial	-0.0222	-0.0068	0.0558	-0.0454	0.0584	0.1302	-0.0275	0.1872	0.0987	-0.0642								
People Mgmt.	0.0114	-0.0085	-0.0205	-0.0129	0.1398	0.1735	0.2166	0.0379	0.0581	0.0395	-0.0059							
Project Mgmt.	0.0276	-0.0090	-0.0087	-0.0356	0.2054	0.2640	0.3036	0.1024	0.1149	0.0666	0.0166	0.1792						
Art	-0.0022	-0.0125	-0.0360	-0.0286	0.0893	0.1215	0.0407	0.0102	0.2244	0.0176	-0.0499	0.0230	0.0729					
Character	0.0397	0.0143	-0.0303	0.0127	-0.1399	-0.1745	-0.0835	0.0134	-0.0999	-0.1319	0.0067	-0.0033	-0.0024	-0.0031				
Cognitive	0.0422	-0.0255	-0.0290	0.0443	0.2547	0.2872	0.1574	0.1650	0.2619	0.1556	0.0224	0.1328	0.2635	0.1919	0.0735			
Customer	-0.0201	-0.0329	-0.1952	0.0213	0.0291	-0.0544	0.0470	-0.0077	-0.0829	-0.0441	-0.0763	0.0023	0.0400	-0.0008	0.1649	0.0527		
Social	0.0177	0.0230	-0.1433	0.0855	-0.0150	-0.0895	0.0424	-0.0207	-0.1240	0.0826	-0.0916	0.0386	0.0592	-0.0237	0.2259	0.0561	0.3581	
Writing	-0.0146	-0.0053	-0.0016	0.0019	0.0997	0.0677	-0.0055	0.0739	0.0702	0.0888	-0.0175	0.0371	0.0400	0.0754	0.0034	0.1032	0.0573	0.0155

Notes: The table shows the Spearman correlation coefficients between formal, informal job indicators and skill requirement variables.

Table 4.3. Descriptive statistics

Variable	Mean (1)	SD (2)	Min (6)	p25 (3)	p50 (4)	p75 (5)	Max (7)
Wage	9,117,274	6,099,864	2,920,000	6,000,000	7,500,000	10,000,000	300,000,000
Formal	0.1502	0.3573	0	0	0	0	1
Informal_1	0.0024	0.0488	0	0	0	0	1
Informal_2	0.2734	0.4457	0	0	0	1	1
Informal_3	0.0237	0.1520	0	0	0	0	1
Education							
High school/Vocational	0.1378	0.3447	0	0	0	0	1
Associate degree	0.6529	0.4761	0	0	1	1	1
Bachelor’s degree	0.2058	0.4043	0	0	0	0	1
Master degree	0.0008	0.0288	0	0	0	0	1
No education required	0.0000	0.0024	0	0	0	0	1
Years of experience							
0-1 year	0.4320	0.4954	0	0	0	1	1
1-2 years	0.3440	0.4751	0	0	0	1	1
2-5 years	0.1967	0.3975	0	0	0	0	1
5-10 years	0.0253	0.1570	0	0	0	0	1
10+ years	0.0019	0.0440	0	0	0	0	1
Position							
New entry	0.0310	0.1733	0	0	0	0	1
Employee	0.8375	0.3689	0	1	1	1	1
Medium-level manager	0.1221	0.3274	0	0	0	0	1
Top manager	0.0094	0.0966	0	0	0	0	1
Location							
Hanoi	0.2540	0.4353	0	0	0	1	1
HCM city	0.3420	0.4744	0	0	0	1	1
Other locations	0.4040	0.4907	0	0	0	1	1
Skill							
Computer	0.3242	0.4681	0	0	0	1	1
Software	0.2086	0.4063	0	0	0	0	1
Financial	0.1404	0.3474	0	0	0	0	1
People Management	0.0380	0.1912	0	0	0	0	1
Project Management	0.0980	0.2974	0	0	0	0	1
Artistic	0.0900	0.2861	0	0	0	0	1
Language	0.2120	0.4087	0	0	0	0	1
Character	0.8566	0.3504	0	1	1	1	1
Cognitive	0.2934	0.4553	0	0	0	1	1
Customer Service	0.3675	0.4821	0	0	0	1	1
Social	0.5949	0.4909	0	0	1	1	1
Writing	0.0238	0.1524	0	0	0	0	1
Attractiveness	0.1260	0.3318	0	0	0	0	1
Obs				**173,106**			

Notes: The table shows summary statistics of all variables used in the regression. Columns (1) and (2) report the mean and standard deviation, respectively. Columns (3) to (7) report the min, 25th percentile, median, 75th percentile and maximum values, respectively.

Table 4.4. CEM covariate balance checks

Panel A: First formality measure							
Multivariate L1 distance before matching: 0.7761							
Multivariate L1 distance after matching: 0.1572							
Univariate imbalance							
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
	L1	**Mean**	**Min**	**25%**	**50%**	**75%**	**Max**
Education	0	0	0	0	0	0	0
Experience	0	0	0	0	0	0	0
Level	0	0	0	0	0	0	0
Location	0	0	0	0	0	0	0
Quarter	0	0	0	0	0	0	0
Industry	0.0075	-0.0035	0	0	0	0	0
Job title	0.0219	-0.0384	0	0	0	0	2
Firm size	0	0	0	0	0	0	0

Panel B: First informality measure							
Multivariate L1 distance before matching: 0.9960							
Multivariate L1 distance after matching: 0.0238							
Univariate imbalance							
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
	L1	**Mean**	**Min**	**25%**	**50%**	**75%**	**Max**
Education	0	0	0	0	0	0	0
Experience	0	0	0	0	0	0	0
Level	0	0	0	0	0	0	0
Location	0	0	0	0	0	0	0
Quarter	0	0	0	0	0	0	0
Industry	0	-0.0119	0	0	0	0	0
Job title	0.0179	0.25	0	0	0	0	-2
Firm size	0	0	0	0	0	0	0

Panel C: Second informality measure							
Multivariate L1 distance before matching: 0.7832							
Multivariate L1 distance after matching: 0.2093							
Univariate imbalance							
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
	L1	**Mean**	**Min**	**25%**	**50%**	**75%**	**Max**
Education	0	0	0	0	0	0	0
Experience	0	0	0	0	0	0	0
Level	0	0	0	0	0	0	0
Location	0	0	0	0	0	0	0
Quarter	0	0	0	0	0	0	0
Industry	0.0243	0.0009	0	0	0	0	0
Job title	0.0552	0.3837	0	5	0	0	0
Firm size	0	0	0	0	0	0	0

Panel D: Third informality measure

Multivariate L1 distance before matching: 0.9355							
Multivariate L1 distance after matching: 0.0785							
Univariate imbalance							
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
	L1	**Mean**	**Min**	**25%**	**50%**	**75%**	**Max**
Education	0	0	0	0	0	0	0
Experience	0	0	0	0	0	0	0
Level	0	0	0	0	0	0	0
Location	0	0	0	0	0	0	0
Quarter	0	0	0	0	0	0	0
Industry	0.0215	0.0244	0	0	0	0	-1
Job title	0.0335	0.0938	0	1	0	0	0
Firm size	0	0	0	0	0	0	0

Notes: The table reports multivariate and univariate imbalance measures for the first formality measure, first informality measure, second and third informality measures in Panels A, B, C and D, respectively. For all panels, column (1) reports the L1 imbalance measure for individual variable. Column (2) reports the difference in the density distributions between treated and control groups at mean. Columns (3) to (7) report the difference in the density distributions between treated and control groups for the 0^{th} (Min), 25^{th}, 50^{th}, 75^{th} and 100^{th} (Max) percentiles, respectively.

Table 4.5. Regression results of Probit model for the characteristics of formal/informal vacancies

	Formal		Informal 1		Informal 2		Informal 3	
	(1) Coefficient	**(2) Marginal effect**	**(3) Coefficient**	**(4) Marginal effect**	**(5) Coefficient**	**(6) Marginal effect**	**(7) Coefficient**	**(8) Marginal effect**
Vocational training	-0.241***	-0.056***	-1.060***	-0.014***	0.054***	0.014***	-0.066**	-0.003**
	(0.015)	(0.004)	(0.086)	(0.002)	(0.013)	(0.004)	(0.029)	(0.001)
Associate degree	-0.266***	-0.061***	-0.887***	-0.013***	-0.094***	-0.024***	-0.070*	-0.003*
	(0.018)	(0.004)	(0.120)	(0.002)	(0.017)	(0.004)	(0.041)	(0.002)
Bachelor’s degree	0.184	0.050			-0.681***	-0.145***		
	(0.121)	(0.035)			(0.137)	(0.022)		
Other educations	-0.366***	-0.081***			0.227***	0.063***	-0.194	-0.008
	(0.084)	(0.016)			(0.068)	(0.020)	(0.208)	(0.008)
1-2 years	-0.042***	-0.009***	-0.256***	-0.002***	0.049***	0.013***	-0.239***	-0.011***
	(0.010)	(0.002)	(0.082)	(0.001)	(0.010)	(0.003)	(0.021)	(0.001)
2-5 years	-0.102***	-0.022***	0.177	0.002	0.024*	0.006*	-0.344***	-0.015***
	(0.014)	(0.003)	(0.112)	(0.001)	(0.013)	(0.003)	(0.035)	(0.001)
5-10 years	-0.331***	-0.063***			0.105***	0.028***	-0.374***	-0.016***
	(0.031)	(0.005)			(0.026)	(0.007)	(0.114)	(0.004)
10+ years	0.557***	0.154***			0.097	0.026		
	(0.075)	(0.024)			(0.081)	(0.022)		
Non-manager	0.175***	0.034***	0.825***	0.004***	-0.052**	-0.014**	-0.012	-0.001
	(0.024)	(0.004)	(0.314)	(0.001)	(0.021)	(0.006)	(0.046)	(0.002)
Medium manager	0.177***	0.035***	0.711*	0.003	-0.248***	-0.064***	-0.255***	-0.010***
	(0.030)	(0.006)	(0.410)	(0.002)	(0.027)	(0.007)	(0.073)	(0.003)
Top manager	0.279***	0.058***			-0.198***	-0.052***		
	(0.066)	(0.015)			(0.058)	(0.015)		
Hanoi	0.031***	0.007***	0.696***	0.006***	-0.078***	-0.021***	0.516***	0.027***
	(0.011)	(0.002)	(0.091)	(0.001)	(0.010)	(0.003)	(0.022)	(0.001)
HCM	-0.053***	-0.011***	-0.199*	-0.001*	-0.058***	-0.015***	-0.014	-0.000
	(0.010)	(0.002)	(0.118)	(0.000)	(0.009)	(0.002)	(0.024)	(0.001)
Computer	0.008	0.002	0.456***	0.004***	0.156***	0.042***	-0.056***	-0.003***
	(0.009)	(0.002)	(0.077)	(0.001)	(0.009)	(0.002)	(0.021)	(0.001)
Software	0.018	0.004	-0.475***	-0.003***	-0.202***	-0.052***	-0.016	-0.001
	(0.012)	(0.003)	(0.105)	(0.001)	(0.011)	(0.003)	(0.033)	(0.001)
Language	0.152***	0.034***	-0.066	-0.001	-0.097***	-0.025***	0.377***	0.021***
	(0.011)	(0.003)	(0.099)	(0.001)	(0.010)	(0.003)	(0.022)	(0.001)
Financial	0.044***	0.009***	0.076	0.001	0.024*	0.006*	-0.398***	-0.015***
	(0.015)	(0.003)	(0.118)	(0.001)	(0.014)	(0.004)	(0.038)	(0.001)
People management	-0.048**	-0.010**	-0.052	-0.000	-0.097***	-0.025***	-0.143**	-0.006**
	(0.022)	(0.004)	(0.314)	(0.002)	(0.021)	(0.005)	(0.063)	(0.002)
Project management	0.122***	0.027***	0.411***	0.004***	0.044***	0.012***	-0.475***	-0.016***
	(0.014)	(0.003)	(0.116)	(0.002)	(0.013)	(0.004)	(0.047)	(0.001)
Art	-0.066***	-0.014***	-0.408***	-0.002***	-0.010	-0.003	-0.247***	-0.010***
	(0.015)	(0.003)	(0.154)	(0.001)	(0.014)	(0.004)	(0.041)	(0.001)
Character	0.211***	0.042***	-0.063	-0.001	0.059***	0.015***	-0.274***	-0.015***
	(0.012)	(0.002)	(0.101)	(0.001)	(0.011)	(0.003)	(0.028)	(0.002)
Cognitive	0.094***	0.020***	-0.271***	-0.002***	-0.032***	-0.008***	0.526***	0.030***
	(0.010)	(0.002)	(0.097)	(0.001)	(0.009)	(0.002)	(0.020)	(0.001)
Customer service	-0.086***	-0.018***	-0.523***	-0.003***	-0.092***	-0.024***	-0.148***	-0.007***
	(0.010)	(0.002)	(0.093)	(0.000)	(0.010)	(0.003)	(0.021)	(0.001)
Social	0.051***	0.011***	-0.098	-0.001	-0.105***	-0.028***	0.358***	0.015***
	(0.009)	(0.002)	(0.065)	(0.001)	(0.009)	(0.002)	(0.024)	(0.001)
Writing	-0.212***	-0.041***	0.269	0.003	0.193***	0.053***	0.191***	0.010***
	(0.028)	(0.005)	(0.253)	(0.003)	(0.025)	(0.007)	(0.049)	(0.003)
Obs	172,767	172,767	80,231	80,231	172,832	172,832	152,223	152,223

Notes: The table presents results of Probit models for job characteristics of formal/informal job ads. The first two columns (1) and (2) report results for the first formal indicator. The next two columns (3) and (4) report results for the first informal indicator. The next two columns (5) and (6) report results for the second informal indicator. The last two columns (7) and (8) report results for the third informal indicator. Odd columns (1), (3), (5) and (7) show the estimated coefficients. Even columns (2), (4), (6) and (8) show the average marginal effect. The dependent variable is a dummy variable that indicates whether a vacancy is classified as formal or informal according to each definition. Time, industry, job title, and firm size dummies are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table 4.6. Regression results on PSM-matched sample

	Formal measure 1 (1)	Informal measure 1 (2)	Informal measure 2 (3)	Informal measure 3 (4)
Formal	0.017***			
	(0.004)			
Informal		-0.061***	-0.008***	-0.060***
		(0.013)	(0.002)	(0.012)
Vocational training	0.014**	-0.018	0.035***	0.078***
	(0.006)	(0.057)	(0.004)	(0.014)
Associate degree	0.116***	-0.132	0.132***	0.185***
	(0.008)	(0.080)	(0.005)	(0.024)
Bachelor's degree	0.417***		0.443***	
	(0.034)		(0.050)	
Other educations	0.099***		0.034**	0.095***
	(0.032)		(0.014)	(0.036)
1-2 years	0.109***	0.114*	0.071***	0.095***
	(0.004)	(0.067)	(0.003)	(0.011)
2-5 years	0.277***	0.483***	0.229***	0.221***
	(0.006)	(0.113)	(0.004)	(0.020)
5-10 years	0.466***		0.445***	0.613***
	(0.015)		(0.009)	(0.107)
10+ years	0.570***		0.520***	
	(0.024)		(0.032)	
Non-manager	0.196***	0.554***	0.202***	0.141***
	(0.014)	(0.187)	(0.009)	(0.024)
Medium manager	0.471***	0.879***	0.447***	0.364***
	(0.016)	(0.214)	(0.010)	(0.048)
Top manager	0.815***		0.838***	
	(0.029)		(0.024)	
Hanoi	0.058***	-0.009	0.061***	0.065***
	(0.005)	(0.090)	(0.003)	(0.010)
HCM	0.070***	0.019	0.056***	0.063***
	(0.006)	(0.094)	(0.003)	(0.011)
Computer	-0.015***	0.049	-0.008***	0.039***
	(0.004)	(0.084)	(0.003)	(0.012)
Software	0.028***	-0.015	0.021***	0.001
	(0.005)	(0.070)	(0.003)	(0.017)
Language	0.029***	0.272***	0.058***	0.033*
	(0.005)	(0.100)	(0.004)	(0.017)
Financial	0.003	0.103	-0.003	-0.029
	(0.006)	(0.100)	(0.004)	(0.028)
People management	0.043***	-0.093	0.034***	-0.137***
	(0.009)	(0.157)	(0.007)	(0.028)
Project management	0.026***	-0.095	0.041***	0.111***
	(0.006)	(0.071)	(0.004)	(0.029)
Art	0.012*	0.083	0.031***	0.072***
	(0.006)	(0.050)	(0.005)	(0.021)
Character	-0.025***	-0.058	-0.022***	-0.048***
	(0.005)	(0.063)	(0.003)	(0.013)

Cognitive	0.012***	0.191**	0.016***	0.007
	(0.004)	(0.093)	(0.003)	(0.013)
Customer service	0.009**	-0.163	0.031***	0.017
	(0.004)	(0.099)	(0.003)	(0.012)
Social	0.014***	-0.083	0.009***	0.007
	(0.004)	(0.072)	(0.003)	(0.014)
Writing	-0.016*	-0.176	0.033***	0.068***
	(0.010)	(0.179)	(0.007)	(0.023)
Obs	50,604	699	93,299	7,680
R2	0.736	0.739	0.711	0.719

Notes: The table presents the regression results of the wage equation on the PSM-matched sample. Columns (1), (2), (3) and (4) show results for the first formality measure, the first, second and third informality measures, respectively. Time, industry, job title, and firm size fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table 4.7. Regression results on CEM-matched sample

	Formal measure 1 (1)	Informal measure 1 (2)	Informal measure 2 (3)	Informal measure 3 (4)
Formal	0.019***			
	(0.004)			
Informal		-0.022	-0.011***	-0.076***
		(0.021)	(0.003)	(0.010)
Vocational training	0.048***	0.169**	0.014**	0.053***
	(0.007)	(0.070)	(0.006)	(0.014)
Associate degree	0.155***	0.007	0.111***	0.214***
	(0.010)	(0.157)	(0.008)	(0.028)
Other educations	0.066		0.205***	
	(0.067)		(0.045)	
1-2 years	0.055***	0.408***	0.066***	0.051***
	(0.005)	(0.132)	(0.004)	(0.012)
2-5 years	0.219***	0.851***	0.243***	0.085**
	(0.007)	(0.203)	(0.005)	(0.036)
5-10 years	0.369***		0.462***	0.400
	(0.027)		(0.018)	(0.274)
10+ years	0.551***		0.352***	
	(0.067)		(0.053)	
Non-manager	0.281***	0.242	0.297***	0.138***
	(0.031)	(0.231)	(0.024)	(0.037)
Medium manager	0.587***	-0.154	0.557***	0.926***
	(0.035)	(0.326)	(0.027)	(0.135)
Top manager	0.829***		1.007***	
	(0.092)		(0.091)	
Hanoi	0.008	0.129	0.036***	0.018
	(0.006)	(0.089)	(0.004)	(0.014)
HCM	0.027***	0.369***	0.046***	0.099***
	(0.006)	(0.136)	(0.004)	(0.014)
Computer	-0.026***	-0.053	-0.015***	0.020*
	(0.005)	(0.047)	(0.003)	(0.011)
Software	0.038***	0.031	0.024***	-0.005
	(0.005)	(0.064)	(0.004)	(0.019)
Language	0.010*	0.078	0.060***	0.035**
	(0.006)	(0.066)	(0.004)	(0.014)
Financial	0.029***	-0.229**	0.020***	-0.010
	(0.006)	(0.093)	(0.005)	(0.022)
People management	0.078***	-0.143	0.015	0.008
	(0.011)	(0.169)	(0.010)	(0.052)
Project management	-0.006	0.098	0.005	-0.079***
	(0.007)	(0.075)	(0.006)	(0.026)
Art	0.012*	0.082	0.022***	0.043**
	(0.007)	(0.055)	(0.006)	(0.021)
Character	-0.023***	-0.061	-0.032***	0.002
	(0.006)	(0.098)	(0.004)	(0.012)
Cognitive	0.012**	-0.002	0.008**	0.001
	(0.005)	(0.053)	(0.004)	(0.011)

Customer service	0.011**	0.198***	0.012***	0.022**
	(0.005)	(0.075)	(0.004)	(0.011)
Social	0.032***	-0.035	0.030***	0.018
	(0.004)	(0.046)	(0.003)	(0.013)
Writing	-0.022**	-0.156	0.021**	0.059**
	(0.010)	(0.168)	(0.008)	(0.024)
Obs	30,208	336	40,458	4,180
R2	0.390	0.773	0.427	0.274

Notes: The table presents the regression results of the wage equation on CEM-matched sample. Columns (1), (2), (3) and (4) show results for the first formality measure, the first, second and third informality measures, respectively. Time, industry, job title, and firm size fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table 4.8. Regression results on text-matched sample

	Formal measure 1 (1)	Informal measure 1 (2)	Informal measure 2 (3)	Informal measure 3 (4)
Formal	0.016*			
	(0.009)			
Informal		-0.214***	-0.020***	-0.062***
		(0.051)	(0.006)	(0.023)
Vocational training	-0.097***	0.589***	0.002	0.083***
	(0.014)	(0.110)	(0.005)	(0.021)
Associate degree	0.026		0.058***	-0.197***
	(0.020)		(0.012)	(0.040)
Bachelor’s degree	-0.101*		0.818***	
	(0.055)		(0.125)	
Other educations	-0.085		0.061**	-0.048
	(0.099)		(0.025)	(0.079)
1-2 years	0.099***	0.227***	0.002	0.055**
	(0.009)	(0.036)	(0.006)	(0.021)
2-5 years	0.323***		0.151***	0.066**
	(0.018)		(0.014)	(0.033)
5-10 years	0.313***		0.336***	0.570**
	(0.046)		(0.034)	(0.227)
10+ years	0.513***		0.697***	
	(0.095)		(0.069)	
Non-manager	0.050*		0.071***	0.114**
	(0.028)		(0.016)	(0.050)
Medium manager	0.381***		0.353***	0.364***
	(0.039)		(0.035)	(0.124)
Top manager	0.559***		0.559***	
	(0.094)		(0.096)	
Hanoi	0.059***	0.445***	0.031***	0.078***
	(0.007)	(0.155)	(0.005)	(0.022)
HCM	0.059***	0.754***	0.028***	0.046***
	(0.011)	(0.157)	(0.005)	(0.016)
Computer	0.068***	0.292***	-0.003	-0.039*
	(0.011)	(0.103)	(0.008)	(0.021)
Software	-0.038*	-0.267*	0.006	0.049
	(0.021)	(0.139)	(0.009)	(0.032)
Language	-0.057***	0.258***	0.048***	0.104*
	(0.016)	(0.098)	(0.011)	(0.055)
Financial	0.040**	-0.359	-0.000	-0.055
	(0.017)	(0.246)	(0.008)	(0.067)
People management	0.042		-0.004	0.135**
	(0.038)		(0.022)	(0.054)
Project management	0.075***		0.036**	0.003
	(0.023)		(0.016)	(0.040)
Art	0.088***	-0.312***	0.021	-0.037
	(0.020)	(0.022)	(0.014)	(0.042)
Character	0.081***	0.364***	-0.042***	0.013
	(0.014)	(0.064)	(0.011)	(0.031)

Cognitive	-0.063***	0.277***	0.006	-0.039
	(0.011)	(0.077)	(0.010)	(0.035)
Customer service	-0.044***		0.016	0.039*
	(0.011)		(0.011)	(0.021)
Social	-0.022**		0.015**	0.011
	(0.011)		(0.007)	(0.024)
Writing	-0.029*		0.026	-0.281***
	(0.017)		(0.029)	(0.091)
Obs	7,598	373	12,038	1,800
R2	0.778	0.885	0.830	0.678

Notes: The table presents the regression results of the wage equation on the text-matched sample. Columns (1), (2), (3) and (4) show results for the first formality measure, the first, second and third informality measures, respectively. Time, industry, job title, and firm size fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table 4.9. Results for the heterogeneous formal-informal sector wage gap by education levels

	Formal measure 1 (1)	Informal measure 1 (2)	Informal measure 2 (3)	Informal measure 3 (4)
Formal	-0.010			
	(0.006)			
Vocational training	-0.022***	-0.024	0.058***	0.075***
	(0.007)	(0.070)	(0.005)	(0.013)
Associate degree	0.105***	-0.131*	0.160***	0.196***
	(0.008)	(0.078)	(0.006)	(0.021)
Bachelor's degree	0.407***		0.539***	
	(0.032)		(0.082)	
Other educations	0.034		0.151***	-0.050
	(0.039)		(0.017)	(0.082)
Formal×Vocational training	0.052***			
	(0.007)			
Formal ×Associate degree	-0.004			
	(0.009)			
Formal ×Bachelor's degree	0.260***			
	(0.054)			
Formal ×Other educations	0.114*			
	(0.066)			
Informal		-0.053***	0.057***	-0.096***
		(0.009)	(0.005)	(0.011)
Informal×Vocational training		0.016	-0.062***	-0.008
		(0.065)	(0.005)	(0.015)
Informal ×Associate degree		0.034	-0.081***	-0.081***
		(0.092)	(0.007)	(0.026)
Informal ×Bachelor's degree			0.001	
			(0.095)	
Informal ×Other educations			-0.092***	0.199*
			(0.028)	(0.103)
1-2 years	0.093***	0.117*	0.071***	0.065***
	(0.004)	(0.069)	(0.002)	(0.009)
2-5 years	0.274***	0.452***	0.247***	0.208***
	(0.005)	(0.117)	(0.003)	(0.018)
5-10 years	0.479***		0.497***	0.437***
	(0.013)		(0.008)	(0.078)
10+ years	0.520***		0.576***	
	(0.024)		(0.039)	
Non-manager	0.216***	0.599***	0.195***	0.161***
	(0.014)	(0.222)	(0.009)	(0.022)
Medium manager	0.482***	0.966***	0.443***	0.649***
	(0.016)	(0.246)	(0.010)	(0.054)
Top manager	0.855***		0.806***	
	(0.029)		(0.023)	
Hanoi	0.007*	-0.031	0.034***	0.037***
	(0.004)	(0.100)	(0.003)	(0.010)
HCM	0.031***	0.008	0.056***	0.087***
	(0.004)	(0.100)	(0.002)	(0.011)
Computer	-0.038***	0.040	-0.020***	0.013
	(0.004)	(0.079)	(0.002)	(0.009)
Software	0.008*	-0.001	0.036***	-0.021
	(0.004)	(0.071)	(0.003)	(0.015)
Language	0.037***	0.293***	0.080***	0.021**

	(0.004)	(0.100)	(0.003)	(0.009)
Financial	0.029***	0.140	0.015***	-0.010
	(0.005)	(0.098)	(0.004)	(0.023)
People management	0.049***	-0.070	0.030***	-0.108***
	(0.009)	(0.171)	(0.007)	(0.027)
Project management	-0.001	-0.098	0.023***	0.047**
	(0.006)	(0.073)	(0.004)	(0.022)
Art	0.019***	0.116**	0.039***	0.025
	(0.006)	(0.049)	(0.004)	(0.017)
Character	-0.060***	-0.050	-0.052***	-0.010
	(0.005)	(0.067)	(0.003)	(0.011)
Cognitive	0.013***	0.201**	0.012***	0.047***
	(0.004)	(0.099)	(0.003)	(0.009)
Customer service	0.020***	-0.166	0.020***	0.004
	(0.004)	(0.103)	(0.003)	(0.008)
Social	0.027***	-0.089	0.020***	-0.001
	(0.003)	(0.073)	(0.002)	(0.011)
Writing	-0.012	-0.223	0.050***	-0.014
	(0.010)	(0.189)	(0.006)	(0.017)
Obs	51,764	700	94,578	8,147
R2	0.485	0.722	0.513	0.372

Notes: The table presents the heterogeneous formal-informal employment wage gap across different education levels. Columns (1), (2), (3) and (4) show results for the first formality measure, the first, second and third informality measures, respectively. The base category is High school. Other control variables and fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table 4.10. Results for the heterogeneous formal-informal sector wage gap by experience levels

	Formal measure 1 (1)	Informal measure 1 (2)	Informal measure 2 (3)	Informal measure 3 (4)
Formal	0.042***			
	(0.004)			
1-2 years	0.101***	0.020	0.096***	0.084***
	(0.005)	(0.084)	(0.003)	(0.012)
2-5 years	0.292***	0.401***	0.274***	0.314***
	(0.006)	(0.104)	(0.004)	(0.021)
5-10 years	0.548***		0.561***	0.582***
	(0.017)		(0.011)	(0.099)
10+ years	0.601***		0.639***	
	(0.034)		(0.064)	
Formal×1-2 years	-0.018***			
	(0.006)			
Formal ×2-5 years	-0.040***			
	(0.008)			
Formal ×5-10 years	-0.160***			
	(0.023)			
Formal ×10+ years	-0.223***			
	(0.038)			
Informal		-0.073***	0.031***	-0.075***
		(0.010)	(0.003)	(0.008)
Informal×1-2 years		0.272***	-0.047***	-0.046***
		(0.091)	(0.004)	(0.015)
Informal ×2-5 years		0.117	-0.053***	-0.259***
		(0.084)	(0.006)	(0.034)
Informal ×5-10 years			-0.132***	-0.280*
			(0.014)	(0.154)
Informal ×10+ years			-0.128*	
			(0.074)	
Vocational training	0.006	0.003	0.021***	0.074***
	(0.005)	(0.055)	(0.004)	(0.010)
Associate degree	0.107***	-0.082	0.113***	0.161***
	(0.007)	(0.085)	(0.005)	(0.018)
Bachelor’s degree	0.471***		0.546***	
	(0.029)		(0.053)	
Other educations	0.085***		0.103***	0.005
	(0.032)		(0.014)	(0.062)
Non-manager	0.217***	0.586***	0.194***	0.156***
	(0.014)	(0.224)	(0.009)	(0.022)
Medium manager	0.483***	0.957***	0.443***	0.648***
	(0.016)	(0.251)	(0.010)	(0.053)
Top manager	0.862***		0.807***	
	(0.029)		(0.023)	
Hanoi	0.009**	-0.045	0.033***	0.035***
	(0.004)	(0.099)	(0.003)	(0.010)
HCM	0.033***	-0.006	0.057***	0.085***
	(0.004)	(0.100)	(0.002)	(0.011)
Computer	-0.037***	0.054	-0.019***	0.013
	(0.004)	(0.077)	(0.002)	(0.009)
Software	0.008*	0.010	0.035***	-0.009
	(0.004)	(0.071)	(0.003)	(0.014)
Language	0.036***	0.278***	0.080***	0.012

	(0.004)	(0.095)	(0.003)	(0.009)
Financial	0.029***	0.167	0.016***	-0.004
	(0.005)	(0.105)	(0.004)	(0.023)
People management	0.050***	-0.086	0.030***	-0.116***
	(0.009)	(0.182)	(0.007)	(0.028)
Project management	-0.002	-0.062	0.023***	0.037*
	(0.006)	(0.068)	(0.004)	(0.022)
Art	0.021***	0.116**	0.037***	0.014
	(0.006)	(0.048)	(0.004)	(0.017)
Character	-0.060***	-0.040	-0.052***	-0.004
	(0.005)	(0.065)	(0.003)	(0.011)
Cognitive	0.013***	0.189**	0.012***	0.050***
	(0.004)	(0.095)	(0.003)	(0.008)
Customer service	0.020***	-0.159	0.021***	0.000
	(0.004)	(0.100)	(0.003)	(0.008)
Social	0.027***	-0.063	0.020***	0.004
	(0.003)	(0.073)	(0.002)	(0.011)
Writing	-0.013	-0.208	0.050***	-0.015
	(0.010)	(0.199)	(0.006)	(0.017)
Obs	51,764	700	94,578	8,147
R2	0.485	0.734	0.513	0.380

Notes: The table presents the heterogeneous formal-informal employment wage gap across different experience levels. Columns (1), (2), (3) and (4) show results for the first formality measure, the first, second and third informality measures, respectively. The base category is 0-1 year of experience. Other control variables and fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table 4.11. Results for the heterogeneous formal-informal sector wage gap by job positions

	Formal measure 1 (1)	Informal measure 1 (2)	Informal measure 2 (3)	Informal measure 3 (4)
Formal	0.220***			
	(0.027)			
Non-manager	0.321***	0.013	0.316***	0.212***
	(0.022)	(0.195)	(0.014)	(0.029)
Medium manager	0.562***	0.361	0.575***	0.751***
	(0.024)	(0.241)	(0.016)	(0.059)
Top manager	0.951***		0.988***	
	(0.037)		(0.030)	
Formal×Non-manager	-0.210***			
	(0.027)			
Formal×Medium manager	-0.153***			
	(0.028)			
Formal×Top manager	-0.181***			
	(0.045)			
Informal		-1.029***	0.224***	0.024
		(0.304)	(0.017)	(0.042)
Informal×Non-manager		0.983***	-0.230***	-0.132***
		(0.305)	(0.017)	(0.043)
Informal×Medium manager		1.084***	-0.254***	-0.269***
		(0.326)	(0.019)	(0.053)
Informal×Top manager			-0.358***	
			(0.034)	
Vocational training	0.008	-0.015	0.020***	0.075***
	(0.005)	(0.056)	(0.004)	(0.010)
Associate degree	0.108***	-0.120	0.113***	0.167***
	(0.007)	(0.081)	(0.005)	(0.018)
Bachelor’s degree	0.504***		0.542***	
	(0.028)		(0.053)	
Other educations	0.088***		0.104***	0.010
	(0.032)		(0.014)	(0.061)
1-2 years	0.092***	0.089	0.071***	0.064***
	(0.004)	(0.071)	(0.002)	(0.008)
2-5 years	0.273***	0.450***	0.247***	0.225***
	(0.005)	(0.118)	(0.003)	(0.019)
5-10 years	0.478***		0.497***	0.451***
	(0.013)		(0.008)	(0.077)
10+ years	0.504***		0.568***	
	(0.023)		(0.038)	
Hanoi	0.007*	-0.025	0.034***	0.035***
	(0.004)	(0.104)	(0.003)	(0.010)
HCM	0.031***	0.024	0.057***	0.089***
	(0.004)	(0.102)	(0.002)	(0.011)
Computer	-0.038***	0.023	-0.019***	0.014
	(0.004)	(0.079)	(0.002)	(0.009)
Software	0.008*	0.010	0.036***	-0.015
	(0.004)	(0.069)	(0.003)	(0.015)
Language	0.037***	0.294***	0.082***	0.018*
	(0.004)	(0.099)	(0.003)	(0.009)
Financial	0.030***	0.152	0.015***	-0.017
	(0.005)	(0.103)	(0.004)	(0.023)
People management	0.047***	-0.035	0.028***	-0.110***

	(0.009)	(0.161)	(0.007)	(0.027)
Project management	-0.002	-0.089	0.024***	0.037*
	(0.006)	(0.069)	(0.004)	(0.021)
Art	0.018***	0.094**	0.037***	0.022
	(0.006)	(0.046)	(0.004)	(0.017)
Character	-0.061***	-0.085	-0.052***	-0.007
	(0.005)	(0.068)	(0.003)	(0.011)
Cognitive	0.014***	0.181*	0.012***	0.046***
	(0.004)	(0.100)	(0.003)	(0.009)
Customer service	0.021***	-0.172*	0.021***	0.003
	(0.004)	(0.100)	(0.003)	(0.008)
Social	0.026***	-0.074	0.020***	0.001
	(0.003)	(0.076)	(0.002)	(0.011)
Writing	-0.014	-0.235	0.051***	-0.022
	(0.010)	(0.186)	(0.006)	(0.017)
Obs	51,764	700	94,578	8,147
R2	0.485	0.727	0.515	0.374

Notes: The table presents the heterogeneous formal-informal employment wage gap across different job positions. Columns (1), (2), (3) and (4) show results for the first formality measure, the first, second and third informality measures, respectively. The base category is entry level. Other control variables and fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Online Appendix

Table A4.1. List of job titles/occupations

Accountant	Database administrator	Marketing	R&D manager
Accounting advisor	Deliveryman	Marketing director	R&D staff
Accounting manager	Deputy	Marketing manager	Real estate businessman
Actor	Deputy manager	Mathematics teacher	Real estate consultant
Admin	Designer	MC	Receptionist
Admin manager	Developer	Mechanics	Restaurant manager
Animator	Digital marketing instructor	Merchandiser	Sale assistant
Appraiser	Director	Network admin	Sales
Architect	Doctor	EHS staff	Sales consultant
Art teacher	Driver	Nurse	Sales manager
AutoCAD senior designer	Editor	Nursery teacher	Sap consultant
Banker	Engineer	Nutritionist	School vice president
Bartender	English interpreter	Office supporter	Scrum master
Beauty teacher	English teacher	Officer	Secretary
Beauty technician	Equipment deployment staff	Online supporter	Security
Bodyguard	Event organiser	Operator	Senior developer
Branch manager	Financial advisor	Others	Senior merchandiser
Bridge system engineer	Financial investment	Partnership manager	Shop manager
Broker	Ga staff	Partnership staff	Spa advisor
Business analyst	Gymnastics teacher	Personal trainer	Stylist
Business development	Health insurance staff	Pharmacist	Supply chain staff
Business representative	Helpdesk	Photographer	Support staff
Business support staff	Housing resource staff	Photoshop	Swimming pool lifeguard
Businessman	HR director	Planning staff	Teaching assistant
CAD/CAM	HR manager	Planning supervisor	Technical consultant
CAD/CAM senior	HR staff	PR	Technician
Cashier	Import-export staff	PR executive	Technology staff
Chef	Information security	Primary school teacher	Tender
Chess game instructor	Innovation staff	Procedure controller	Tester

Chinese interpreter	Internal auditor	Producer	Tour guide
Cleaner	It staff	Product advisor	Tour leader
Clearance specialist	It system administrator	Product specialist	Tour operator
Compliance officer	It teacher	Product/content reviewer	Training specialist
Comtor	Japanese language staff	Production management	Vietnamese teacher
Content creator	Japanese teacher	Production management	Visa/study abroad advisor
Controller/supervisor	Japanese translator	Production statistics staff	Waiter
Cook	Korean interpreter	Project development staff	Warehouse staff
Costing staff	Legal experts	Project manager	Web developer
Customer service director	Librarian	Purchasing staff	Web management
Customer service manager	Live streamer	QA manager	Worker
Customer service staff	Logistics staff	QA staff	Yoga instructor
Customer support	Manager	QC staff	
Data analyst	Market consultant	QC team leader	
Data scientist	Market research analyst	Quality inspector	

Notes: The table presents definitions of main variables used in our analysis.

Table A4.2. Definitions of main variables

Variable	Definition
High school	A dummy variable that takes the value of 1 if the job ad requires a high school qualification, and 0 otherwise.
Vocational training	A dummy variable that takes the value of 1 if the job ad requires a vocational training qualification, and 0 otherwise.
Associate degree	A dummy variable that takes the value of 1 if the job ad requires an associate degree qualification, and 0 otherwise.
Bachelor’s degree	A dummy variable that takes the value of 1 if the job ad requires a bachelor’s degree qualification, and 0 otherwise.
Other educations	A dummy variable that takes the value of 1 if the job ad requires other educational qualifications, and 0 otherwise.
0-1 year	A dummy variable that takes the value of 1 if the job ad requires less than one year of experience, and 0 otherwise.
1-2 years	A dummy variable that takes the value of 1 if the job ad requires 1-2 years of experience, and 0 otherwise.
2-5 years	A dummy variable that takes the value of 1 if the job ad requires 2-5 years of experience, and 0 otherwise.
5-10 years	A dummy variable that takes the value of 1 if the job ad requires 5-10 years of experience, and 0 otherwise.
10+ years	A dummy variable that takes the value of 1 if the job ad requires more than 10 years of experience, and 0 otherwise.
Entry level/Internship	A dummy variable that takes the value of 1 if the job ad is at an entry level/internship position, and 0 otherwise.
Non-manager	A dummy variable that takes the value of 1 if the job ad is at a non-managerial position, and 0 otherwise.
Medium manager	A dummy variable that takes the value of 1 if the job ad is at a medium-manager position, and 0 otherwise.
Top manager	A dummy variable that takes the value of 1 if the job ad is at a top-manager position, and 0 otherwise.
Hanoi	A dummy variable that takes the value of 1 if the job ad locates in Hanoi, and 0 otherwise.
HCM	A dummy variable that takes the value of 1 if the job ad locates in HCM city, and 0 otherwise.
Other locations	A dummy variable that takes the value of 1 if the job ad locates in other cities/provinces, and 0 otherwise.
Computer	A dummy variable that takes the value of 1 if the job ad requires computer skill, and 0 otherwise.
Software	A dummy variable that takes the value of 1 if the job ad requires software skill, and 0 otherwise.
Language	A dummy variable that takes the value of 1 if the job ad requires language skill, and 0 otherwise.
Financial	A dummy variable that takes the value of 1 if the job ad requires financial skill, and 0 otherwise.
People management	A dummy variable that takes the value of 1 if the job ad requires people management skill, and 0 otherwise.
Project management	A dummy variable that takes the value of 1 if the job ad requires project management skill, and 0 otherwise.
Art	A dummy variable that takes the value of 1 if the job ad requires artistic skill, and 0 otherwise.
Character	A dummy variable that takes the value of 1 if the job ad requires character skill, and 0 otherwise.

Cognitive	A dummy variable that takes the value of 1 if the job ad requires cognitive skill, and 0 otherwise.
Customer service	A dummy variable that takes the value of 1 if the job ad requires customer service skill, and 0 otherwise.
Social	A dummy variable that takes the value of 1 if the job ad requires social skill, and 0 otherwise.
Writing	A dummy variable that takes the value of 1 if the job ad requires writing skill, and 0 otherwise.

Notes: The table presents definitions of the main variables used in our analysis.

Table A4.3. Robustness test on PSM-matched sample using average salary measure

	Formal measure 1 (1)	Informal measure 1 (2)	Informal measure 2 (3)	Informal measure 3 (4)
Formal	0.013***			
	(0.004)			
Informal		0.295***	-0.023***	-0.031***
		(0.021)	(0.002)	(0.012)
Vocational training	-0.019***	-0.008	0.009**	0.063***
	(0.007)	(0.069)	(0.003)	(0.015)
Associate degree	0.074***	0.055	0.106***	0.135***
	(0.008)	(0.091)	(0.005)	(0.024)
Bachelor's degree	0.304***		0.365***	
	(0.035)		(0.056)	
Other educations	0.063**		-0.010	0.037
	(0.031)		(0.014)	(0.039)
1-2 years	0.085***	0.126*	0.069***	0.079***
	(0.004)	(0.067)	(0.003)	(0.011)
2-5 years	0.259***	0.233**	0.223***	0.201***
	(0.006)	(0.108)	(0.004)	(0.019)
5-10 years	0.431***		0.428***	0.599***
	(0.014)		(0.008)	(0.098)
10+ years	0.501***		0.521***	
	(0.024)		(0.029)	
Non-manager	0.213***	0.550*	0.211***	0.154***
	(0.014)	(0.293)	(0.009)	(0.029)
Medium manager	0.488***	0.877**	0.463***	0.352***
	(0.016)	(0.344)	(0.010)	(0.045)
Top manager	0.834***		0.865***	
	(0.030)		(0.024)	
Hanoi	0.046***	-0.022	0.055***	0.074***
	(0.005)	(0.101)	(0.003)	(0.011)
HCM	0.067***	-0.019	0.052***	0.064***
	(0.006)	(0.106)	(0.003)	(0.012)
Computer	-0.023***	-0.035	-0.016***	0.045***
	(0.004)	(0.067)	(0.003)	(0.012)
Software	0.028***	-0.009	0.025***	-0.003
	(0.005)	(0.079)	(0.003)	(0.017)
Language	0.042***	0.147	0.066***	0.036**
	(0.005)	(0.096)	(0.004)	(0.017)
Financial	-0.009*	0.067	-0.011***	-0.072***
	(0.006)	(0.147)	(0.004)	(0.028)
People management	0.037***	-0.303	0.032***	-0.224***
	(0.009)	(0.202)	(0.007)	(0.033)
Project management	0.025***	0.040	0.040***	0.104***
	(0.005)	(0.084)	(0.004)	(0.029)
Art	0.009	-0.132	0.035***	0.030
	(0.006)	(0.111)	(0.004)	(0.022)
Character	-0.026***	0.012	-0.023***	-0.025*
	(0.005)	(0.081)	(0.003)	(0.014)
Cognitive	0.015***	0.142*	0.018***	-0.041***
	(0.004)	(0.084)	(0.003)	(0.013)
Customer service	0.001	-0.100	0.035***	0.041***
	(0.004)	(0.079)	(0.003)	(0.012)
Social	0.012***	0.061	0.005**	0.002
	(0.004)	(0.069)	(0.002)	(0.015)

Writing	-0.045***	0.209	0.028***	0.043**
	(0.010)	(0.205)	(0.007)	(0.021)
Obs	50,604	699	93,299	7,680
R2	0.738	0.733	0.723	0.727

Notes: The table presents the regression results of the wage equation on the PSM-matched sample using the average salary measure. Columns (1), (2), (3) and (4) show results for the first formality measure, the first, second and third informality measures, respectively. Time, industry, job title, and firm size fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table A4.4. Robustness test on PSM-matched sample using maximum salary measure

	Formal measure 1 (1)	Informal measure 1 (2)	Informal measure 2 (3)	Informal measure 3 (4)
Formal	0.011**			
	(0.005)			
Informal		0.476***	-0.032***	-0.014
		(0.029)	(0.002)	(0.014)
Vocational training	-0.036***	-0.009	-0.005	0.061***
	(0.008)	(0.092)	(0.004)	(0.019)
Associate degree	0.050***	0.165	0.093***	0.119***
	(0.010)	(0.115)	(0.006)	(0.028)
Bachelor’s degree	0.228***		0.323***	
	(0.039)		(0.062)	
Other educations	0.047		-0.034**	0.015
	(0.033)		(0.016)	(0.049)
1-2 years	0.078***	0.116	0.069***	0.074***
	(0.005)	(0.084)	(0.003)	(0.013)
2-5 years	0.254***	0.104	0.221***	0.188***
	(0.007)	(0.126)	(0.004)	(0.022)
5-10 years	0.416***		0.417***	0.606***
	(0.015)		(0.009)	(0.102)
10+ years	0.464***		0.518***	
	(0.028)		(0.029)	
Non-manager	0.224***	0.525	0.218***	0.156***
	(0.015)	(0.372)	(0.009)	(0.034)
Medium manager	0.496***	0.840*	0.475***	0.341***
	(0.017)	(0.444)	(0.011)	(0.052)
Top manager	0.840***		0.878***	
	(0.033)		(0.024)	
Hanoi	0.039***	-0.015	0.052***	0.072***
	(0.006)	(0.118)	(0.004)	(0.014)
HCM	0.063***	-0.045	0.050***	0.063***
	(0.006)	(0.127)	(0.004)	(0.015)
Computer	-0.027***	-0.058	-0.020***	0.047***
	(0.004)	(0.082)	(0.003)	(0.015)
Software	0.029***	-0.012	0.025***	-0.006
	(0.005)	(0.108)	(0.003)	(0.020)
Language	0.051***	0.078	0.072***	0.037*
	(0.005)	(0.113)	(0.004)	(0.019)
Financial	-0.012*	0.039	-0.012***	-0.093***
	(0.006)	(0.192)	(0.004)	(0.031)
People management	0.036***	-0.432	0.030***	-0.264***
	(0.010)	(0.266)	(0.008)	(0.039)
Project management	0.024***	0.111	0.038***	0.089***
	(0.006)	(0.109)	(0.004)	(0.034)
Art	0.007	-0.257	0.040***	0.015
	(0.007)	(0.158)	(0.005)	(0.025)
Character	-0.024***	0.049	-0.022***	-0.013
	(0.006)	(0.101)	(0.003)	(0.016)
Cognitive	0.016***	0.132	0.020***	-0.070***
	(0.004)	(0.100)	(0.003)	(0.014)
Customer service	-0.003	-0.103	0.038***	0.052***
	(0.005)	(0.102)	(0.004)	(0.014)
Social	0.012***	0.134	0.003	-0.003
	(0.004)	(0.084)	(0.003)	(0.018)

Writing	-0.058***	0.413*	0.026***	0.042*
	(0.012)	(0.245)	(0.007)	(0.024)
Obs	50,604	699	93,299	7,680
R2	0.709	0.743	0.693	0.703

Notes: The table presents the regression results of the wage equation on the PSM-matched sample using the maximum salary measure. Columns (1), (2), (3) and (4) show results for the first formality measure, the first, second and third informality measures, respectively. Time, industry, job title, and firm size fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table A4.5. Robustness test on CEM-matched sample using average salary measure

	Formal measure 1 (1)	Informal measure 1 (2)	Informal measure 2 (3)	Informal measure 3 (4)
Formal	0.030***			
	(0.004)			
Informal		0.301***	-0.032***	-0.056***
		(0.027)	(0.003)	(0.011)
Vocational training	0.017**	0.108	0.005	0.008
	(0.008)	(0.115)	(0.006)	(0.017)
Associate degree	0.122***	0.273	0.099***	0.116***
	(0.011)	(0.185)	(0.008)	(0.028)
Other educations	0.078		0.022	
	(0.061)		(0.044)	
1-2 years	0.040***	0.010	0.066***	0.010
	(0.005)	(0.099)	(0.004)	(0.013)
2-5 years	0.205***	0.542***	0.237***	0.115***
	(0.007)	(0.133)	(0.005)	(0.029)
5-10 years	0.390***		0.449***	0.249
	(0.031)		(0.019)	(0.189)
10+ years	0.339***		0.318***	
	(0.092)		(0.038)	
Non-manager	0.249***	0.313	0.287***	0.156***
	(0.029)	(0.387)	(0.024)	(0.055)
Medium manager	0.541***	0.060	0.522***	0.523***
	(0.032)	(0.426)	(0.026)	(0.078)
Top manager	0.885***		1.110***	
	(0.094)		(0.104)	
Hanoi	0.005	0.213	0.037***	-0.005
	(0.006)	(0.148)	(0.004)	(0.015)
HCM	0.013**	0.385**	0.056***	0.060***
	(0.006)	(0.188)	(0.004)	(0.016)
Computer	-0.052***	-0.207***	-0.036***	-0.013
	(0.005)	(0.062)	(0.003)	(0.012)
Software	0.048***	-0.017	0.041***	0.039*
	(0.005)	(0.051)	(0.004)	(0.021)
Language	0.002	0.004	0.058***	0.016
	(0.006)	(0.072)	(0.004)	(0.014)
Financial	0.032***	-0.300**	0.028***	-0.043**
	(0.006)	(0.124)	(0.005)	(0.019)
People management	0.050***	0.022	0.015	-0.025
	(0.011)	(0.183)	(0.011)	(0.048)
Project management	-0.024***	0.042	-0.004	0.023
	(0.007)	(0.084)	(0.006)	(0.025)
Art	0.017**	-0.170*	0.029***	0.006
	(0.007)	(0.095)	(0.006)	(0.023)
Character	-0.036***	-0.132	-0.036***	0.020
	(0.006)	(0.085)	(0.004)	(0.014)
Cognitive	0.038***	0.177***	0.014***	-0.047***
	(0.005)	(0.067)	(0.004)	(0.012)
Customer service	0.020***	0.196**	0.037***	0.070***
	(0.005)	(0.095)	(0.004)	(0.012)
Social	0.023***	0.084	0.023***	0.034**
	(0.005)	(0.054)	(0.004)	(0.013)
Writing	-0.021*	0.166	0.012	0.070**
	(0.011)	(0.148)	(0.008)	(0.032)

Obs	30,208	336	40,458	4,180
R2	0.398	0.679	0.453	0.313

Notes: The table presents the regression results of the wage equation on the CEM-matched sample using the average salary measure. Columns (1), (2), (3) and (4) show results for the first formality measure, the first, second and third informality measures, respectively. Time, industry, job title, and firm size fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table A4.6. Robustness test on CEM-matched sample using maximum salary measure

	Formal measure 1 (1)	Informal measure 1 (2)	Informal measure 2 (3)	Informal measure 3 (4)
Formal	0.036***			
	(0.004)			
Informal		0.428***	-0.049***	-0.048***
		(0.045)	(0.003)	(0.013)
Vocational training	-0.008	0.207	0.003	-0.013
	(0.010)	(0.188)	(0.008)	(0.021)
Associate degree	0.098***	0.112	0.093***	0.077**
	(0.012)	(0.457)	(0.009)	(0.033)
Other educations	0.111*		-0.105**	
	(0.057)		(0.042)	
1-2 years	0.039***	-0.105	0.069***	-0.007
	(0.006)	(0.109)	(0.004)	(0.015)
2-5 years	0.201***	0.669***	0.236***	0.098***
	(0.008)	(0.252)	(0.006)	(0.031)
5-10 years	0.378***		0.435***	0.122
	(0.032)		(0.020)	(0.170)
10+ years	0.239**		0.306***	
	(0.110)		(0.034)	
Non-manager	0.233***	0.482	0.293***	0.153**
	(0.031)	(0.581)	(0.025)	(0.065)
Medium manager	0.508***	0.550	0.538***	0.422***
	(0.034)	(0.695)	(0.028)	(0.086)
Top manager	0.866***		1.170***	
	(0.083)		(0.109)	
Hanoi	0.009	0.487***	0.043***	0.001
	(0.007)	(0.156)	(0.005)	(0.019)
HCM	0.009	0.389*	0.063***	0.059***
	(0.007)	(0.217)	(0.005)	(0.019)
Computer	-0.069***	-0.320***	-0.050***	-0.028*
	(0.005)	(0.076)	(0.004)	(0.014)
Software	0.047***	-0.015	0.047***	0.073***
	(0.006)	(0.102)	(0.005)	(0.026)
Language	0.001	-0.045	0.057***	0.025
	(0.006)	(0.077)	(0.005)	(0.018)
Financial	0.038***	-0.177	0.028***	-0.085***
	(0.007)	(0.257)	(0.006)	(0.023)
People management	0.033***	0.310**	0.006	-0.015
	(0.012)	(0.155)	(0.012)	(0.051)
Project management	-0.022***	0.156	-0.015**	0.062**
	(0.008)	(0.169)	(0.006)	(0.031)
Art	0.033***	-0.308***	0.039***	-0.017
	(0.008)	(0.109)	(0.007)	(0.030)
Character	-0.039***	-0.281	-0.039***	0.039**
	(0.006)	(0.196)	(0.005)	(0.016)
Cognitive	0.048***	0.166*	0.022***	-0.062***
	(0.006)	(0.092)	(0.004)	(0.014)
Customer service	0.024***	0.236**	0.050***	0.099***
	(0.006)	(0.114)	(0.005)	(0.014)
Social	0.022***	0.278***	0.023***	0.038**
	(0.005)	(0.068)	(0.004)	(0.017)
Writing	-0.027**	0.647***	0.002	0.026
	(0.012)	(0.164)	(0.010)	(0.044)
Obs	30,208	336	40,458	4,180
R2	0.361	0.604	0.407	0.291

Notes: The table presents the regression results of the wage equation on the CEM-matched sample using the maximum salary measure. Columns (1), (2), (3) and (4) show results for the first formality measure, the first, second and third informality measures, respectively. Time, industry, job title, and firm size fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table A4.7. Robustness test on text-matched sample using average salary measure

	Formal measure 1 (1)	Informal measure 1 (2)	Informal measure 2 (3)	Informal measure 3 (4)
Formal	0.003			
	(0.009)			
Informal		-0.241***	-0.020***	-0.052**
		(0.075)	(0.006)	(0.023)
Vocational training	-0.154***	0.491***	-0.016***	-0.020
	(0.017)	(0.127)	(0.005)	(0.030)
Associate degree	-0.039*		0.036***	-0.318***
	(0.022)		(0.011)	(0.046)
Bachelor’s degree	-0.238***		0.692***	
	(0.052)		(0.107)	
Other educations	-0.212		-0.004	-0.185**
	(0.148)		(0.025)	(0.085)
1-2 years	0.072***	0.173***	0.014**	0.056**
	(0.009)	(0.029)	(0.006)	(0.023)
2-5 years	0.300***		0.140***	0.093***
	(0.019)		(0.013)	(0.033)
5-10 years	0.242***		0.346***	0.705***
	(0.048)		(0.031)	(0.248)
10+ years	0.298***		0.611***	
	(0.058)		(0.056)	
Non-manager	0.046		0.082***	0.204**
	(0.033)		(0.018)	(0.094)
Medium manager	0.344***		0.326***	0.424***
	(0.045)		(0.034)	(0.142)
Top manager	0.627***		0.544***	
	(0.073)		(0.087)	
Hanoi	0.042***	0.562***	0.027***	0.087***
	(0.007)	(0.126)	(0.006)	(0.030)
HCM	0.053***	0.649***	0.020***	0.043**
	(0.011)	(0.126)	(0.005)	(0.019)
Computer	0.075***	-0.030	-0.001	0.019
	(0.011)	(0.119)	(0.008)	(0.030)
Software	-0.041**	-0.112	-0.018*	0.033
	(0.021)	(0.172)	(0.010)	(0.034)
Language	-0.030*	0.391***	0.068***	0.127*
	(0.017)	(0.106)	(0.011)	(0.068)
Financial	0.086***	-0.170	-0.023***	-0.120*
	(0.019)	(0.242)	(0.009)	(0.072)
People management	0.059		0.019	0.017
	(0.041)		(0.021)	(0.071)
Project management	0.071***		0.021	-0.021
	(0.026)		(0.013)	(0.052)
Art	0.055***	-0.087***	0.033***	-0.113**
	(0.020)	(0.006)	(0.012)	(0.057)
Character	0.081***	0.173*	-0.031***	0.002
	(0.015)	(0.097)	(0.011)	(0.036)
Cognitive	-0.044***	0.174	0.026***	-0.067*
	(0.012)	(0.120)	(0.009)	(0.036)
Customer service	-0.069***		0.048***	0.073***
	(0.012)		(0.011)	(0.026)
Social	-0.027**		-0.009	-0.080***
	(0.012)		(0.008)	(0.028)

Writing	-0.002		0.065***	-0.429***
	(0.016)		(0.019)	(0.098)
Obs	7,598	373	12,038	1,800
R2	0.773	0.973	0.833	0.643

Notes: The table presents the regression results of the wage equation on the text-matched sample using the average salary measure. Columns (1), (2), (3) and (4) show results for the first formality measure, the first, second and third informality measures, respectively. Time, industry, job title, and firm size fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Table A4.8. Robustness test on text-matched sample using maximum salary measure

	Formal measure 1 (1)	Informal measure 1 (2)	Informal measure 2 (3)	Informal measure 3 (4)
Formal	-0.004			
	(0.010)			
Informal		-0.247***	-0.021***	-0.049*
		(0.095)	(0.007)	(0.028)
Vocational training	-0.180***	0.381**	-0.024***	-0.062
	(0.021)	(0.160)	(0.006)	(0.039)
Associate degree	-0.074***		0.027**	-0.366***
	(0.026)		(0.013)	(0.056)
Bachelor's degree	-0.316***		0.626***	
	(0.059)		(0.100)	
Other educations	-0.270		-0.045	-0.251**
	(0.175)		(0.035)	(0.100)
1-2 years	0.061***	0.124***	0.017**	0.058**
	(0.011)	(0.028)	(0.007)	(0.029)
2-5 years	0.295***		0.135***	0.102**
	(0.021)		(0.014)	(0.041)
5-10 years	0.222***		0.352***	0.800***
	(0.055)		(0.032)	(0.270)
10+ years	0.177***		0.552***	
	(0.057)		(0.055)	
Non-manager	0.037		0.088***	0.244**
	(0.038)		(0.021)	(0.119)
Medium manager	0.311***		0.320***	0.475***
	(0.051)		(0.038)	(0.168)
Top manager	0.655***		0.526***	
	(0.071)		(0.085)	
Hanoi	0.029***	0.613***	0.027***	0.099**
	(0.008)	(0.150)	(0.007)	(0.039)
HCM	0.049***	0.613***	0.016**	0.045*
	(0.013)	(0.150)	(0.007)	(0.024)
Computer	0.082***	-0.251*	-0.003	0.050
	(0.012)	(0.151)	(0.009)	(0.041)
Software	-0.044*	0.031	-0.033***	0.034
	(0.023)	(0.218)	(0.011)	(0.042)
Language	-0.017	0.492***	0.080***	0.127
	(0.019)	(0.133)	(0.013)	(0.081)
Financial	0.119***	-0.016	-0.036***	-0.154*
	(0.023)	(0.304)	(0.011)	(0.082)
People management	0.074		0.037	-0.013
	(0.045)		(0.025)	(0.084)
Project management	0.068**		0.014	-0.043
	(0.029)		(0.015)	(0.069)
Art	0.036	0.002	0.045***	-0.151**
	(0.023)	(0.002)	(0.014)	(0.072)
Character	0.080***	0.033	-0.022*	0.001
	(0.019)	(0.126)	(0.012)	(0.046)
Cognitive	-0.039***	0.115	0.033***	-0.086**
	(0.014)	(0.156)	(0.010)	(0.043)
Customer service	-0.082***		0.065***	0.083**
	(0.015)		(0.012)	(0.033)
Social	-0.028**		-0.023**	-0.129***
	(0.014)		(0.009)	(0.037)

Writing	0.023		0.088***	-0.488***
	(0.018)		(0.022)	(0.115)
Obs	7,598	373	12,038	1,800
R2	0.744	0.994	0.787	0.628

Notes: The table presents the regression results of the wage equation on the text-matched sample using the maximum salary measure. Columns (1), (2), (3) and (4) show results for the first formality measure, the first, second and third informality measures, respectively. Time, industry, job title, and firm size fixed effects are included but not reported. Robust standard errors are shown in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Chapter 5. Conclusion

This thesis consists of three empirical studies, which focus on several essential determinants of using online job vacancy data. In the first study, we examine the effect of physical attractiveness on earning using a rich dataset of online job ads collected from one of the largest job portals in Vietnam from February 2019 to July 2920. To make our vacancies comparable, we use three different matching techniques. First, we use the propensity score matching method to match vacancies on all observable job characteristics, including job title, location, education level, experience level, job position, skills required and firm size. Second, we use coarsened exact matching to match vacancies based on job title, location, education level, experience level, job position and firm size. Finally, we adopt a text-based matching approach to match job ads based on the textual content of the vacancies.

Our estimation results reveal that physical appearance has different impacts on the offered wage across men-target and women-target ads. In particular, among women-target postings, jobs that feature a beauty requirement offer, on average 3.3% to 5.2% higher salary than those that do not. In contrast, among men-target postings, jobs that feature a beauty requirement do not offer a higher wage or even offer 5.2% lower wages than those that do not. These results suggest that physical appearance is more important for women than men. Such differences might stem from the differences in traditional roles between women and men or differences in the workplace power between the two genders. In addition, we find evidence of wage premia for beautiful women not only among jobs involving a high level of social/customer interactions where beauty is expected to enhance the job performance but also among other occupations where beauty is not.

Furthermore, the wage premium varies across different education, experience levels and job positions. For jobs requesting the lowest education category and entry-level jobs, beauty can bring an earning premium of approximately four percent for beautiful women. Then the higher the job level, the smaller the beauty effect is. Additionally, we show that the wage effect of beauty is heterogeneous across geographical areas. In particular, there exists a beauty premium for women in the largest cities and a beauty penalty for women in smaller towns. There is no significant evidence of beauty premium for male workers in both large cities and small towns.

In brief, this empirical study shows evidence of the beauty premium for good-looking women even after controlling for detailed job characteristics. Drawing from the key findings, we call for policymakers to pay more attention to the issue of discrimination based on physical appearance in the labour market and design effective policies to tackle this issue. Clearly, it can be challenging to estimate the level of discrimination using traditional labour market data sources due to the lack of important labour market dimensions, and the time and cost of data collection. Thus, our study emphasizes the usefulness of online job vacancy data for labour economics studies.

The second empirical study explores the use of foreign currency, the US dollar, in particular, to advertise wages in online job vacancies. Vietnam is a highly dollarised economy, and people prefer the US dollar to the local currency for many reasons, including hedging against the depreciation of local currency (exchange rate risk). In this paper, we investigate the complementarity between monetary compensation and non-monetary compensation (i.e., exchange rate benefit). Further, we examine various heterogeneities across education levels, experience levels and job positions. Finally, we examine the skill requirements and job characteristics of dollarised jobs.

This study contributes to several strands of literature. First, to the best of our knowledge, our paper is the first to examine the wage dollarisation phenomenon in the online labour market, which is the use of the US dollar in quoting wages. Second, we provide additional evidence on the complementarity between wage and non-wage benefits using a new type of benefits – exchange rate benefits – that changes over time depending on exchange rate volatilities. Third, setting wages in foreign currency can be seen as similar to the hidden wage strategy. Fourth, we contribute to the growing literature by employing online vacancies data to study the labour market dynamics by linking the labour market dynamics with a macroeconomic condition, which is the payment dollarisation in our study.

Our empirical analysis results in three main findings. First, using different matching models (i.e., propensity score matching, coarsened exact matching and text matching), we find evidence of a positive relationship between monetary and non-monetary compensations. More specifically, vacancies advertising wages in US dollars can offer both a hedge against exchange rate risks and 40% to 46% higher wages than other jobs. The finding might be explained by the efficiency wage theory, which states that firms are willing to pay both higher wages and higher benefits to motivate workers, hence increasing their productivity. We also rule out an alternative explanation that dollarised jobs are advertised by foreign firms, which are more likely to pay a higher wage than domestic firms, by running the regressions for job postings by foreign firms and domestic firms separately.

Second, we find that while there are no significant differences in the link between wage and exchange rate benefits across education levels, this complementary relationship between wage and non-wage benefits varies across different experience and job levels. Particularly, the positive link between wages and benefits mostly concentrates among jobs at lower positions and requiring less work experience.

Moreover, the paper explores the skill requirements and job characteristics of jobs offering exchange rate benefits and find that setting wage in US dollars is similar to hiding wage to the extent that they both request higher education qualifications, higher work experience and at higher positions. Furthermore, we also document that the wage dollarisation phenomenon concentrates in the two biggest cities in Vietnam. In addition, firms that advertise wages in US dollars tend to demand foreign language, writing and artistic skills.

Last but not least, this empirical research aims to inform policymakers about the extent of wage dollarisation in the labour market in order to tackle the persistent dollarisation phenomenon in the economy. Second, it provides more understanding of a new channel through which firms use to seek highly skilled workers. Lastly, it documents new sources of labour market inequalities. Instead of focusing on explaining wage inequality by the level of human capital or differences in productivity, we focus on the role of the labour demand side – firms – in creating wage and benefits inequalities.

In the third paper of this thesis, we examine the wage gap between formal and informal jobs using online job vacancies data over a seventeen-month period from February 2019 to December 2020. To ensure that our job postings are homogeneous, we use different matching methods, including the PSM, CEM and text-matching, to match them on all job characteristics and requirements, including job title, location, education level, experience level, job position, skills required and firm size.

We rely on four distinct and comprehensive definitions to define formal and informal job postings. We use a legal-based approach to define informal employment: we identify informal jobs as those that do not follow the labour law (showing discrimination based on genders, marital status and disability in the recruiting process). Additionally, we identify formal jobs as those explicitly offer social insurance, security protection, and maternity benefits to employees. Hence, not following the labour law is used as a signal of informality. One of the contributions of this study is providing new measures to define informality based on the signals contained in job ads because informal employment (especially among salaried jobs) is getting more sophisticated and harder to identify.

According to our estimation results, we find a significant wage gap between formal and informal jobs within the salaried work sector. Particularly, we find significant wage premia of 1.6 to 1.9 percent for formal jobs and wage penalties of 0.8 to 21.4 percent for informal ones. These results reveal that informal jobs tend to offer lower wages than their formal counterparts. This finding can be explained by the segmented labour market theory, which argues that informal jobs represent a survival mechanism for employees who cannot enter the formal job sector. Another explanation is the compensating differentials theory, which states that informal sector workers might be paid lower earnings in exchange for the flexibility in working hours, place of work or for the reason of tax and other social security contributions avoidance.

In addition, the study documents that the differentials in wages between formal and informal jobs vary across different education, experience levels and job positions. That is, the formal-informal employment wage gap is small among low-education jobs and gets larger as the required education increases. The wage gap is also considerably heterogeneous across job ads of different work experience and job positions.

The results of this study yield several policy implications. First, it aims to inform policymakers about the degree of informality in the online labour market and the wage gap between formal and informal jobs. The government might need to introduce a stricter legal framework and impose a higher level of enforcement to reduce informal sector employment and protect workers in this sector. Yet, one might encounter numerous challenges when estimating the formal-informal sector wage differentials using existing labour market data, which might lack various labour market dimensions and might not be promptly and regularly updated. Hence, this study highlights the advantages of using online job vacancy data as a wealthy and real-time data source for prompt analyses of the labour market dynamics.

Technical Appendix

This section includes a detailed technical appendix to accompany the methodology sections of the three empirical studies in this thesis, which describes more details of the matching techniques employed throughout the thesis.

Matching is a class of techniques of pre-processing data to control for some potentially confounding biases by reducing the covariate imbalances between the treated and control groups. Matching is not a method of estimation but a way to pre-process data so that estimations of the average treatment effect based on the matched sample will be less model-dependent than when estimations are based on the full sample. Matching involves dropping observations that do not have close matches on pre-treatment covariates in both the treated and control groups.

There are several matching techniques for causal inference, among which Propensity Score Matching and Coarsened Exact Matching are the two most popular techniques. First, Propensity Score Matching (Rosenbaum and Rubin, 1983) allows us to construct an artificial control group by matching each treated unit with a non-treated unit of similar characteristics based on a single score – propensity score.

The matching process first involve estimating the propensity score using either Logit or Probit model. In particular, the propensity score represents the probability of being assigned in the treatment group. The propensity score can be used to decrease or eliminate the selection bias by balancing covariates (the characteristics of units) between treated and control groups. Based on this propensity score, units in the treated group are matched with those in the control group of the same propensity score or within a small range of it.

While Propensity Score Matching (PSM) is a widely used and well-established technique in the literature, Coarsened Exact Matching (Iacus et al., 2012) is a relatively new technique that allows us to non-parametrically create a matched sample to estimate the treatment effect. In general, the CEM algorithm involves the following three steps. Firstly, CEM temporarily coarsen each covariate X into different bins. For example, the number of years of education can be coarsened into categories such as high school, associate degree, bachelor's degree, master's degree and higher. Another example is an individual's age, which can be coarsened into 0-10, 10-20, 20-40, 40-60 and 60+ age bins. Secondly, CEM sorts all observations in treated and control groups into strata, each of which has the same values of the coarsened covariate X. Finally, CEM discards the units in stratum that do not contain at least one treated and one control unit.

If the coarsened bins have zero width, then CEM will return the exact matching solution. However, there might be too few observations that remain in the final matched sample. In contrast, if the coarsened bins are too wide, then only few observations will be discarded, but the differences in values/characteristics of units within the same strata can be too large to be considered as similar. In this thesis, since all the matching covariates are categorical variables, our CEM algorithms are equivalent to exact matching and yields matched samples that have fewer observations than our PSM algorithms.

Since our data are text data, we also rely another matching technique, which is text matching. Text matching is increasingly employed in recent research papers as it can outperform other matching methods in performing tasks related to the text (Roberts et al., 2020). For instance, text matching is applied in comparing the technological similarity between patents (Arts et al., 2018), matching job seekers’ CVs and vacancies (Chala et al., 2018) or measuring product similarity based on the product’s description text (Hoberg and Phillips, 2016).

Throughout three studies in the thesis, PSM is our preferred matching technique and all further analyses are conducted on the propensity score matched samples. The reason for choosing PSM over CEM and text matching methods is that this technique yields the highest number of observations in the matched sample.

References

Abueg, L. C., Hubilla, P. G. O., Lozano, F. M., Valdivieso, P. I. L., & Valencia, M. M. S. (2020). Penalties and some counterfactuals to beauty premium: Evidence from a job search simulation experiment. *DLSU Business & Economics Review*, *30*(1), 15-29.

Acemoglu, D., Autor, D., Hazell, J., & Restrepo, P. (2020). AI and jobs: Evidence from online vacancies. *National Bureau of Economic Research*. No. w28257.

Adair, P. (2021). Non-observed economy vs. shadow economy and informal employment in Poland: A range of mismatching estimates. *Comparative Economic Studies in Europe*, 249-278. Palgrave Macmillan, Cham.

Adams-Prassl, A., Balgova, M., & Qian, M. (2020). Flexible work arrangements in low wage jobs: Evidence from job vacancy data.

Adhikari, D. B. (2011). Income generation in informal sector: A case study of the street vendors of Kathmandu Metropolitan City. *Economic Journal of Development Issues*, 1-14.

Adrjan, P., & Lydon, R. (2019). Clicks and jobs: Measuring labour market tightness using online data. *Central Bank of Ireland Economic Letter Series*, *2019*(6).

Aiba, D., & Tha, R. (2017). Dollarization and enterprise's behaviors: The case of Cambodia. *Hitotsubashi Economics, 10*(2), 49-80.

Alekseeva, L., Azar, J., Gine, M., Samila, S., & Taska, B. (2021). The demand for AI skills in the labor market. *Labour Economics*, *71*, 102002.

Altonji, J. G., & Usui, E. (2007). Work hours, wages, and vacation leave. *ILR Review*, *60*(3), 408-428.

Anastasi, A. (1981). Sex differences: Historical perspectives and methodological implications. *Developmental Review*, *1*(3), 187-206.

Andreoni, J., & Petrie, R. (2008). Beauty, gender and stereotypes: Evidence from laboratory experiments. *Journal of Economic Psychology*, *29*(1), 73-93.

Angel-Urdinola, D. F., & Tanabe, K. (2012). Micro-determinants of informal employment in the Middle East and North Africa region. *World Bank.*

Arceo-Gomez, E. O., & Campos-Vazquez, R. M. (2019). Double discrimination: Is discrimination in job ads accompanied by discrimination in callbacks?. *Journal of Economics, Race, and Policy*, *2*(4), 257-268.

Arts, S., Cassiman, B., & Gomez, J. C. (2018). Text matching to measure patent similarity. *Strategic Management Journal*, *39*(1), 62-84.

Arunachalam, R., & Shah, M. (2012). The prostitute's allure: The return to beauty in commercial sex work. *The BE Journal of Economic Analysis & Policy*, *12*(1).

Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics*, *10*(2), 150-161.

Autor, D. H., & Handel, M. J. (2013). Putting tasks to the test: Human capital, job tasks, and wages. *Journal of Labor Economics*, *31*(S1), S59-S96.

Azar, J., Marinescu, I., Steinbaum, M., & Taska, B. (2020). Concentration in US labor markets: Evidence from online vacancy data. *Labour Economics*, *66*, 101886.

Babina, T., Fedyk, A., He, A. X., & Hodson, J. (2021). Artificial intelligence, firm growth, and product innovation. *Firm Growth, and Product Innovation*.

Babin, J. J., Hussey, A., Nikolsko-Rzhevskyy, A., & Taylor, D. A. (2020). Beauty premiums among academics. *Economics of Education Review*, *78*, 102019.

Baker-Sperry, L., & Grauerholz, L. (2003). The pervasiveness and persistence of the feminine beauty ideal in children's fairy tales. *Gender and Society*, *17*(5), 711-726.

Banfi, S., & Villena-Roldan, B. (2019). Do high-wage jobs attract more applicants? Directed search evidence from the online labor market. *Journal of Labor Economics*, *37*(3), 715-746.

Bargain, O., & Kwenda, P. (2014). The informal sector wage gap: New evidence using quantile estimations on panel data. *Economic Development and Cultural Change, 63*(1), 117-153.

Bar-Tal, D., & Saxe, L. (1976). Physical attractiveness and its relationship to sex-role stereotyping. *Sex Roles*, *2*(2), 123-133.

Baughman, R., DiNardi, D., & Holtz-Eakin, D. (2003). Productivity and wage effects of “family-friendly” fringe benefits. *International Journal of Manpower*.

Benjamin, N. C., & Mbaye, A. A. (2012). The informal sector, productivity, and enforcement in west Africa: A firm-level analysis. *Review of Development Economics*, 16(4), 664-680.

Berggren, N., Jordahl, H., & Poutvaara, P. (2010). The looks of a winner: Beauty and electoral success. *Journal of Public Economics*, *94*(1-2), 8-15.

Berri, D. J., Simmons, R., Van Gilder, J., & O'Neill, L. (2011). What does it mean to find the face of the franchise? Physical attractiveness and the evaluation of athletic performance. *Economics Letters*, *111*(3), 200-202.

Bertrand, M., & Mullainathan, S. (2001). Do people mean what they say? Implications for subjective survey data. *American Economic Review*, *91*(2), 67-72.

Biddle, J. E., & Hamermesh, D. S. (1998). Beauty, productivity, and discrimination: Lawyers' looks and lucre. *Journal of Labor Economics*, *16*(1), 172-201.

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, 436-455.

Borjas, G. J., & Fisher, E. O. N. (2001). Dollarization and the Mexican labor market. *Journal of Money, Credit and Banking*, 626-647.

Borland, J., & Leigh, A. (2014). Unpacking the beauty premium: What channels does it operate through, and has it changed over time?. *Economic Record*, *90*(288), 17-32.

Borooah, V. K. (2019). Wage inequality. *Disparity and Discrimination in Labour Market Outcomes in India*, 163-195. Palgrave Macmillan, Cham.

Bozoyan, C., & Wolbring, T. (2011). Fat, muscles, and wages. *Economics and Human Biology*, *9*(4), 356-363.

Braithwaite, J. (1995). From second economy to informal sector: The Russian labor market in transition. *The World Bank*.

Braun, S., Peus, C., & Frey, D. (2015). Is beauty beastly?. *Zeitschrift für Psychologie*.

Brenčič, V. (2012). Wage posting: Evidence from job ads. *Canadian Journal of Economics/Revue Canadienne d'Économique*, *45*(4), 1529-1559.

Brenčič, V., & Norris, J. B. (2012). Employers' on-line recruitment and screening practices. *Economic Inquiry*, *50*(1), 94-111.

Brenzel, H., Gartner, H., & Schnabel, C. (2014). Wage bargaining or wage posting? Evidence from the employers' side. *Labour Economics*, *29*, 41-48.

Brown, C. (1980). Equalizing differences in the labor market. *The Quarterly Journal of Economics*, *94*(1), 113-134.

Bryson, A., Dorsett, R., & Purdon, S. (2002). The use of propensity score matching in the evaluation of active labour market policies.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31-72.

Calvo, M. G., & Gramont, M. C. A. V. (1992). Currency substitution in developing countries: An introduction.

Carneiro, F., & Henley, A. (2001). Modelling formal vs. informal employment and earnings: Micro-econometric evidence for Brazil. *U of Wales at Aberystwyth Management & Business Working Paper*, (2001-15).

Carnevale, A. P., & Rose, S. J. (2014). The undereducated American.

Castillo, P. (2006). Optimal monetary policy and endogenous price dollarization. *Banco Central de Reserva del Perú*.

Cellerino, A. (2003). Psychobiology of facial attractiveness. *Journal of Endocrinological investigation*, *26*(3; SUPP), 45-48.

Chaker, N. N., Walker, D., Nowlin, E. L., & Anaza, N. A. (2019). When and how does sales manager physical attractiveness impact credibility: A test of two competing hypotheses. *Journal of Business Research*, *105*, 98-108.

Chala, S. A., Ansari, F., Fathi, M., & Tijdens, K. (2018). Semantic matching of job seeker to vacancy: A bidirectional approach. *International Journal of Manpower*.

Chen, M. A. (2012). The informal economy: Definitions, theories and policies. *WIEGO Working Paper*.

Cheng, H., Li, X., Dong, Y., & Qi, S. (2020). Competition and favoritism in bank loan markets. *Pacific-Basin Finance Journal*, *59*, 101214.

Chidakwa, A. M., & Munhupedzi, R. N. (2017). Investigating the impact of dollarization on economic growth: A case of Zimbabwe. *Expert Journal of Finance*, *5*(1).

Chowdhury, A. R., Areias, A. C., Imaizumi, S., Nomura, S., & Yamauchi, F. (2018). Reflections of employers' gender preferences in job ads in India: An analysis of online job portal data. *The World Bank.*

Christl, M., & Köppl–Turyna, M. (2020). Gender wage gap and the role of skills and tasks: Evidence from the Austrian PIAAC data set. *Applied Economics*, *52*(2), 113-134.

Colclough, C., Kingdon, G., & Patrinos, H. (2010). The changing pattern of wage returns to education and its implications. *Development Policy Review*, *28*(6), 733-747.

Combes, P. P., & Gobillon, L. (2015). The empirics of agglomeration economies. *Handbook of Regional and Urban Economics 5*, 247-348.

Connell, R., 1995. Masculinities. *Sydney: Polity Press*.

Costa Dias, M., Joyce, R., & Parodi, F. (2018). The gender pay gap in the UK: Children and experience in work. *IFS Working Papers*.

Cuadros-Meñaca, A. (2020). Remittances, health insurance, and pension contributions: Evidence from Colombia. *World Development*, 127, 104766.

Dale-Olsen, H. (2006). Wages, fringe benefits and worker turnover. *Labour Economics*, *13*(1), 87-105.

Daza, J. L. (2005). Informal economy, undeclared work and labour administration. *Dialogue*.

De Graaf-Zijl, M. (2012). Job satisfaction and contingent employment. *De Economist*, *160*(2), 197-218.

De Zamaroczy, M., & Sa, S. (2002). Macroeconomic adjustment in a highly dollarized economy: The case of Cambodia. *IMF Working Papers*, *2002*(092).

Deming, D., & Kahn, L. B. (2018). Skill requirements across firms and labor markets: Evidence from job postings for professionals. *Journal of Labor Economics*, *36*(S1), S337-S369.

Deming, D. J., & Noray, K. L. (2018). STEM careers and the changing skill requirements of work. National Bureau of Economic Research. No. w25065.

Deng, W., Li, D., & Zhou, D. (2020). Beauty and job accessibility: new evidence from a field experiment. *Journal of Population Economics*, *33*(4), 1303-1341.

Deryugina, T., & Shurchkov, O. (2015). Now you see it, now you don’t: The vanishing beauty premium. *Journal of Economic Behavior and Organization*, *116*, 331-345.

Di Paolo, A., & Tansel, A. (2015). Returns to foreign language skills in a developing country: The case of Turkey. *The Journal of Development Studies*, *51*(4), 407-421.

Dickens, W., & Lang, K. (1985). Testing dual labor market theory: A reconsideration of the evidence.

Djumalieva, J., & Sleeman, C. (2018). An open and data-driven taxonomy of skills extracted from online job adverts. *Developing Skills in A Changing World of Work: Concepts, Measurement and Data Applied in Regional and Local Labour Market Monitoring Across Europe*, *425*.

Doorley, K., & Sierminska, E. (2015). Myth or fact? The beauty premium across the wage distribution in Germany. *Economics Letters*, *129*, 29-34.

Drenik, A., & Perez, D. J. (2021). Domestic price dollarization in emerging economies. *Journal of Monetary Economics*, *122*, 38-55.

Duma, N. (2014). Asia’s most dollarized economy: Causes and consequences. *Cambodia Entering a New Phase of Growth*, 27-49.

Dustmann, C., & Meghir, C. (2005). Wages, experience and seniority. *The Review of Economic Studies*, *72*(1), 77-108.

Ekman, L. (2013). Currency competition and the denomination of wages in the Palestinian territories.

El Badaoui, E., Strobl, E., & Walsh, F. (2010). The formal sector wage premium and firm size. *Journal of Development Economics*, *91*(1), 37-47.

Enrico, M. (2011). Local labor markets. *Handbook of Labor Economics*, *4*, 1237-1313.

Eriksson, T., & Kristensen, N. (2014). Wages or fringes? Some evidence on trade-offs and sorting. *Journal of Labor Economics*, *32*(4), 899-928.

Faberman, R. J., & Menzio, G. (2018). Evidence on the relationship between recruiting and the starting wage. *Labour Economics*, *50*, 67-79.

Faberman, J., & Kudlyak, M. (2016). What does online job search tell us about the labor market?. *FRB Chicago Economic Perspectives*, *40*(1).

Faberman, R. J., & Kudlyak, M. (2019). The intensity of job search and search duration. *American Economic Journal: Macroeconomics*, *11*(3), 327-57.

Fakih, A. (2014). Availability of family-friendly work practices and implicit wage costs: New evidence from Canada.

Faryna, O., Pham, T., Talavera, O., & Tsapin, A. (2022). Wage and unemployment: Evidence from online job vacancy data. *Journal of Comparative Economics*, *50*(1), 52-70.

Feingold, A. (1992). Good-looking people are not what we think. *Psychological Bulletin*, *111*(2), 304.

Feliciano, Z. M., & Lipsey, R. E. (2006). Foreign ownership, wages, and wage changes in US industries, 1987–92. *Contemporary Economic Policy*, *24*(1), 74-91.

Forsythe, E., Kahn, L. B., Lange, F., & Wiczer, D. (2020). Labor demand in the time of COVID-19: Evidence from vacancy postings and UI claims. *Journal of Public Economics*, *189*, 104238.

Fortin, N. M. (2005). Gender role attitudes and the labour-market outcomes of women across OECD countries. *Oxford Review of Economic Policy*, *21*(3), 416-438.

Franzoi, S. L., & Herzog, M. E. (1987). Judging physical attractiveness: What body aspects do we use?. *Personality and Social Psychology Bulletin*, *13*(1), 19-33.

French, M. T. (2002). Physical appearance and earnings: Further evidence. *Applied Economics*, *34*(5), 569-572.

French, M. T., Robins, P. K., Homer, J. F., & Tapsell, L. M. (2009). Effects of physical attractiveness, personality, and grooming on academic performance in high school. *Labour Economics*, *16*(4), 373-382.

Galindo, A., Izquierdo, A., & Montero, J. M. (2007). Real exchange rates, dollarization and industrial employment in Latin America. *Emerging Markets Review*, *8*(4), 284-298.

Gallagher, M., Giles, J., Park, A., & Wang, M. (2015). China’s 2008 labor contract law: Implementation and implications for China’s workers. *Human Relations*, *68*(2), 197-235.

Gariety, B. S., & Shaffer, S. (2001). Wage differentials associated with flextime. *Monthly Labor Review.*, *124*, 68.

Gehrsitz, M. (2014). Looks and labor: Do attractive people work more?. *Labour*, *28*(3), 269-287.

Gilmore, D. C., Beehr, T. A., & Love, K. G. (1986). Effects of applicant sex, applicant physical attractiveness, type of rater and type of job on interview decisions. *Journal of Occupational Psychology*, *59*(2), 103-109.

Ginsburgh, V. A., & Prieto-Rodriguez, J. (2011). Returns to foreign languages of native workers in the European Union. *ILR Review*, *64*(3), 599-618.

Glaeser, E. L., & Mare, D. C. (2001). Cities and skills. *Journal of Labor Economics*, *19*(2), 316-342.

Gonzalez, L., & Loureiro, Y. K. (2014). When can a photo increase credit? The impact of lender and borrower profiles on online peer-to-peer loans. *Journal of Behavioral and Experimental Finance*, *2*, 44-58.

Goldsmith, A. H., Hamilton, D., & Darity Jr, W. (2006). Shades of discrimination: Skin tone and wages. *American Economic Review*, *96*(2), 242-245.

Gosling, A. S., Westbrook, J. I., & Spencer, R. (2004). Nurses' use of online clinical evidence. *Journal of Advanced Nursing*, *47*(2), 201-211.

Grinis, I. (2019). The STEM requirements of "Non-STEM" jobs: Evidence from UK online vacancy postings. *Economics of Education Review*, *70*, 144-158.

Grove, W. A., Hussey, A., & Jetter, M. (2011). The gender pay gap beyond human capital heterogeneity in noncognitive skills and in labor market tastes. *Journal of Human Resources*, *46*(4), 827-874.

GSO and ILO (2018). 2016 Report on Informal Employment in Viet Nam. *Hanoi: Hong Duc Publishing House*.

Guéguen, N. (2012). Hair color and courtship: Blond women received more courtship solicitations and redhead men received more refusals. *Psychological Studies*, *57*(4), 369-375.

Guéguen, N., & Jacob, C. (2012). Lipstick and tipping behavior: When red lipstick enhances waitresses tips. *International Journal of Hospitality Management*, *31*(4), 1333-1335.

Hale, G., Regev, T., & Rubinstein, Y. (2021). Do looks matter for an academic career in economics?.

Hall, R. E., & Krueger, A. B. (2012). Evidence on the incidence of wage posting, wage bargaining, and on-the-job search. *American Economic Journal: Macroeconomics*, *4*(4), 56-67.

Hamermesh, D. S., & Biddle, J. E. (1994). Beauty and the labor market. *Economic Review*, *1174*, 1186.

Hamermesh, D. S. (2006). Changing looks and changing "discrimination": The beauty of economists. *Economics Letters*, *93*(3), 405-412.

Hamermesh, D. S., & Abrevaya, J. (2013). Beauty is the promise of happiness?. *European Economic Review*, *64*, 351-368.

Hamermesh, D. S., & Parker, A. (2005). Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review*, *24*(4), 369-376.

Harmon, C., Oosterbeek, H., & Walker, I. (2003). The returns to education: Microeconomics. *Journal of Economic Surveys*, *17*(2), 115-156.

Harper, B. (2000). Beauty, stature and the labour market: A British cohort study. *Oxford Bulletin of Economics and Statistics*, *62*, 771-800.

Hart, K. (1973). Informal income opportunities and urban employment in Ghana. *The Journal of Modern African Studies*, *11*(1), 61-89.

Haveman, H. A., & Beresford, L. S. (2012). If you're so smart, why aren't you the boss? Explaining the persistent vertical gender gap in management. *The ANNALS of the American Academy of Political and Social science*, *639*(1), 114-130.

Hayashi, R., & Matsuda, N. (2020). Covid-19 impact on job postings: Real-time assessment using Bangladesh and Sri Lanka online job portals.

Haynes, J. B., & Sessions, J. G. (2013). Work now, pay later? An empirical analysis of the pension–pay trade off. *Economic Modelling*, *30*, 835-843.

Heilman, M. E., & Saruwatari, L. R. (1979). When beauty is beastly: The effects of appearance and sex on evaluations of job applicants for managerial and nonmanagerial jobs. *Organizational Behavior and Human Performance*, *23*(3), 360-372.

Henley, A., Arabsheibani, G. R., & Carneiro, F. G. (2009). On defining and measuring the informal sector: Evidence from Brazil. *World Development*, *37*(5), 992-1003.

Hensvik, L., Le Barbanchon, T., & Rathelot, R. (2021). Job search during the COVID-19 crisis. *Journal of Public Economics*, *194*, 104349.

Hershbein, B., & Kahn, L. B. (2018). Do recessions accelerate routine-biased technological change? Evidence from vacancy postings. *American Economic Review*, *108*(7), 1737-72.

Heuermann, D., Halfdanarson, B., & Suedekum, J. (2010). Human capital externalities and the urban wage premium: Two literatures and their interrelations. *Urban Studies*, *47*(4), 749-767.

Heywood, J. S., Siebert, W. S., & Wei, X. (2007). The implicit wage costs of family friendly work practices. *Oxford Economic Papers*, *59*(2), 275-300.

Hijzen, A., Martins, P. S., Schank, T., & Upward, R. (2013). Foreign-owned firms around the world: A comparative analysis of wages and employment at the micro-level. *European Economic Review*, *60*, 170-188.

Hirsch, B., Jahn, E. J., Manning, A., & Oberfichtner, M. (2022). The urban wage premium in imperfect labor markets. *Journal of Human Resources*, *57*(S), S111-S136.

Hoberg, G., & Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, *124*(5), 1423-1465.

Hönn, M., & Göz, G. (2007). The ideal of facial beauty: A review. *Journal of Orofacial Orthopedics/Fortschritte der Kieferorthopädie*, *68*(1), 6-16.

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, *20*(1), 1-24.

Islam, A., & Smyth, R. (2012). The economic returns to good looks and risky sex in the Bangladesh commercial sex market. *The BE Journal of Economic Analysis & Policy*, *12*(1).

Javorcik, B., Kett, B., Stapleton, K., & O'Kane, L. (2019). The Brexit vote and labour demand: evidence from online job postings.

Jefferson, Y. (2004). Facial beauty--establishing a universal standard. *International Journal of Orthodontics (Milwaukee, Wis.)*, *15*(1), 9-22.

Jeffreys, S. (2014). Beauty and misogyny: Harmful cultural practices in the West. *Routledge*.

Johnson, S. K., Podratz, K. E., Dipboye, R. L., & Gibbons, E. (2010). Physical attractiveness biases in ratings of employment suitability: Tracking down the “beauty is beastly” effect. *The Journal of Social Psychology*, *150*(3), 301-318.

Johnson, N. B., & Provan, K. G. (1995). The relationship between work/family benefits and earnings: A test of competing predictions. *The Journal of Socio-Economics*, *24*(4), 571-584.

Johnston, D. W. (2010). Physical appearance and wages: Do blondes have more fun?. *Economics Letters*, *108*(1), 10-12.

Jones, T. R., & Price, J. (2017). Information and the beauty premium in political elections. *Contemporary Economic Policy*, *35*(4), 677-683.

Jurajda, Ŝ. (2005). Gender segregation and wage gap: An east-west comparison. *Journal of the European Economic Association*, *3*(2-3), 598-607.

Kaplan, R. M. (1978). Is beauty talent? Sex interaction in the attractiveness halo effect. *Sex Roles*, *4*(2), 195-204.

Kammoun, M., & Power, G. J. (2022). Capital market reactions to project finance loans. *Finance Research Letters*, 45, 102115.

Kapeliushnikov, R. I. (2012). Informal employment in Russia: What do the alternative definitions say. *Higher School of Economics Working Paper WP3/2012/04*.

Karabchuk, T., & Soboleva, N. (2014). Informal employment and subjective well-being in Europe: Evidence from the European social survey data. *Comparative Southeast European Studies*, *62*(4), 397-420.

Karabchuk, T., & Zabirova, A. (2018). Informal employment in service industries: estimations from nationally representative Labour Force Survey data of Russian Federation. *The Service Industries Journal*, 38(11-12), 742-771.

Khamis, M. (2012). Is informal sector work an alternative to workfare benefits? The case of pre-program expansion and economic crisis. *Review of Development Economics*, *16*(4), 579-593.

Kilbourne, B. S., England, P., Farkas, G., Beron, K., & Weir, D. (1994). Returns to skill, compensating differentials, and gender bias: Effects of occupational characteristics on the wages of white women and men. *American Journal of Sociology*, *100*(3), 689-719.

Kreisman, D., & Rangel, M. A. (2015). On the blurring of the color line: Wages and employment for Black males of different skin tones. *Review of Economics and Statistics*, *97*(1), 1-13.

Kubo, K., Sam, V., & Chea, Y. (2021). Currency exchange under payments dollarization: Converting a source of efficiency loss into a catalyst for financial development in Cambodia. *Journal of the Asia Pacific Economy*, 1-25.

Kuhn, P., & Shen, K. (2013). Gender discrimination in job ads: Evidence from China. *The Quarterly Journal of Economics*, *128*(1), 287-336.

Kunze, A. (2005). The evolution of the gender wage gap. *Labour Economics*, *12*(1), 73-97.

Kureková, L. M., Beblavý, M., & Thum-Thysen, A. (2015). Using online vacancies and web surveys to analyse the labour market: A methodological inquiry. *IZA Journal of Labor Economics*, *4*(1), 18.

Kuwabara, K., & Thébaud, S. (2017). When beauty doesn't pay: Gender and beauty biases in a peer-to-peer loan market. *Social Forces*, *95*(4), 1371-1398.

Lackó, M. (2000). Hidden economy–an unknown quantity? Comparative analysis of hidden economies in transition countries, 1989–95. *Economics of Transition*, *8*(1), 117-149.

Landry, C. E., Lange, A., List, J. A., Price, M. K., & Rupp, N. G. (2010). Is a donor in hand better than two in the bush? Evidence from a natural field experiment. *American Economic Review*, *100*(3), 958-83.

Lee, S., Pitesa, M., Pillutla, M., & Thau, S. (2015). When beauty helps and when it hurts: An organizational context model of attractiveness discrimination in selection decisions. *Organizational Behavior and Human Decision Processes*, *128*, 15-28.

Lehmann, H., & Zaiceva, A. (2013). Informal employment in Russia: Incidence, determinants and labor market segmentation.

Lukiyanova, A. L. (2013). Earnings inequality and informal economy in Russia (2000-2010). *HSE Economic Journal*, *17*(2), 212-238.

Machin, S., & Puhani, P. A. (2003). Subject of degree and the gender wage differential: Evidence from the UK and Germany. *Economics Letters*, *79*(3), 393-400.

Malchow-Møller, N., Markusen, J. R., & Schjerning, B. (2013). Foreign firms, domestic wages. *The Scandinavian Journal of Economics*, *115*(2), 292-325.

Maloney, W. F. (1999). Does informality imply segmentation in urban labor markets? Evidence from sectoral transitions in Mexico. *The World Bank Economic Review*, *13*(2), 275-302.

Mang, C. (2012). Online job search and matching quality. *IFO Working Paper*.

Marcouiller, D., de Castilla, V. R., & Woodruff, C. (1997). Formal measures of the informal-sector wage gap in Mexico, El Salvador, and Peru. *Economic Development and Cultural Change*, *45*(2), 367-392.

Marinescu, I., & Rathelot, R. (2018). Mismatch unemployment and the geography of job search. *American Economic Journal: Macroeconomics*, *10*(3), 42-70.

Marinescu, I., & Wolthoff, R. (2020). Opening the black box of the matching function: The power of words. *Journal of Labor Economics*, *38*(2), 535-568.

Matsuda, N., Ahmed, T., & Nomura, S. (2019). Labor market analysis using big data: The case of a Pakistani online job portal. *World Bank Policy Research Working Paper*, (9063).

Mavisakalyan, A. (2018). Do employers reward physical attractiveness in transition countries?. *Economics and Human Biology*, *28*, 38-52.

Mazumdar, D. (1983). Segmented labor markets in LDCs. *The American Economic Review*, *73*(2), 254-259.

McCaig, B., & Pavcnik, N. (2015). Informal employment in a growing and globalizing low-income country. *American Economic Review*, 105(5), 545-50.

Michelacci, C., & Suarez, J. (2006). Incomplete wage posting. *Journal of Political Economy*, *114*(6), 1098-1123.

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy*, *66*(4), 281-302.

Mincer, J. (1974). Schooling, experience, and earnings. *Human Behavior and Social Institutions* No. 2.

Mitra, A. (2002). Mathematics skill and male–female wages. *The Journal of Socio-Economics*, *31*(5), 443-456.

Mobius, M. M., & Rosenblat, T. S. (2006). Why beauty matters. *American Economic Review*, *96*(1), 222-235.

Mocan, N., & Tekin, E. (2010). Ugly criminals. *The Review of Economics and Statistics*, *92*(1), 15-30.

Modestino, A. S., Shoag, D., & Ballance, J. (2016). Downskilling: changes in employer skill requirements over the business cycle. *Labour Economics*, *41*, 333-347.

Morales, L. F., Ospino, C., & Amaral, N. (2021). Online Vacancies and its Role in Labor Market Performance. *Borradores de Economía*; No. 1174.

Moretti, E. (2014). Are cities the new growth escalator?. *World Bank Policy Research Working Paper*, (6881).

Neuman, S., & Oaxaca, R. L. (2004). Wage decompositions with selectivity-corrected wage equations: A methodological note. *The Journal of Economic Inequality*, *2*(1), 3-10.

Nguyen, T. L., Collins, G. S., Spence, J., Daurès, J. P., Devereaux, P. J., Landais, P., & Le Manach, Y. (2017). Double-adjustment in propensity score matching analysis: choosing a threshold for considering residual imbalance. *BMC Medical Research Methodology*, 17(1), 1-8.

Ningrum, P. K., Pansombut, T., & Ueranantasun, A. (2020). Text mining of online job advertisements to identify direct discrimination during job hunting process: A case study in Indonesia. *Plos One*, *15*(6), e0233746.

Niu, Y., Hirudayaraj, M., Sims, C. H., & Kawashima, Y. (2018). Lookism in the Chinese nursing industry: A content analysis of online recruitment advertisements. *Journal of Asia-Pacific Business*, *19*(2), 96-113.

Nomura, S., Imaizumi, S., Areias, A. C., & Yamauchi, F. (2017). Toward labor market policy 2.0: The potential for using online job-portal big data to inform labor market policies in India. *World Bank Policy Research Working Paper*, (7966).

Normand, S. L. T., Landrum, M. B., Guadagnoli, E., Ayanian, J. Z., Ryan, T. J., Cleary, P. D., & McNeil, B. J. (2001). Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. *Journal of Clinical Epidemiology*, *54*(4), 387-398.

Odajima, K. (2017). Dollarization in Cambodia: Behavior of households and enterprises in a highly dollarized environment. *Dollarization and De-dollarization in Transitional Economies of Southeast Asia* (pp. 33-72). Palgrave Macmillan, Cham.

Odajima, K., & Aiba, D. (2019). Dollarization in Cambodia: A review of recent empirical findings from a nation-wide survey. *アジア研究*, *65*(1), 24-44.

O'Higgins, N., & Viegelahn, C. (2021). Informal employment in Viet Nam: Trends and determinants.

Oreffice, S., & Quintana-Domeque, C. (2016). Beauty, body size and wages: evidence from a unique data set. *Economics and Human Biology*, *22*, 24-34.

Pailhé, A., & Solaz, A. (2019). Is there a wage cost for employees in family-friendly workplaces? The effect of different employer policies. *Gender, Work and Organization*, *26*(5), 688-721.

Parrett, M. (2015). Beauty and the feast: Examining the effect of beauty on earnings using restaurant tipping data. *Journal of Economic Psychology*, *49*, 34-46.

Patacchini, E., Ragusa, G., & Zenou, Y. (2015). Unexplored dimensions of discrimination in Europe: Homosexuality and physical appearance. *Journal of Population Economics*, *28*(4), 1045-1073.

Peng, L., Cui, G., Chung, Y., & Zheng, W. (2020). The faces of success: Beauty and ugliness premiums in e-commerce platforms. *Journal of Marketing*, *84*(4), 67-85.

Persico, N., Postlewaite, A., & Silverman, D. (2004). The effect of adolescent experience on labor market outcomes: The case of height. *Journal of Political Economy*, *112*(5), 1019-1053.

Pfann, G. A., Biddle, J. E., Hamermesh, D. S., & Bosman, C. M. (2000). Business success and businesses' beauty capital. *Economics Letters*, *67*(2), 201-207.

Pfeifer, C. (2012). Physical attractiveness, employment and earnings. *Applied Economics Letters*, *19*(6), 505-510.

Póvoa, A. C. S., Pech, W., Viacava, J. J. C., & Schwartz, M. T. (2020). Is the beauty premium accessible to all? An experimental analysis. *Journal of Economic Psychology*, *78*, 102252.

Puga, D. (2010). The magnitude and causes of agglomeration economies. *Journal of Regional Science*, *50*(1), 203-219.

Raghubir, P., & Srivastava, J. (2002). Effect of face value on product valuation in foreign currencies. *Journal of Consumer Research*, *29*(3), 335-347.

Rashad, I. (2008). Height, health, and income in the US, 1984–2005. *Economics and Human Biology*, *6*(1), 108-126.

Ravina, E. (2017). An assessment of the mating motive explanation of the beauty premium in market-based settings. *Behavioral and Brain Sciences*, *40*.

Reingen, P. H., & Kernan, J. B. (1993). Social perception and interpersonal influence: Some consequences of the physical attractiveness stereotype in a personal selling setting. *Journal of Consumer Psychology*, *2*(1), 25-38.

Roberts, M. E., Stewart, B. M., & Nielsen, R. A. (2020). Adjusting for confounding with text matching. *American Journal of Political Science*, *64*(4), 887-903.

Robins, P. K., Homer, J. F., & French, M. T. (2011). Beauty and the labor market: accounting for the additional effects of personality and grooming. *Labour*, *25*(2), 228-251.

Roca, J. D. L., & Puga, D. (2017). Learning by working in big cities. *The Review of Economic Studies*, *84*(1), 106-142.

Rooth, D. O. (2009). Obesity, attractiveness, and differential treatment in hiring a field experiment. *Journal of Human Resources*, *44*(3), 710-735.

Rosen, H. S. (1976). Taxes in a labor supply model with joint wage-hours determination. *Econometrica: Journal of the Econometric Society*, 485-507.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41-55.

Sabia, J. J., & Rees, D. I. (2012). Body weight and wages: Evidence from Add Health. *Economics and Human Biology*, *10*(1), 14-19.

Salter, S. P., Mixon Jr, F. G., & King, E. W. (2012). Broker beauty and boon: A study of physical attractiveness and its effect on real estate brokers' income and productivity. *Applied Financial Economics*, *22*(10), 811-825.

Scholz, J. K., & Sicinski, K. (2015). Facial attractiveness and lifetime earnings: Evidence from a cohort study. *Review of Economics and Statistics*, *97*(1), 14-28.

Sengupta, A., & Das, P. (2014). Gender wage discrimination across social and religious groups in India: Estimates with unit level data. *Economic and Political Weekly*, 71-76.

Siregar, R., & Chan, N. (2014). Factors behind foreign currency holding by household in Cambodia.

Sørensen, K. L., & Vejlin, R. (2014). Return to experience and initial wage level: Do low wage workers catch up?. *Journal of Applied Econometrics*, *29*(6), 984-1006.

Soto, R. (2009). Dollarization, economic growth, and employment. *Economics Letters*, *105*(1), 42-45.

Staneva, A. V., & Arabsheibani, G. R. (2014). Is there an informal employment wage premium? Evidence from Tajikistan. *IZA Journal of Labor and Development*, *3*(1), 1-24.

Štefánik, M. (2012). Internet job search data as a possible source of information on skills demand. *Building on Skills Forecasts—Comparing Methods and Applications*, 246.

Steinmetz, S., Bianchi, A., Tijdens, K., & Biffignandi, S. (2014). Improving web survey quality. *A Data Quality Perspective, John Wiley & Sons Ltd, New York, NY*, 273-298.

Stiglitz, J. E. (1976). The efficiency wage hypothesis, surplus labour, and the distribution of income in LDCs. *Oxford Economic Papers*, *28*(2), 185-207.

Stinebrickner, R., Stinebrickner, T., & Sullivan, P. (2019). Beauty, job tasks, and wages: A new conclusion about employer taste-based discrimination. *Review of Economics and Statistics*, *101*(4), 602-615.

Tansel, A., Keskin, H. I., & Ozdemir, Z. A. (2020). Is there an informal employment wage penalty in Egypt? Evidence from quantile regression on panel data. *Empirical Economics*, *58*(6), 2949-2979.

Turrell, A., Speigner, B., Djumalieva, J., Copple, D., & Thurgood, J. (2018). Using job vacancies to understand the effects of labour market mismatch on UK output and productivity.

Valev, N. T. (2010). The hysteresis of currency substitution: Currency risk vs. network externalities. *Journal of International Money and Finance*, *29*(2), 224-235.

Vu, L., & Rammohan, A. (2022). Is there an informal employment penalty in food security? Evidence from rural Vietnam. *The European Journal of Development Research*, 1-25.

Wada, R., & Tekin, E. (2010). Body composition and wages. *Economics and Human Biology*, *8*(2), 242-254.

Wade, M. R., & Parent, M. (2002). Relationships between job skills and performance: A study of webmasters. *Journal of Management Information Systems*, *18*(3), 71-96.

Williams, C. C., & Kayaoglu, A. (2020). COVID-19 and undeclared work: Impacts and policy responses in Europe. *The Service Industries Journal*, *40*(13-14), 914-931.

Williams, C.C. and Round, J. (2008), "Re-theorising the nature of informal employment: some lessons from Ukraine", *International Sociology*, *23*(3), 367-388.

Williams, C.C. and Windebank, J. (1998), Informal employment in the advanced economies: Implications for work and welfare, *Routledge*, London.

Wilson, R. K., & Eckel, C. C. (2006). Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly, 59*(2), 189-202.

Woodhams, C., Lupton, B., & Xian, H. (2009). The persistence of gender discrimination in China–evidence from recruitment advertisements. *The International Journal of Human Resource Management*, *20*(10), 2084-2109.

Yamaguchi, S. (2018). Changes in returns to task-specific skills and gender wage gap. *Journal of Human Resources*, *53*(1), 32-70.

Ziegler, L. (2021). Skill demand and wages. Evidence from linked vacancy data.