Zunia Saif Tirmazee*

Unequal pay for equal education!
A case of gender wage gap from Punjab, Pakistan[1]

This study aimed to quantify the returns to tertiary educational attainment and measures the extent to which these returns differ for men and women. The article provides new empirical evidence of the returns to tertiary education literature by introducing a unique instrument, namely the supply of education to deal with endogeneity. The analysis is implemented using a pooled cross-section of five rounds of the Pakistan Social and Living Standards survey with 10,000 observations. The results show that the marginal returns to acquiring one extra year of education beyond matriculation are higher for women than for men. This result could partially explain the reversal of the gender gap in enrolments from secondary and lower to the post-secondary level of education in Punjab. The first stage results highlight the significance of investing in physical infrastructure for the greater accumulation of human capital.

Keywords: gender wage gap, returns, supply of education, tertiary education.

1. Introduction

The gender gaps in primary and lower secondary enrolment in Punjab are smaller, but still, boys outnumber girls at both levels. At higher secondary (intermediate) level the gender gap shrinks, and at BA/BS/postgraduate level female students outnumber males (PSLM, 2014). This paper looks at the advantages in the labour

---

* Lahore School of Economics, Pakistan. ORCID: 0000-0002-6158-8130.

market to the increased higher educational attainment and the extent to which these returns differ for males and females, to see whether these differences can to some extent explain the reversal in the gender gaps in education/enrolment. The returns to any level of education broadly fall into three categories (i) private financial, (ii) private non-financial returns such as availability of both better jobs and better working conditions, and (iii) social ones. In this paper, however, the author focused only on the private financial returns by examining their effect of higher education attainment on the wages and the gender gap.

To estimate the causal link between the returns to tertiary education in Punjab and the gender gap in these returns the study used the instrumental variable (IV) technique. Making use of the exogenous variation in the supply of higher education institutes the author employed the total number of available tertiary educational institutes at the district level in Punjab as an instrument. The steady increase in tertiary education institutions reflects improved access to tertiary education for male and female students for two reasons. Firstly, the expansion of tertiary education facilities exogenously decreases the costs associated with attaining more education, by making access to college education cheaper for individuals when a new college is constructed in the local area. Secondly, having a college in one's area also may reduce the mobility concerns which are an important hurdle, especially for females in attaining education (Cheema et al., 2019). In addition to IV, the study also used the region-and-time fixed effects in the first stage to control for region-specific and time-varying unobserved factors that may cause omitted variable bias if not accounted for.

The first stage of this analysis was a regression of educational attainment on the number of tertiary education institutes available in a district. In estimating the first stage, the two identifying assumptions were: a) the relationship between changes in college availability and changes in educational attainment is not reflective of changes in development in general and b) the exact timing of college opening in a given district is not driven by demand for education.

To show that the changes in college availability and in educational attainment are related regardless of the level of development, the author showed that first stage results are robust to controlling for the development of a region using various community-level indicators of development. Additionally, the responsiveness of years of education to variation in the supply of tertiary educational institutes was evident only for the relevant age cohorts, i.e. from 16 to 32 years of age. The results of the first stage were null for sample observations that lie just above and just below the relevant age cohort. This again showed that the first stage results are not reflective of development in general, because if they were there should be a significant positive relationship between the two regardless of the age brackets due to the confounding effect of regional development.

The main findings of the analysis are that there is a positive significant relationship between the estimated years of education from the first stage and the income levels.

Another important result is that males on average earn significantly more than females regardless of the level of education, however an extra year of education brings higher returns to women compared to men. This implies that the gender earnings gap tends to fall with rising education distribution. Whatever the reason for the differential in the earnings of two genders, discrimination or difference in their respective productivities, the results show that as the years of education attained increase, the earnings differential narrows.

The first stage results show that the greater availability of colleges at district level is significantly associated with higher educational attainment at individual level. Moreover, the impact of the increased supply of tertiary education institutions on tertiary education attainment is higher in low Human Development Index (HDI)[2] districts than in the high HDI districts of Punjab. This is an important result from a policy perspective as it shows that investing in the physical infrastructure in less developed regions yields the greatest returns.

This work is a contribution to the study of the labour market in Pakistan, as tertiary education is still an understudied area there. To that end, the analysis makes a significant contribution to the literature on tertiary education by introducing a unique instrument, the district level supply of education, i.e. the number of Arts and Science Intermediate, Degree, and Post-Graduate Colleges for male and female students in a given district at a given point in time in Punjab. A related contribution of this study is that a pooled cross-section with a very large number of observations has not been used to study the dynamics of returns, particularly to tertiary education for Pakistan. This was achieved by making use of five rounds (2006, 2008, 2010, 2012, and 2014) of the household level survey, Pakistan Social and Living Standards Measurement (PSLM), covering a decade. This provided a very large number of observations, approximately 10,000 in this case, and sufficient data points to study this research question.

This analysis derives its significance from the important lessons it bears for policy. For instance, the returns to tertiary education as projected over the life cycle reflect the expectations that influence current student decisions to participate in higher education. If the returns increase with years of education, then there is a positive signal from the labour market, which should effectively lead to greater investment in human capital accumulation. The literature shows that households do respond to information regarding returns to education (Jensen, 2012; Attanasio and Kaufmann, 2009). The results regarding the gender gap in returns to tertiary education can be taken as confirmation of this hypothesis as higher marginal returns could be a reason why female enrolment has been increasing in tertiary education over the past decade (Table 1).

---

[2] The "Human Development Index measures development by quantifying three dimensions of human life – education, health, and standard of living" (Pakistan Human Development Index report 2017) http://hdr.undp.org/sites/default/files/reports/2847/pk-nhdr.pdf

The first stage of this analysis also has at least two very important policy implications. Firstly, the results show the importance of investing in physical capital for the accumulation of human capital. A concerted effort to plan the expansion of the supply of education, especially in areas where there is a dearth of tertiary educational institutions, may allow for the possibility to accumulate greater years of education for individuals who are on the margins. Secondly, the increased availability of tertiary education institutions could also have substantial positive spillover effects when females with tertiary education enter the labour force as school teachers facilitating the supply of more low-cost private schools (Andrabi et al., 2008).

The rest of this paper is organised as follows. Section 2 provides a review of the literature. Section 3 gives a detailed account of the methodology. Section 4 presents the data and descriptive statistics. Section 5 presents the results of the empirical analysis, Section 6 discusses the robustness check, and finally, Section 7 concludes.

2. Literature review

This review of literature sheds light on three issues. Firstly, the issue of gender inequality in the labour market outcomes, secondly, the household's decision to invest in education, and thirdly, the approaches adopted in various studies to establish a causal link between educational attainment and the financial returns to education.

The labour force participation rate of female graduates in Punjab between the age of 25-35 is only 32% compared to that of males at 96% (Labour Force Survey, 2018). The wages of women with higher education are about 68% of the wages of equally qualified men (PSLM, 2014). In this study, the author probed the likely causes of the gender wage gap by breaking it down into explained and unexplained gaps. The main finding was that almost one-third of this gap can be explained by the difference in the human capital of men and women and their nature of work. The remaining two-thirds of this gap remains unexplained and can be attributed to either discrimination or omitted variables (Tirmazee, 2021). Numerous explanations, such as occupational segregation (Levanon et al., 2009; Blau and Kahn, 1992), work interruptions (Epstein, 1988; Neumark and Korenman, 1992), education and training (Blau and Kahn, 2017; Becker, 2010; Mincer, 1962, 1974), temporal flexibility (Goldin, 2014) or unionization (ILO, 2018), have been advanced in the literature for the gender pay gap. In the context of Pakistan, it has been argued that the gender pay gap is because most women either work as unpaid family workers or if in paid employment they are often employed in low-skilled, low-paid jobs (Khan, 2017).

Human capital accumulation, an important determinant of labour market outcomes, reflects the preferences of the demand side and the capacity of the supply side to meet the demand for education. On the demand side, the parents' decision to enroll their children in another year of schooling involves comparing the cost and

benefits of this additional investment, where the benefits may include better wages, healthcare, higher standard of living, etc., and the costs may include tuition fees, the opportunity cost of the child's time, transportation, etc. The simple economic model of comparing the cost of schooling with the benefits implies that if the benefits of schooling such as monetary and non-monetary returns of acquiring education rise, the optimal investment in schooling may increase.

For instance, according to Banerjee et al. (2013), underinvestment in education can be due to the inadequate availability of resources to fund this, or due to the inadequate information that leads households to underestimate the returns to education. If the problem is inadequate resources to fund investment, then there is a need for governments to subsidise education but if the problem is that of inadequate information then there is a need to inform households of the true benefits of education.

The literature does suggest that households respond to information regarding returns to education. For instance, Jensen (2012) provided evidence of how increasing awareness regarding potential job opportunities owing to the rapid expansion of the Business Process Outsourcing (BPOs) industry in India led to a significant rise in investment in the education of younger girls by households. Similarly, Attanasio and Kaufmann (2009) provided evidence on the significance of individual perceptions regarding future returns to schooling using data for Mexico. In their analysis to model college or school choice, they found that mothers' expectations and individual's own expectations matter for college enrolment. In other settings where different instruments were used to provide information such as the author's own calculated returns to education (Jensen, 2010) or a short video showcasing the ways of acquiring financial resources to fund education (Dinkelman and Martinez, 2014), it was seen that households do update their beliefs and react accordingly. Given how households react to information on returns to education, the author believes that this analysis is crucial as it directly yields information on these returns and also reflects the expectations that influence parents' and students' decisions regarding investing time, money, and effort in education.

Since the main objective in this paper was to estimate the causal link between labour market returns and human capital accumulation, the review of various approaches for estimating this causal link follows. The general approach in the literature for tackling the question of private returns to schooling has been an estimation of the Mincerian wage function (Mincer, 1974), which is a simple regression linking schooling with the wages earned. Below there is a simple Mincerian wage equation:

$$Ln\left(Earnings\right)_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + e_i \qquad (1)$$

where $Ln(Earnings)_i$ is the log of yearly earnings of person i and X_i is a vector of individual i's characteristics, and S_i is the accumulated years of education. Ordinary least squares (OLS) yields biased estimates of the parameters of the above equation due to unobserved heterogeneity. In addition, the classic Mincerian wage equation

does not allow to test for the heterogeneity of effects, i.e. how the observed relationship between years of education and the returns to education differs across various subsets of the population. Some possible solutions suggested in the literature to deal with the shortcomings of the Mincerian wage equation are as follows.

Firstly, studies have directly tried to account for the 'ability bias' by including appropriate measures that in some way are a proxy for unobserved ability, such as IQ level or various test scores. However, there are always concerns regarding the extent to which these proxies accurately measure ability, as a multitude of measures for ability has resulted in the past in inconsistent signs for these variables (Dickens and Lang, 1993). One popular method to control for innate ability has been the use of siblings (twins) (Ashenfelter and Zimmerman, 1997; Bingley et al., 2009; Bonjour et al., 2002; Isacsson, 1999; Miller et al., 1995), under the assumption that using twins or siblings allows to differentiate the innate ability since a lot of what determines an individual's ability is common across members of the same household, especially twins. In this way, by eliminating unobserved individual ability by differencing first, one can obtain an unbiased estimator of the return to education by exploiting the differences between education levels and earnings of siblings (Krueger and Ashenfelter, 1992). However, studies regarding twins are often criticised because between-twin differences in schooling are not randomly assigned, but instead are endogenously chosen especially when they depend upon the individual's own aptitude and ability or parental preferences regarding the allocation of expenditure between different children.

A natural experiment is yet another interesting way of tackling the ability bias, where an exogenous event is taken as instrument for the level of education. Some popular natural experiments have been minimum school leaving laws (Harmon and Walker, 1995; Dickson and Smith, 1995), the month of birth (Angrist and Keueger, 1991), and proximity to the school (Card, 1993a, 1999), where the probability of acquiring an extra year of schooling increases/decreases due to the random occurrence of an event that is completely independent of unobserved individual characteristics.

There is another strand of literature that involves identifying an exogenous variable (instrument) that must be correlated with the education level, but is not correlated with the returns to a particular level of education and unobserved ability. In this respect, family background variables such as parental education, spouse's education (Aslam, 2009; Söderbom et al., 2006; Trostel et al., 2002), average education level of the household and/or birth order of an individual (Bertoni and Brunello, 2016; Kantarevic and Mechoulan, 2006) have been used as instruments in the literature. However, the issue remains that because of intergenerational transmission of ability, family background does not completely assure there being no correlation between unobserved ability and the family background variable at hand. Thus the popularly used demand-side family background variables as exogenous determinants of the level of education are often criticised by labour economists as only partially attenuating the ability bias.

As demand-side instruments such as family background, are now widely criticized (Dickson and Smith, 1995), the focus has shifted to the sources of variation in schooling from the supply side such as school-leaving laws or proximity to schools, etc. in search of identifying the source of exogenous variations in education attainment. Using physical capital as an instrument for educational attainment is another way that the literature addresses the endogeneity problem in estimating the returns to education (Duflo, 2001; Maluccio et al., 1998; Card, 1993b). This approach also allows for answer a relevant policy question, namely whether increases in physical infrastructure create opportunities to increase human capital or not. Using physical capital as an instrument is also important from a policy point of view, as increases in human capital ultimately affect the lives and living conditions of citizens, thereby reducing poverty. There is evidence in the literature to suggest that the availability of schools positively affects school enrolment rates owing to the increased and easier access to opportunities to attain education (Khan, 2021; Mazumder et al., 2019; Lavy, 1996; Lillard and Willis, 1994). The availability of schools additionally is linked to improving socio-economic conditions (Carneiro et al., 2013; Case and Deaton, 1999; Currie and Moretti, 2003).

Moreover, Valero and Van Reenen (2019) show that human capital accumulation as well as innovation is an important mediating factor between universities and regional growth.

3. Methodology

In this paper, the author estimates the earnings function (1) using the instrumental variables (henceforth IV methodology). The first stage of the IV procedure is as follows:

$$S_i = \pi_0 + \pi_1 Z_i + \varepsilon_i, \tag{2}$$

where the null hypothesis to be tested is to confirm the instrument (Z_i) relevance $\pi_1 = 0$. The author explains instrument (Z_i) and the first stage (2) in the next section.

3.1. Identification strategy: IV estimation

This study made use of a supply-side IV, i.e. the number of arts and science intermediate, degree and post-graduate colleges for males and females per 10,000 individuals in a district in a given year in Punjab. The region has recently seen tremendous growth in the number of colleges, both private and public, with an increasing number of both genders graduating from these colleges, as shown in Table 1 below. This paper aimed to use this expansion in tertiary education as a means of improving access to tertiary education for male and female students. Moreover, these colleges are also a substitute for private colleges, thereby ensuring ease of access.

Table 1

Number of intermediate, degree colleges, and post graduate classes by gender,
their enrolment, and teaching staff in Punjab

Year	No. of colleges			Enrolment			Teaching staff		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
	Total	Boys	Girls	Total	Boys	Girls	Total	Boys	Girls
2005-06	672	339	333	619	273	346	19131	10677	8454
2007-08	744	379	365	676	306	370	20255	11448	8807
2009-10	901	461	440	724	339	385	23096	12645	10451
2012-13	994	492	502	837	416	421	26312	14490	11822
2014-15	1095	543	552	937	455	482	26823	14997	11826

Source: Punjab Development Statistics (Various issues).

The instrument was calculated as follows:

$$Total\ no.\ of\ colleges\ per\ 10{,}000\ individuals_{dk} =$$
$$= \frac{Total\ no.\ of\ colleges\ in\ a\ district_{dk}}{District\ population_{dk}} \times 10{,}000, \tag{3}$$

where d is any district in Punjab, k is the year in which individual i was in the normal age range for going to college; more about k in the next section.

3.2. First stage

To empirically test if the expansion of tertiary education translates into a greater accumulation of tertiary education, the first stage of this analysis was a regression of years of education attained by individual i on the number of colleges per 10,000 individuals available in a district in year k in which individual i was at the age of going to college. Since the study pooled cross-section spanning over a decade, this also allowed to include in the sample individuals who in the latest year did not fall into the relevant college-going age range, i.e. 16 to 24, which is the standard range for going to college. The final sample included students between the ages 16 to 32, as one typically enters college at an age of 16. Hence, anyone of this age in any of the included rounds of PSLM, i.e. 2006, 2008, 2010, 2012, 2014, was included in the sample. Similarly, the upper bound for the sample was 32 years, as anyone of that age in the latest year of the analysis, i.e. 2014, would be 24 in 2006, and would have just finished their Master's degree, thus the maximum that these data allowed to include regarded individuals of 32 years of age in 2014. A complete description of the sample in tabular form is given in Table 2, which shows if an individual at a particular age in a particular round of PSLM can be included in the sample or not, which depends on whether they were in the college-going age range, namely 16-24, in any of the included rounds. The highlighted cells (blue) are the age ranges from each round included in the final sample.

Table 2

Sample description

Age in	Year when	Age in	Year when	Age in	Year when	Age in	Year when	Age in	Year when
2014	24	2012	24	2010	24	2008	24	2006	24
32	2006	32	2004	32	2002	32	2000	32	1998
31	2007	31	2005	31	2003	31	2001	31	1999
30	2008	30	2006	30	2004	30	2002	30	2000
29	2009	29	2007	29	2005	29	2003	29	2001
28	2010	28	2008	28	2006	28	2004	28	2002
27	2011	27	2009	27	2007	27	2005	27	2003
26	2012	26	2010	26	2008	26	2006	26	2004
25	2013	25	2011	25	2009	25	2007	25	2005
24	2014	24	2012	24	2010	24	2008	24	2006
23	2015	23	2013	23	2011	23	2009	23	2007
22	2016	22	2014	22	2012	22	2010	22	2008
21	2017	21	2015	21	2013	21	2011	21	2009
20	2018	20	2016	20	2014	20	2012	20	2010
19	2019	19	2017	19	2015	19	2013	19	2011
18	2020	18	2018	18	2016	18	2014	18	2012
17	2021	17	2019	17	2017	17	2015	17	2013
16	2022	16	2020	16	2018	16	2016	16	2014

Source: author's own analysis.

The first stage of this analysis is as follows:

$$S_{idt} - \pi_0 + \pi_1 Z_{dk} + \mu_d + \alpha_t + \varepsilon_{idtk}, \tag{4}$$

where S_{idt} stands for years of education, Z_{dk} is number of colleges per 10,000 individuals as calculated in equation (3), available in district d that individual i is from when he/she was at the age k of going to college, μ_d is the district fixed effects which are critical to control for any unobserved time-invariant district attributes that affect college availability and educational attainment in a district. This allowed for a more robust test proving the first identifying assumption; α_t is the year fixed effects to control for time-specific trends that allow to look for any change over time in the choice or preference of individuals, e.g. increase in demand for education or more progressive thinking overtime, etc.

The main hypothesis tested in the first stage was that exposure to a greater number of colleges does not affect educational attainment; $H_0: \pi_1 = 0$. This allowed for correlation of errors within the district, namely the use of cluster-robust standard errors.

3.2.1. Instrument validity

Instrument relevance. The proposed instrument is relevant given these colleges are widely dispersed across the entire province, ensuring greater access to education for both males and females. It is important to highlight here that the very policy that guides the setting up of these colleges ensures greater access. These colleges are set up by the Higher Education Department (HED, a ministerial department responsible for higher education) to improve access to education. The criteria considered before setting up a college in a locality are: (i) there is enough population in the area and (ii) the number of students who pass, out of SSC and intermediate levels from that area, and (iii) land available for college building (Higher Education Commission, 2007). With a motive of setting up an educational facility in every neighbourhood, these colleges find their way to localities where there is enough population to take advantage of this facility, and where a college was not already present in that location.

Instrument exogeneity. To satisfy the exclusion restriction one needs to prove that other district or community level attributes are uncorrelated with the supply of colleges. To prove exogeneity it is necessary to show that the increased availability of opportunities to acquire education is not reflective of the better overall development of a region. Later, the results showed that for the relevant age range the effect, for the most part, is driven by the variation in the availability of colleges even after controlling for the indicators of development. Additionally, controlling for district fixed effects allowed to control for unobserved time-invariant district attributes that may affect both the college presence and the educational attainment and may bias the estimated coefficient of the instrument in the first stage.

The other conjecture to strengthen the exogeneity condition relies on the assumption that the exact timing of college opening in a given district is not driven by demand for education, therefore the preferences or demand of citizens for greater opportunities to acquire education is not a concern here. Hence, one can exploit the fact that the contemporaneous supply of colleges in a district is not driven by the demand for education, but is reflective of pent-up demand as it takes time and involves incurring financial costs to respond to the demand for educational institutions, and consequently to set up an educational facility. To ensure that any time-varying trends or preferences are controlled for year, fixed effects were included in the study.

To further strengthen the argument that demand is not so much of a concern here, the role that political influence plays is also worth discussing. One can imagine that if in a certain district the member of the parliament from that district belongs to the opposition party, they would find it difficult to get funding/approval for a new college in the district, whereas if they are from the ruling party, they may get funding/approval for a new college in the district even in the absence of demand. Thus, the link between demand and the opening of a new college is weak. Therefore, the role of political connections needs to be also considered when thinking about the second identifying assumption.

Additionally, the suggestion that the supply of colleges is not driven by demand is further confirmed by the fact that private colleges (which one could assume are a product of demand) cater to only one-fourth of the total student body that goes to these degree colleges to attain tertiary education, and this has consistently been the case for all the years included in the analysis (Statistics of Arts and Science, 2015). However, so as not to be completely ignorant of the effect of demand, the study controlled for district and year fixed effects to account for preferences and changing trends.

3.2.2. Identifying assumptions

Therefore, in running the first stage the two identifying assumptions were:

1. The relationship between changes in college availability and changes in educational attainment is not reflective of changes in development otherwise. This was addressed in the analysis later by showing that the first stage results hold only for the relevant age range that could have benefited from the increased availability of college and are robust to the district fixed effects.
2. The exact timing of college opening in a given district is not driven by the demand for education. This was addressed by controlling for district and year-fixed effects.

3.3. Second stage

The second stage made use of the estimated years of education from the first stage to estimate the returns to years of education. To find the gender gap in returns to education the study also included in the second stage an indicator for gender, and also interacted the gender indicator with the years of education. The second stage specification was as follows:

$$Ln\left(Earnings\right)_{idt} = \beta_0 + \beta_1 \hat{S}_{idt} + \beta_2 Male_{idt} + \beta_3 \hat{S}_{idt} Male_{idt} + \mu_d + \alpha_t + \varphi_i + \omega_i + e_{idt}, \quad (5)$$

where $Ln(Earnings)_{idt}$ is the log of yearly earnings of person i, in district d, in year t, $\hat{S}_{idt}$ is the estimated years of education for person i, in district d, in year t, $Male_{idt}$ is an indicator variable for gender. It is one for males and zero for females, $\hat{S}_{idt} Male_{idt}$ is the interaction of $\hat{S}_{idt}$ and $Male_{idt}$, μ_d is the district fixed effects to control for unobserved time-invariant district attributes, α_t is the year fixed effects to control for time trends, φ_i is the sector fixed effects to control for unobserved time-invariant industry attributes that may influence both an individual's educational attainment and their eventual returns. For instance, many women in Pakistan end up joining either the education sector or the health sector. Finally, ω_i is the occupation fixed effects.

The main hypothesis proposed to test in the second stage is that higher educational attainment does not affect earnings; $H_0 : \beta_1 = 0$. The study allowed for correlation of errors within district, namely using cluster-robust standard errors.

4. Data and the descriptive statistics

To carry out this analysis the author used a pooled cross-section of five rounds of PSLM for 2006, 2008, 2010, 2012, and 2014. The data for the supply of education were collected from the *Punjab Development Statistics and Statistics of Arts and Science Intermediate and Degree Colleges* for the above stated years.[3] Table 3 below is a concise snapshot of the data used for estimating the impact of tertiary education on yearly wages and how it differs by gender across the years included in the analysis. The table shows that males in the sample are on average older, have more years of experience, and earn more than females, there is also a higher chance that the men in the sample, compared to the women are married. However, the highest education level attained by women is higher than by men, a confirmation of the statistics presented in Table 1 that female enrolment in tertiary education has risen so much that it has been higher than for males in recent years. All of these differences between genders are significant at a 1% level of significance as shown by the t-values of difference in the means test. These differences between genders have been the same across all the years. The important thing to note from this table is how women are improving in terms of their prime human capital determinants in that the difference between males and females in years of experience has been shrinking over time. Secondly, women, in trying to catch up with men in terms of human capital, have surpassed them as far as the years of education attained is concerned.

As shown in Table 3, men in this sample on average earn significantly more than women, the same phenomenon is evident in the kernel density graph shown in Figure 1. The data suggest that the distribution of yearly wages for males for different levels of education peaks to the right of the series' mean compared to that of women who have a much wider distribution. It is also noteworthy that the shape of the wage distribution for women changes from bimodal to just like that of men from Intermediate to Master's level. This points to the fact that there is a greater inequality of wages between men and women at lower educational levels. At higher educational levels such Master's, women tend to be doing much better in catching up with men. There are three lessons to learn from these figures: (i) it appears that women do not tend to work in jobs that offer very low wages(as the distribution for women starts at a level higher than for men), (ii) for all the three levels of education, women outnumber men at the lower end of the distribution which means that no matter what level of education they achieve, women earn less than men, and (iii) just above the mean, the distribution for women is lower than men – suggesting that at the higher end of wage distribution, women are outnumbered by men.

---

[3] Data available on the Punjab Bureau of Statistics website. http://www.bos.gop.pk/ developmentstat

Table 3

Summary statistics

Variable	Female	Male	Difference	Count
	(1)	(2)	(3)	(4)
Year: 2006				
Age	25.002	26.397	1.395***	2,488
Experience (Years)	6.897	8.943	2.046***	2,488
Married	0.309	0.43	0.121***	2,488
Years of Education	13.15	12.478	-0.672***	2,488
Yearly Earnings	69,877.63	116157	46,279.520***	2,483
Observations	573	1,915	2,488	
Year: 2008				
Age	24.955	26.48	1.525***	2,524
Experience (Years)	6.594	8.968	2.373***	2,524
Married	0.262	0.438	0.176***	2,524
Years of Education	13.383	12.541	-0.841***	2,524
Yearly Earnings	88,985.24	158628	69,642.836***	2,524
Observations	572	1,952	2,524	
Year: 2010				
Age	25.11	26.589	1.479***	2,488
Experience (Years)	5.685	7.883	2.197*** 2,488	
Married	0.261	0.45	0.189***	2,488
Years of Education	14.536	13.759	-0.776***	2,488
Yearly Earnings	111171.02	204334	93,162.688***	2,485
Observations	655	1,833	2,488	
Year: 2012				
Age	25.126	26.595	1.468***	2,743
Experience (Years)	6.751	8.605	1.854***	2,743
Married	0.293	0.49	0.198***	2,743
Years of Education	13.415	12.999	-0.416***	2,743
Yearly Earnings	143537.5	236195	92,657.758***	2,739
Observations	816	1,927	2,743	
Year: 2014				
Age	25.248	26.158	0.910***	2,298
Experience (Years)	7.023	8.28	1.257***	2,298
Married	0.352	0.479	0.127***	2,298
Years of Education	13.257	12.887	-0.370***	2,298
Yearly Earnings	131008.49	228949	97,940.266***	2,291
Observations	690	1,608	2,298	

Source: author's own calculations.
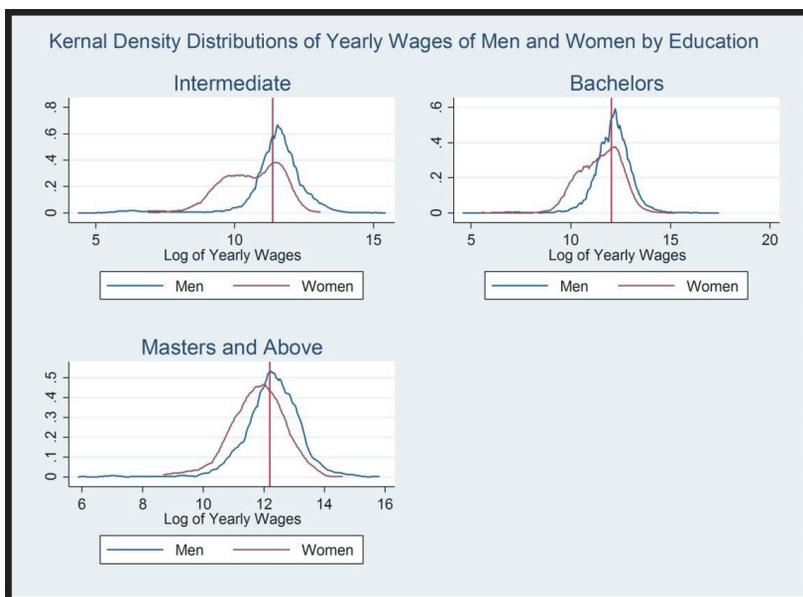
Fig. 1. Wage densities by education for men and women

Source: author's own calculations using various rounds of PSLM data.

As the wage distribution of women is 'bulkier' toward the lower end of the distribution, and begins to decline on the higher end before the men's wage distribution declines, there is a clear indication that women earn less than men. Although the raw data indicates a gender gap in wages earned by men and women who have acquired more than ten years of education, the next section reconfirms this observation using OLS and instrumental variables regression.

5. Results

The results of the first stage are reported in Table 4, where one can see that the chosen instrument predicts the highest education level attained by an individual. The total number of educational institutions per 10,000 individuals in a district in the year when an individual was at the age of going to college significantly affects the years of education. As hypothesised, the greater the number of tertiary education institutions the person is exposed to when they were at the age of going to college, significantly greater the likelihood of going to college and therefore attaining a longer period of education. Moreover, the study also controlled for the unobserved time-invariant factors by checking for district fixed effects and time-varying year fixed effects. The author's argument that increasing the schooling inputs per capita improves access to education subsequently leading to an increase in the highest level of education attained seems to be valid.

To prove that the first stage results were not driven by the overall development in a district, the first stage controlled additionally for community-level development indicators, which are: a source of drinking water, grocery store, public transport, primary school, secondary school, hospital, and a population welfare centre available within thirty minutes' distance of the household. In column 2 in Table 4, one can see that the instrument still continues to hold its significance; although the coefficient's size was reduced, it remains positive and significant.

To prove that the changes in the years of education do not reflect changes in development in general, and that for the sample, i.e. within the given age range additional years of education attained above matriculation are significantly affected by college availability in the district, the first stage was run for individuals who did not fall in the desired sample age range, i.e. aged 16-32, in any of the included years. The logic behind doing this was to see if the relationship between years of education and number of colleges is spurious. If so, then one should see years of education increasing regardless of college presence, even for individuals who are not at the age of going to college.

Table 4

First stage regression

Dependent variables	Years of Education	
	(1)	(2)
Total number of colleges per 10,000 individuals	4.52**	0.936*
	(1.821)	(0.558)
Year FE	YES	YES
District FE	YES	YES
Development Indicators	NO	YES
Observations	11,677	11,677
Mean of the dependent variable	12.86	12.86

Note: Robust standard errors in parentheses. SE's clustered by districts. Controls in the first stage include experience, experience squared, married. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Source: author's own calculations.

Therefore, the analysis was run on very narrow age bands around the upper (32 yrs) and lower (16 yrs) age cut-off to make the comparison between roughly similar groups. Thus, the comparison at the lower end of age distribution was made between individuals from the age bracket 12-15 and those from 16 to 19. At the upper end of the age distribution the comparison was made between individuals from the

age bracket 29-32, and those from 33 to 36. For each of these four regressions other than the basic controls included in Table 4, development indicators were also controlled for. The coefficient plots of these regressions are presented in Figure 2. Therefore, for individuals who are at the age of going to college, it can be argued safely that college presence matters. If this was a spurious result, then one would also see some impact in the similar groups, and human capital increasing regardless of college presence, thus reflecting the general progression of the society. Additionally, year fixed effects were also incorporated in generating Figure 2, to control for trends that change over time, such as a preference for longer periods of education, etc.
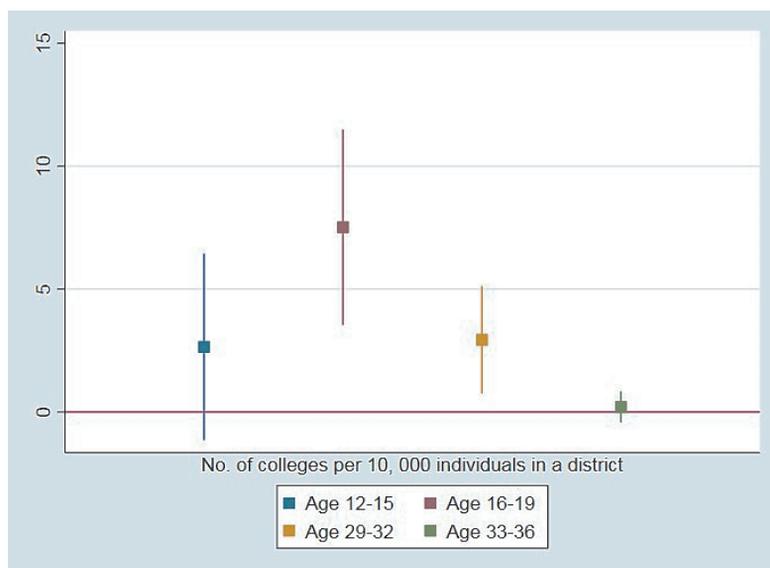
Fig. 2. Estimates of beta coefficients of number of colleges for tighter time windows controlling for community development

Note: Community development indicators controlled for are: access to piped water, grocery store, public transport, primary school, middle school, hospital, and population welfare centre.

Source: author's own analysis.

The results for the second stage show that there are positive returns to attaining tertiary education, and that there is a gender gap in those returns in favour of men. The results for the second stage are presented in Table 5. Column 1 shows that years of education beyond matriculation significantly affect one's wage. Moreover, *Male* – the indicator for gender, shows that men have higher wages than women on average. However, an additional year of education brings a significantly greater increment in wages of women compared to those of men, as the coefficient on the interaction term of *Male* and *Years of education* is significant and negative. This result is in line with earlier findings in the literature, which also suggest that as years of schooling

increase, gender wage gap tends to fall (Blau and Kahn, 2017; Blundell et al., 2000). These higher marginal returns for women could result from the dual impact, i.e a direct effect of human capital accumulation on returns which is also true for men, but in addition, there is an indirect impact for women which is due to the attenuation of the impact of discrimination, tastes and circumstances (DTC) (Dougherty, 2005).

Tirmazee (2021) confirmed this finding by showing that the gender wage gap is highest at the lowest end of the wage distribution, and is contributed in large part by the unexplained gap. The inverse relationship between DTC (hence the wage gap) and the years of schooling could probably be as more educated women have a degree or a formal qualification required in a job that offers a standardised wage, or highly educated women may be able to deal well with discrimination, or may even be able to find better job openings for herself where her characteristics are rewarded fairly (Dougherty, 2005). In column 2 in Table 5, one can see that all of these results are robust to controlling for development indicators.

Table 5

Second stage regression

Dependent variables	Log of Yearly Earnings	
	(1)	(2)
Years of Education	0.827***	1.041**
	(0.080)	(0.434)
Male	1.885***	2.292***
	(0.140)	(0.807)
Male* Yrs of Education	-0.553***	-0.757*
	(0.073)	(0.410)
Observations	11,677	11,677
Mean	11.62	11.62
Year FE	YES	YES
District FE	YES	YES
Industry FE	YES	YES
Occupation FE	YES	YES
Development Indicators	NO	YES

Note: Robust standard errors in parentheses. SEs clustered by districts. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Controls include: experience, experience squared, marital status, marital status interacted with gender. Column (2) additionally controls for development indicators.

Source: author's own analysis.

An important consideration in estimating the second stage is the possibility of difference in the quality of education imparted in male and female colleges. However, that is not problematic, as all of these colleges are set up by the provincial ministerial

department HED under uniform guidelines and are of similar quality. As mentioned above, these colleges are also cheaper, more accessible options made available for students who cannot afford to study in private colleges. The teaching staff in these colleges are recruited through a central standard procedure and are rotated periodically between colleges. All the colleges must meet the minimum requirements of available legal and physical infrastructure as outlined by the Higher Education Commission in the PU-01 proforma for setting up higher education Institutions.[4]

6. Robustness checks

6.1. Selection's correction

Next, the author aimed to correct for selection bias given the very selected sample of individuals who are in paid employment and also have acquired more than ten years of education. The fact that the decision to join the workforce for women is not random, given their expected gender role, they are required to look after the family and the household. The need to maintain a work-life balance may affect women's decision to participate in the workforce, and depending on that, may also affect their choice of job, profession, or industry. There could also be some self-selection on the men's side, if they deliberately decide to enter paid work rather than be self-employed. This possibility is equally valid for women as well. All of these decisions made by men and women are dependent upon their socio-economic conditions and may ultimately affect the returns they earn in the labour market. The extent to which a certain level of education is financially beneficial for men and women is thus not independent of the choices made, and therefore can be expected to interact with these work-related choices to either attenuate or augment the returns to education.

Another source of the sample selection bias is the very selected sample of individuals who have more than ten years of education, since a very privileged group of people continue to post-secondary education in Pakistan. Although women have caught up with men at tertiary educational level, at lower levels there is still a substantial gender gap which means that there is a very narrow group of women who enter into tertiary education.

To correct for both selection biases, the study used the Heckmann two-step procedure (Heckman, 1976, 1979). Correcting for selection in paid employment involved first estimating the probability of participating in the workforce using the probit model. The exclusion restriction of the participating equation included dependent children and adults aged less than seven years and above sixty years respectively in the household. The probability of salaried employment estimated

---

[4] https://hec.gov.pk/english/services/universities/Documents/887_HEC2_Criteria_of_university_institutions.pdf

from the participating equation was used to estimate the inverse Mills ratio or the selectivity term (lambda) which was later used as one of the controls in the second stage.

To correct for selection in higher education, the author ran a second selection function where the probability of having acquired higher education was regressed on all the controls in the main wage equation, along with the average education level of the household as an exclusion restriction.[5] The probability of acquiring higher education estimated from this selection function was used to estimate another inverse Mills ratio that was also used as an additional control in the second stage.

To correct for endogeneity between years of education and wages along with correcting for the two selection biases mentioned above, the study incorporated selectivity terms (IMR1 and IMR2) as controls in the IV regression. The results are shown in columns 2 and 3 of Table 6. Here too, the coefficients' sizes, their signs, and significance are similar to those obtained in the simple IV regression in Tables 4 and 5. The results from this section suggest that when corrected for the selection bias, the main results continue to hold and therefore it is safe to assume for Punjab's labour market, that for both men and women who have more than ten years of education additional years of education bring greater returns, but that those additional years are comparatively more beneficial for women than for men, leading to a reduction in the gender wage gap at each successive level of education.

6.2. Are the results driven by affluent districts?

There is also some a-priori evidence from the data to suggest that in districts where the enrolment per capita is higher, the Human Development Index[6] is also higher for the year 2015. The defined concept of Human Development Index, reflects the increase in the capabilities of people by providing them with increased opportunities and the 'freedom of choice' to avail those opportunities. The choropleth maps of the province of Punjab below show that there is indeed a correlation between the HDI (Figure 3) and the enrolment in higher education (Figure 4) across districts.

This is especially true for districts in the north and the centre. Thus, for instance, Rawalpindi (north) and Lahore (centre) with a very high Human Development Index, also show very high levels of enrolment per capita in higher education. Similarly, districts in the south such as Rajanpur, Bahawalnagar, and Bahawalpur, and those in the West, such as Dera Ghazi Khan, and Muzaffargarh, have both lower Human Development Indices and also lower enrolements per capita.

---

[5] Although a better exclusion restriction could have been the parental education (Asadullah and Xiao, 2019), the data allowed this information only for a very small subsample, therefore the study used the average education level of the household excluding one's own.

[6] The HDI for Punjab has been taken from the Pakistan Human Development Index report 2017.

Z. S. Tirmazee

Table 6

Correcting for selection

Dependent variables	Log of Yearly Earnings	Yrs. of Education
	Second Stage	First Stage
	(1)	(2)
Yrs. of Education	1.128** (0.445)	
Male	2.526*** (0.821)	
Male* Yrs. of Education	-0.860** (0.418)	
IMR1	-0.007 (0.135)	
IMR2	-0.320*** (0.060)	
Log Total Number of Colleges		0.951* (0.551)
Year FE	YES	YES
District FE	YES	YES
Industry FE	YES	YES
Occupation FE	YES	YES
Development Indicators	YES	YES
Observations	11,398	11,398

Note: Robust standard errors in parentheses. SE's clustered by districts. ***$p<0.01$,**$p<0.05$, *$p<0.1$. The dependent variable in the participation equation is salaried employment. The exclusion restriction for the participation equation is dependent children and adults aged < 7 yrs and > 60 yrs respectively in the household. IMR1 is the inverse Mills ratio calculated from this equation. The dependent variable in the selection function for selection into higher education is if individuals have acquired more than ten years of education. The exclusion restriction used is the average education level of the household excluding one's own education. IMR2 is the inverse Mills ratio estimated from this second selection function.

Source: author's own analysis.

Therefore, there is an indication from the data that higher human capital accumulation is correlated with better human development. This points to the need of devising ways that can ease the accumulation of human capital and providing physical infrastructure, i.e. schools and colleges is one of the policy options available. The study provides evidence that the responsiveness of human capital accumulation is higher in districts that are relatively worse off as indicated by HDI, making it all the more policy-relevant to invest in physical infrastructure in poorer or worse-off areas.
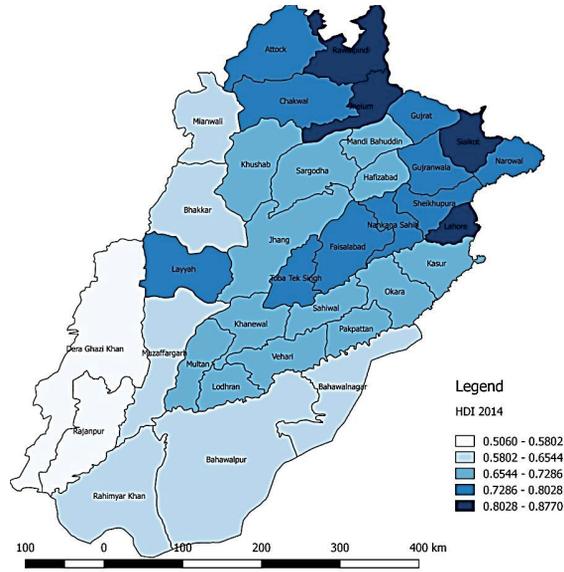
Fig. 3. Human Development Index, 2015

Source: author's own analysis using HDI figures from the UNDP Human Development Index report, 2017.
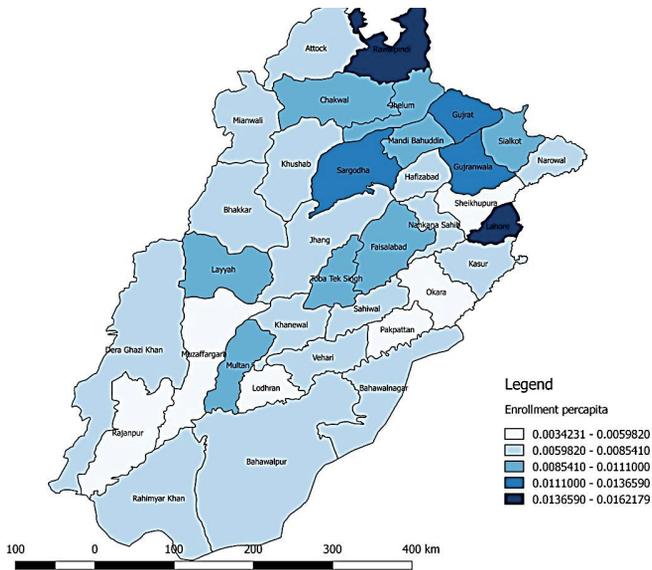
Fig. 4. Enrolment per capita, 2015

Source: author's own analysis using enrolment figures from the Punjab Development Statistics, an annual publication of the Punjab Bureau of Statistics.

Table 7

IV regression: by HDI

Variables	High HDI		Low HDI	
	Second Stage	First Stage	Second Stage	First Stage
	(1)	(2)	(3)	(4)
Yrs. of Education	1.074		1.322***	
	(0.800)		(0.495)	
Male	2.321		2.851***	
	(1.497)		(0.951)	
Male* Yrs. of Education	-0.770		-1.049**	
	(0.760)		(0.476)	
Log Total Number of Colleges		0.447		1.684***
per 10,000 individuals		(0.559)		(0.439)
Observations	7,075	7,075	4,602	4,602
Year FE	YES	YES	YES	YES
District FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Occupation FE	YES	YES	YES	YES

Note: Robust standard errors in parentheses. SEs clustered by districts. ***$p<0.01$, **$p<0.05$, *$p<0.1$.

Source: author's own analysis.

To rule out the possibility that the effect observed in the first stage may be driven by the more affluent districts where the human capital accumulation and physical infrastructure is relatively abundant, the author ran the analysis separately for better-off and worse-off districts. The distinction was made based on the HDI of the districts in 2015. The results are shown in Table 7. The results of the first stage are significant for poorer low HDI districts while they are insignificant for the richer high HDI districts. This shows that there is a greater impact of investing in physical infrastructure where the opportunities are already lagging. The results for the second stage also are significant for the low-HDI regions. The earnings for men are higher on average but an extra year of education brings comparatively greater returns for women than for men.

6.3. Are the results driven by men?

To rule out the possibility that the effect observed in the first stage may be driven by the male colleges, and that the availability of colleges may be increasing their enrolment relatively more than those for females, the colleges were split by gender

and by HDI levels to see the impact of each on the educational attainment. To distinguish the responsiveness of female human capital to the presence of physical infrastructure from that of males, the author ran the IV regression by breaking down the total number of colleges into male and female colleges, taking the two as separate instruments in the first stage.

Table 8

IV regression: men and women colleges

Variables	High HDI		Low HDI	
	Second Stage	First Stage	Second Stage	First Stage
	(1)	(2)	(3)	(4)
Yrs. of Education	1.024***		1.002***	
	(0.157)		(0.203)	
Male	2.21***		2.186***	
	(0.285)		(0.375)	
Male* Yrs. of Education	-0.716***		-0.738***	
	(0.148)		(0.190)	
Log Total Number of Girls' Colleges		0.0004		0.004*
per 10,000 girls		(0.001)		(0 .002)
Log Total Number of Boys' Colleges		0.001		0.001
per 10,000 boys		(0.001)		(0.001)
Observations	6,937	6,937	4,764	4,764
Year FE	YES	YES	YES	YES
District FE	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES
Occupation FE	YES	YES	YES	YES

Note: Robust standard errors in parentheses. SEs clustered by districts. ***p<0.01, **p<0.05, *p<0.1.

Source: author's own analysis.

The results shown in Table 8 demonstrate that when the colleges are split up by gender, only the coefficient for female colleges is significant and positive in the first stage in the poorer districts; the coefficient for male colleges is not significant in any of the regressions. In the richer districts, the significance of female colleges also disappears, which is an even stronger indication of the result obtained in section 6.2. College availability seems to make a difference to the marginalised group in lagging areas which include districts in the south and west of the province. These districts are also much more conservative in their values regarding educating women, let alone

allowing them go to a distant college in a neighbouring district or provincial capital. Therefore, increased college availability in a district may ease the mobility constraint for a lot of female students, making gender-segregated tertiary education institutes available for them (where others from similar backgrounds may be able to acquire higher education degrees). The results of the second stage stay the same.

Conclusion

This paper is an investigation of the gender gap in the returns to tertiary education in Punjab using the Instrumental Variables technique. The study used exogenous variation in the expansion of the supply of higher education institutions to men and women to identify and compare the returns to tertiary education for both genders in Punjab, Pakistan. A large number of colleges in a given district can affect the probability of moving from secondary to tertiary education, since their accessibility improves by alleviating two constraints, namely the high cost of acquiring a higher degree, and mobility if a college is built in one's locality. To carry out this analysis the author used a pooled cross-section constructed from five rounds of the PSLM survey for 2006, 2008, 2010, 2012, and 2014.

The results of this analysis suggest that there is a significant positive relationship between years of education beyond matriculation and the earnings of individuals. Moreover, the marginal returns to acquiring one extra year of education are higher for women than for men, suggesting that gender inequality tends to fall as human capital accumulation improves. The first stage results confirm that it is important to invest in building infrastructure to increase educational attainment. Having controlled for other development indicators and shown that this relationship holds for the appropriate age range (16-32, the sample should cover college-going age during one or more years included in the study), we remove doubts regarding the first stage results as not being causal.

Some important policy lessons to be learned from this analysis are: firstly, the significance of investing in higher education, both because it increases the prospects of graduates in the labour market by increasing labour market returns, and also it decreases gender inequality in the labour market returns. Secondly, the significance of investing in the physical infrastructure such as universities or higher education institutions, as this facilitates the accumulation of human capital by making it less costly for households to invest in it when a higher education institution is built in their locality. Cheema et al. (2019) described the glass walls hindering women to take up training, and showed that once an education centre is housed in their village, it significantly increases their take-up rates.

Thirdly, the responsiveness of human capital investment to investment in physical capital is greatest in the less developed regions of Punjab, hence the greatest returns can be achieved by targeting the expansion of educational institutes to lagging

regions. Using physical capital as an instrument is also important, as an increase in human capital ultimately affects the lives and living conditions of citizens, reducing poverty. There is evidence in the literature to suggest that the availability of schools positively affects school enrolment rates owing to the increased and easier access to opportunities to attain education. In addition, the availability of schools is linked to improving socio-economic conditions (Carneiro et al., 2013; Case and Deaton, 1999; Duflo, 2001; Currie and Moretti, 2003). Moreover, Valero and Van Reenen (2019) showed that human capital accumulation, as well as innovation, is an important mediating factor between universities and regional growth.

Fourthly, a more indirect lesson – a spin-off of the first two lessons – is the spillover effects of investing in physical infrastructure, i.e. the government, by establishing tertiary education institutions in the less developed regions could promote the growth of private low-cost high schools in the area, as graduates from the higher education institutions, enter the labour market to increase the supply of teachers at primary and secondary levels of schooling. This is possible because of an increase in the supply of school teachers graduating from these tertiary education institutions (Andrabi et al., 2008).

There are some important directions that this work could be extended into, that are currently either beyond the scope of this study or because of data being unavailable. The first important direction for the future research is to find out which subject streams or fields tend to reduce the gender gap in the returns to tertiary education the most. Secondly, the distance to a tertiary education institution is a better reflection of improvement in access and could be a better instrument. In this case, the data does not allow to include the distance from a household to a college. Another important research dimension is to examine, while tertiary education is expanding, what is happening to the quality of higher education imparted across different institutions, and what implications could this have for the gender gap in labour market outcomes. Lastly, the fate of graduates largely depends on the balance between supply and demand. An interesting extension of this analysis could be to find out how much of the results are driven because of the expansion of access, or because of the expansion of the demand for graduates in the labour market.

References

Andrabi, T., Das, J. and Khwaja, A. I. (2008). A dime a day: The possibilities and limits of private schooling in Pakistan. *Comparative Education Review*, 52(3), 329–355.

Angrist, J. D. and Keueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4), 979–1014.

Asadullah, M. N. and Xiao, S. (2019). Labor market returns to education and English language skills in the People's Republic of China: An update. *Asian Development Review*, 36(1), 80–111.

Ashenfelter, O. and Zimmerman, D. J. (1997). Estimates of the returns to schooling from sibling data: Fathers, sons, and brothers. *Review of Economics and Statistics*, 79(1), 1–9.

Aslam, M. (2009). Education gender gaps in Pakistan: Is the labor market to blame? *Economic Development and Cultural Change*, 57(4), 747–784.

Attanasio, O. and Kaufmann, K. (2009). Educational choices, subjective expectations, and credit constraints. Technical report, National Bureau of Economic Research.

Banerjee, A., Glewwe, P., Powers, S. and Wasserman, M. (2013). *Expanding access and increasing student learning in post-primary education in developing countries: A review of the evidence.* Citeseer.

Becker, G. S. (2010). *The economics of discrimination*. University of Chicago press.

Bertoni, M. and Brunello, G. (2016). Later-borns don't give up: The temporary effects of birth order on European earnings. *Demography*, 53(2), 449–470.

Bingley, P., Christensen, K., and Walker, I. (2009). The returns to observed and unobserved skills over time: Evidence from a panel of the population of Danish twins. *Danish National Institute for Social Research*.

Blau, F. D. and Kahn, L. M. (1992). The gender earnings gap: Learning from international comparisons. *The American Economic Review*, 82(2), 533–538.

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of economic literature*, 55(3):789–865.

Blundell, R., Dearden, L., Goodman, A. and Reed, H. (2000). The returns to higher education in Britain: Evidence from a British cohort. *The Economic Journal*, 110(461), F82–F99.

Bonjour, D., Cherkas, L., Haskel, J., Hawkes, D. and Spector, T. (2002). Estimating returns to education using a new sample of UK twins. *Queen Mary and Westfield College, London*.

Card, D. (1993a). Using geographic variation in college proximity to estimate the return to schooling. Technical report, National Bureau of Economic Research.

Card, D. (1993b). Using geographic variation in college proximity to estimate the return to schooling.

Card, D. (1999). The causal effect of education on earnings. In *Handbook of Labour Economics*, volume 3, pages 1801–1863. Elsevier.

Carneiro, P., Meghir, C. and Parey, M. (2013). Maternal education, home environments, and the development of children and adolescents. *Journal of the European Economic Association*, 11 (suppl 1), 123–160.

Case, A. and Deaton, A. (1999). School inputs and educational outcomes in South Africa. *The Quarterly Journal of Economics*, 114(3), 1047–1084.

Cheema, A., Khwaja, A. I., Naseer, F. and Shapiro, J. N. (2019). Glass walls: Experimental evidence on access constraints faced by women. Technical report, Mimeo, Harvard University.

Currie, J. and Moretti, E. (2003). Mother's education and the intergenerational transmission of human capital: Evidence from college openings. *The Quarterly journal of economics*, 118(4), 1495–1532.

Dickens, W. T. and Lang, K. (1993). Labor market segmentation theory: Reconsidering the evidence. *Labor economics: Problems in analyzing labor markets*, 141–180.

Dickson, M. and Smith, S. (1995). What determines the return to education: An extra year or a hurdle cleared? *Economics of Education Review*, 30(6), 1167–1176.

Dinkelman, T., Martinez, A. C. (2014). Investing in schooling in Chile: The role of information about financial aid for higher education. *Review of Economics and Statistics*, 96(2), 244–257.

Dougherty, C. (2005). Why are the returns to schooling higher for women than for men? *The Journal of Human Resources*, 40(4), 969–988.

Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review*, 91(4), 795–813.

Epstein, C. F. (1988). *Deceptive distinctions: Sex, gender, and the social order*. Yale University Press.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American EconomicReview*, 104(4), 1091–1119.

Harmon, C. and Walker, I. (1995). Estimates of the economic return to schooling for the United Kingdom. *The American Economic Review*, 85(5), 1278–1286.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables, and a simple estimator for such models. In *Annals of economic and social measurement*, 5(4), 475–492, NBER.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153–161.

Higher Education Commission (2007). Guidelines for the establishment of a new university or an institution of higher education.

Isacsson, G. (1999). Estimates of the return to schooling in Sweden from a large sample of twins. *Labour Economics*, 6(4), 471–489.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics*, 125(2), 515–548.

Jensen, R. (2012). Do labor market opportunities affect young women's work and family decisions? Experimental evidence from India. *The Quarterly Journal of Economics*, 127(2), 753–792.

Kantarevic, J. and Mechoulan, S. (2006). Birth order, educational attainment, and earnings an investigation using the PSID. *Journal of Human Resources*, 41(4), 755–777.

Khan, F. (2017). Barriers to pay equality in Pakistan. *International Labor Organization*, Geneva, Switzerland.

Khan, S. (2021). Unmet expectations: The impacts of school construction on female outcomes in rural Punjab, Pakistan.

Krueger, A. and Ashenfelter, O. (1992). Estimates of the economic return to schooling from a new sample of twins. *National Bureau of Economic Research.*

Lavy, V. (1996). School supply constraints and children's educational outcomes in rural Ghana. *Journal of Development Economics*, 51(2), 291–314.

Levanon, A., England, P. and Allison, P. (2009). Occupational feminization and pay: Assessing causal dynamics using 1950–2000 U.S. census data. *Social Forces*, 88(2), 865–891.

Lillard, L. A. and Willis, R. J. (1994). Intergenerational educational mobility: Effects of family and state in Malaysia. *Journal of Human Resources*, 29(4), 1126–1166.

Maluccio, J. A. (1998). Endogeneity of schooling in the wage function. Technical report, International Food Policy Research Institute (IFPRI).

Mazumder, B., Rosales-Rueda, M. and Triyana, M. (2019). Intergenerational human capital spillovers: Indonesia's school construction and its effects on the next generation. In *AEA Papers and Proceedings*, 109, 243–249.

Miller, P., Mulvey, C. and Martin, N. (1995). What do twin studies reveal about the economic returns to education? A comparison of Australian and U.S. findings. *The American Economic Review*, 85(3), 586–599.

Mincer, J. (1962). Labor force participation of married women: A study of labor supply. In *Aspects of labor economics*, 63–105, Princeton University Press.

Mincer, J. (1974). Schooling, experience and earnings. Series *Human Behavior Social Institutions*, NBER.

Neumark, D. and Korenman, S. (1992). Sources of bias in women's wage equations: Results using sibling data, NBER Working Papers 4019, National Bureau of Economic Research, Inc.

Söderbom, M., Teal, F., Wambugu, A. and Kahyarara, G. (2006). The dynamics of returns to education in Kenyan and Tanzanian manufacturing. *Oxford Bulletin of Economics and Statistics*, 68(3), 261–288.

Tirmazee, Z. (2021). Too much of a good thing? Increasing gender wage disparity in face of rapidly expanding post-secondary education in Punjab, Pakistan. Technical report.

Trostel, P., Walker, I. and Woolley, P. (2002). Estimates of the economic return to schooling for 28 countries. *Labour economics*, 9(1), 1–16.

Valero, A. and Van Reenen, J. (2019). The economic impact of universities: Evidence from across the globe. *Economics of Education Review*, 68, 53–67.