RETURNS TO COMPUTER USE IN BANGLADESH: AN ECONOMETRIC ANALYSIS

Presented by

Syed Yusuf Saadat

Young Scholars Seminar Series

Centre for Policy Dialogue

Dhaka: 20 November 2019

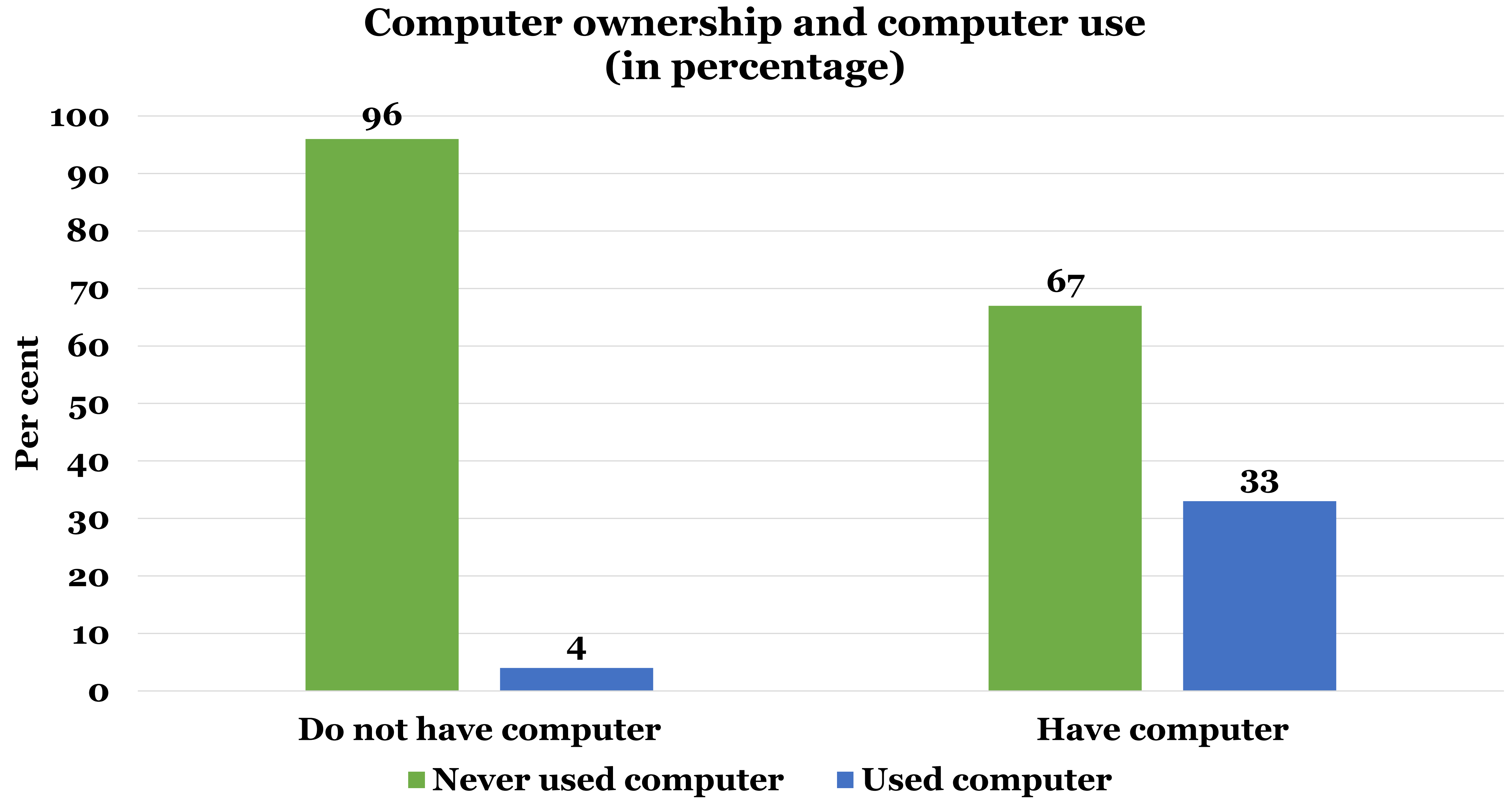
Draft;
not for citation

Draft;
not for citation

INTRODUCTION

SUSTAINABLE DEVELOPMENT GOALS

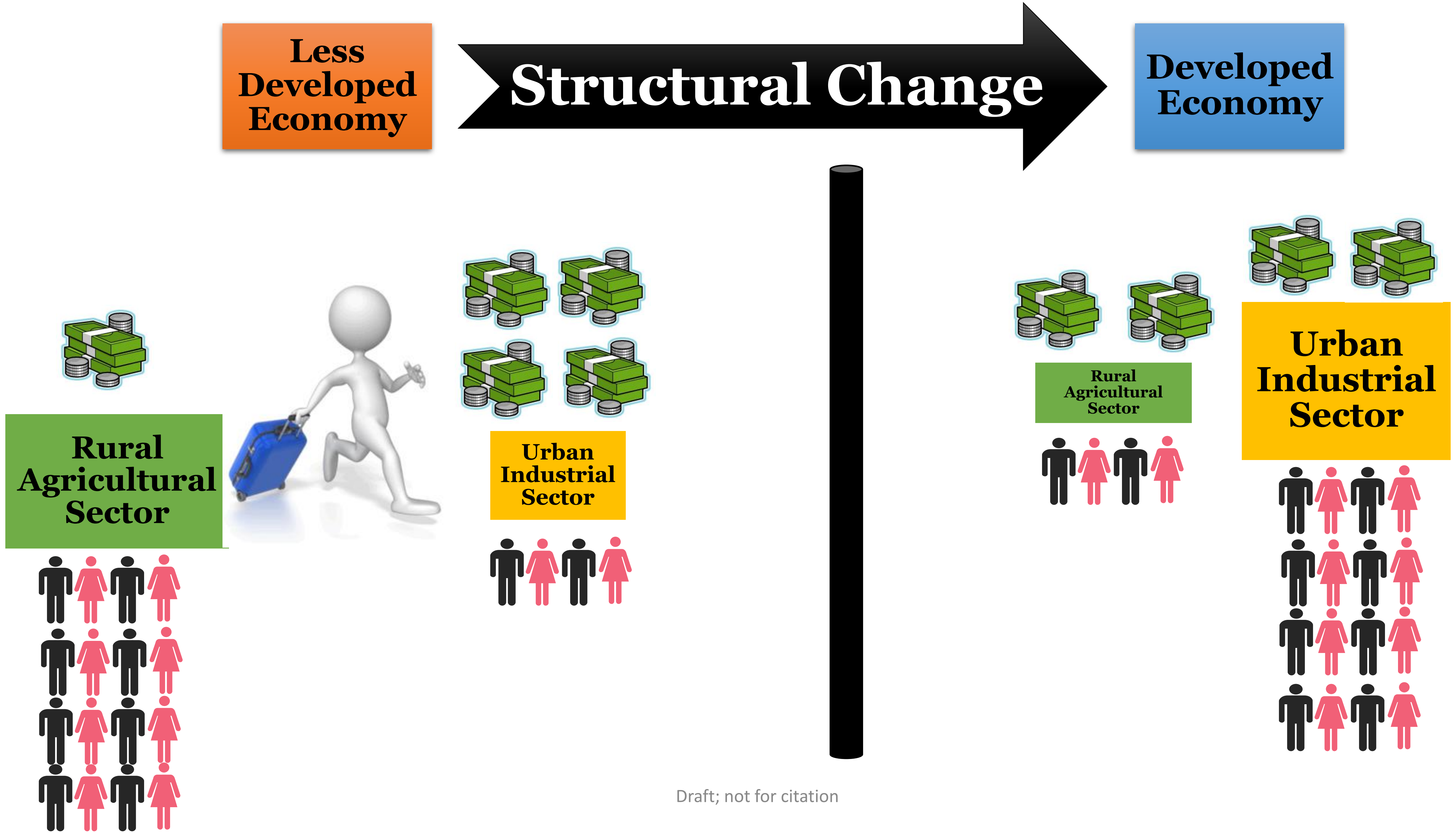
4 QUALITY EDUCATION	8 DECENT WORK AND ECONOMIC GROWTH	10 REDUCED INEQUALITIES	17 PARTNERSHIPS FOR THE GOALS
TARGET 4·4	TARGET 8·6	TARGET 10·1	TARGET 17·8
INCREASE THE NUMBER OF PEOPLE WITH RELEVANT SKILLS FOR FINANCIAL SUCCESS	PROMOTE YOUTH EMPLOYMENT, EDUCATION AND TRAINING	REDUCE INCOME INEQUALITIES	STRENGTHEN THE SCIENCE, TECHNOLOGY AND INNOVATION CAPACITY FOR LEAST DEVELOPED COUNTRIES



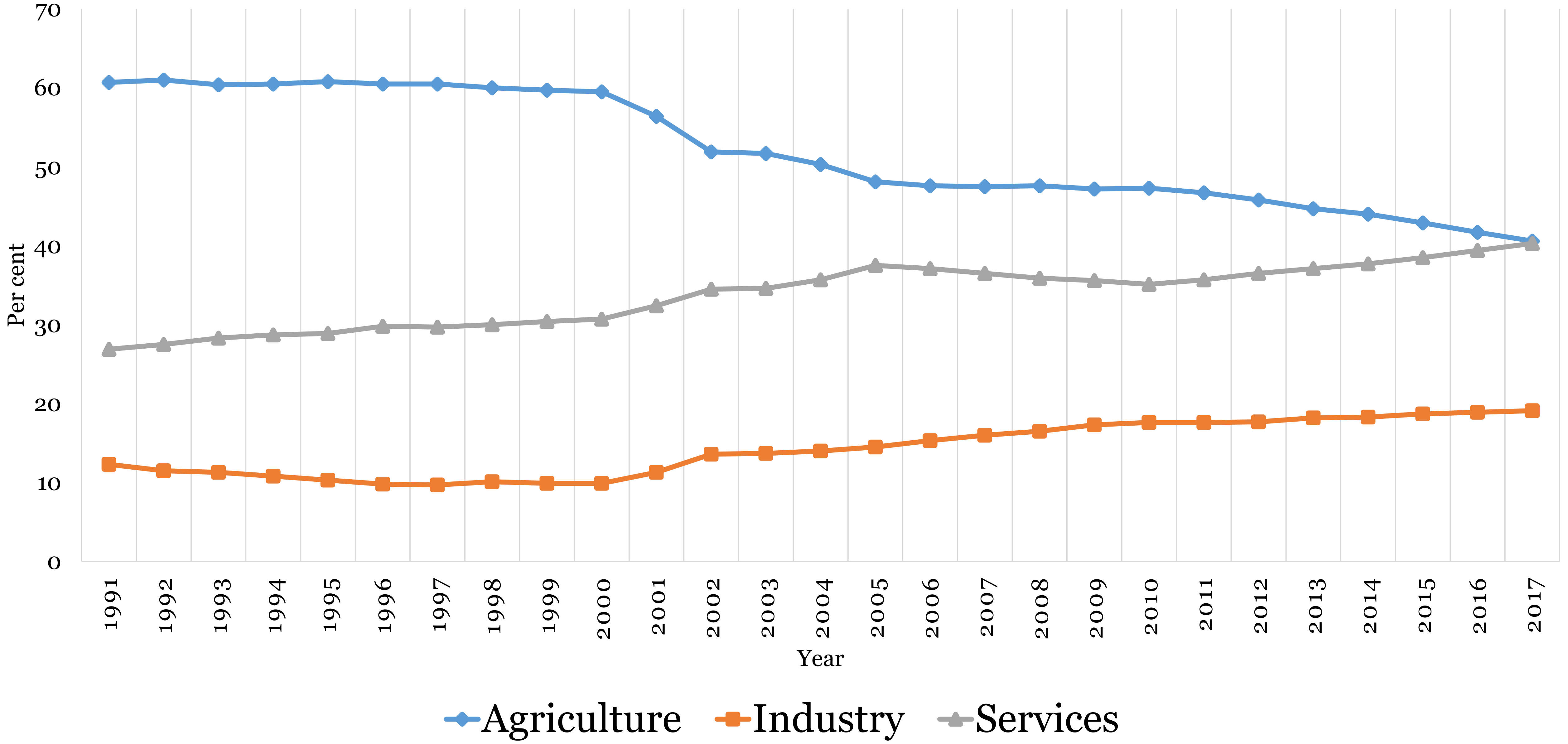
Source: Bangladesh Computer Use and Access Survey 2013, Bangladesh Bureau of Statistics (BBS)

Region \ Wages	Never used computer (in BDT)	Used computer (in BDT)	Wage differential (in %)
Barishal	11,343	15,731	38.68
Chattogram	11,217	14,358	28.00
Dhaka	10,881	14,788	35.91
Khulna	10,709	13,122	22.53
Rajshahi	10,620	13,425	26.41
Rangpur	10,264	13,551	32.02
Sylhet	10,320	13,401	29.85
Rural	10,287	12,925	25.64
Urban	11,254	14,434	28.26
National	10,812	14,162	30.98

Source: Bangladesh Labour Force Survey 2013 and Bangladesh Computer Use and Access Survey 2013, Bangladesh Bureau of Statistics (BBS)


Less Developed Economy
Structural Change
Developed Economy
Rural Agricultural Sector
Urban Industrial Sector
Rural Agricultural Sector
Urban Industrial Sector

Sector-wise employment (as percentage of total employment) in Bangladesh (1991-2017)

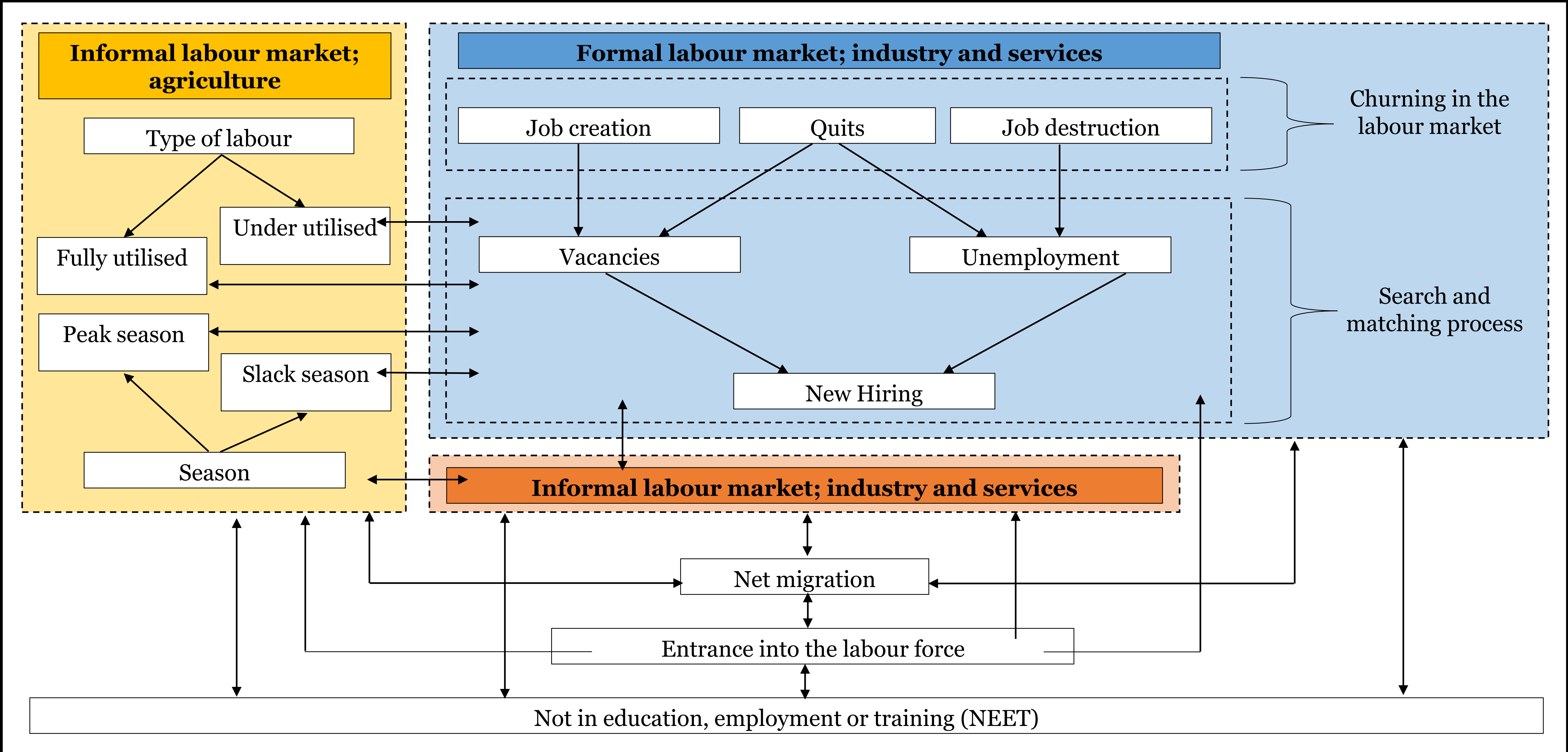


Source: ILO modelled estimates, ILOSTAT

Draft;
not for citation

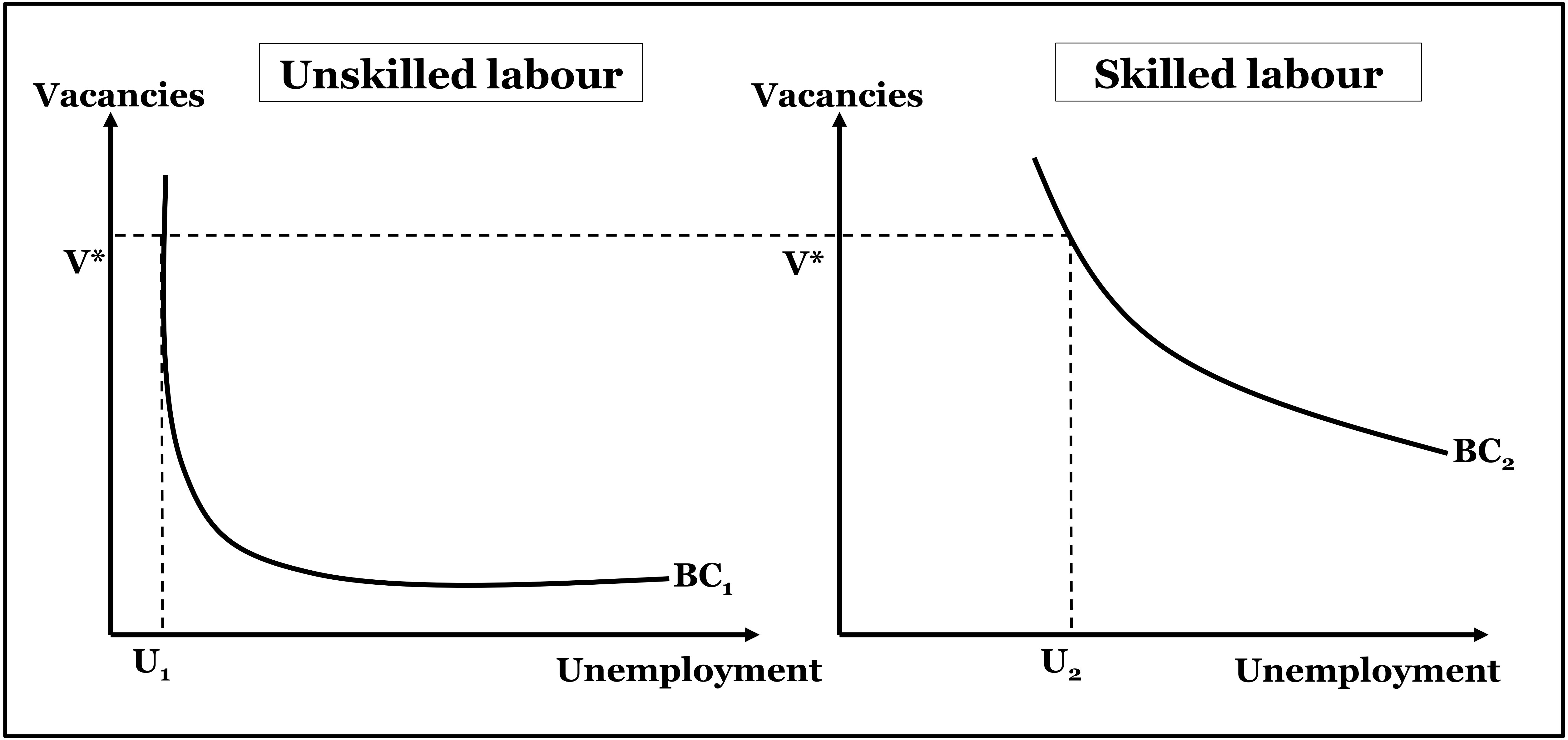
Theoretical Framework

Dynamics of Labour Market of Bangladesh



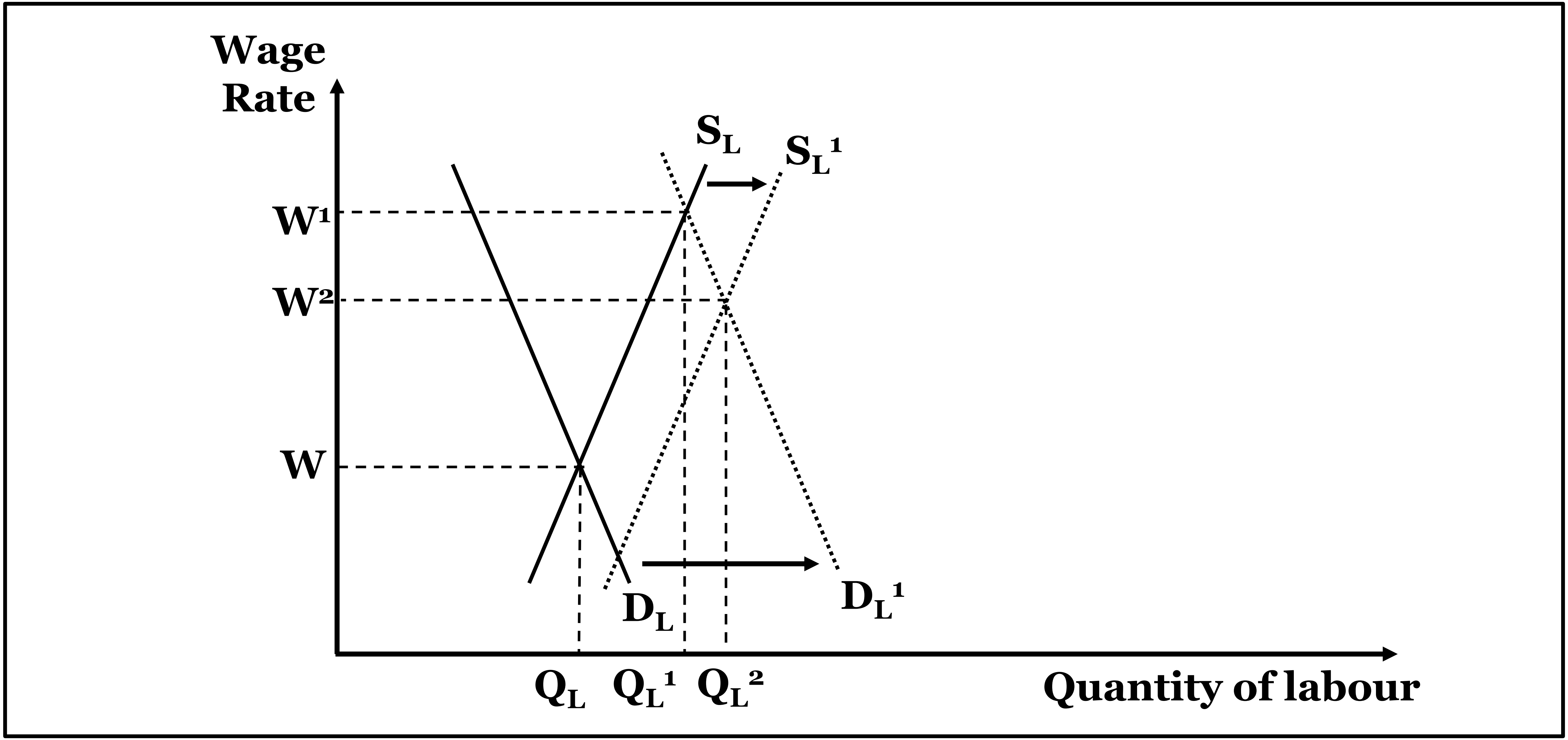
Source: Authors' illustration based on Bleakley & Fuhrer (1997).
Note: i) For simplicity of exposition, formal employment in agriculture sector is not shown.

Beveridge Curves for Unskilled and Skilled Labour Markets



Source: Authors' illustration based on (Beveridge, 1944)

Labour Market for Workers with Computer Skills



Source: Authors' illustration

Draft; not for citation

Literature Review

40 per cent of the returns to education could be attributed to the propagation of computer use. (Krueger, 1993)

Spread of computer technology may explain 30 to 50 per cent of the increase in the rate of growth of the wage of skilled workers. (Autor, et al., 1996)

Use of white collar tools yielded a wage premium, whilst the use of blue collar tools resulted in a wage penalty. (DiNardo & Pischke, 1997)

Men who used computers in the year 2000 earned 5 per cent more, whilst women who used computers in the year 2000 earned 14 per cent more. (Dolton & Makepeace, 2004)

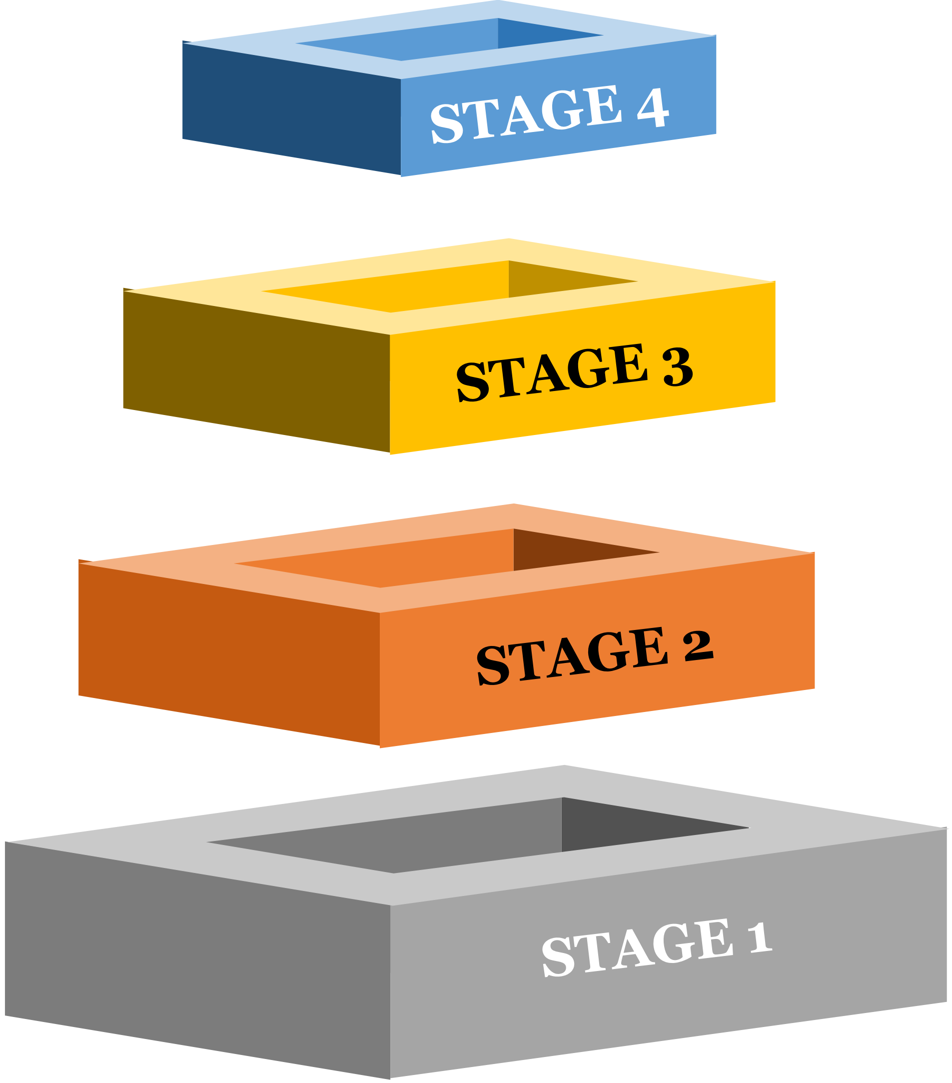
Returns to Computer Use in Previous Studies

Author	Year	Country	Returns to computer use
Alan B. Krueger	1993	United States of America	10% to 15%
David Autor, Lawrence F. Katz, Alan B. Krueger	1984	United States of America	17%
David Autor, Lawrence F. Katz, Alan B. Krueger	1989	United States of America	19%
David Autor, Lawrence F. Katz, Alan B. Krueger	1993	United States of America	20%
John E. DiNardo, Jörn-Steffen Pischke	1979	West Germany	11%
John E. DiNardo, Jörn-Steffen Pischke	1985-1986	West Germany	16%
John E. DiNardo, Jörn-Steffen Pischke	1991-1992	West Germany	17%
Barton Hughes Hamilton	1980	United States of America	13% to 25%
Peter Dolton, Gerry Makepeace	1991 to 2000	United Kingdom	14% for men; 9% for women

DATA

Cross sectional data	**All variables from LFS 2013, except CPI (which is from Bangladesh Bank)**

Sampling Strategy



Systematic random sampling of clusters of 24 households from each of the 1512 PSUs/EAs. 36,242 households are selected at this stage.

Random selection of PSUs/EAs from all of the 64 districts and 21 regional strata.

7 divisions: Barisal, Chittagong, Dhaka, Khulna, Rajshahi, Rangpur, and Sylhet. 3 kinds of localities: City Corporation, Urban, and Rural.

1512 primary sampling units (primary sampling units) or enumeration areas (EAs). Each PSU/EA had approximately 80-120 households.

Sample Size Calculation Formula

$$n = \left[\frac{(1-p)}{p} * \left(\frac{z\left(\frac{\alpha}{2}\right)}{r} \right)^2 \right] * deff$$

where,

p = apriori proportion of the required characteristics in the population
$z\left(\frac{\alpha}{2}\right)$ = value of the standard normal variate allowing $100(1-\alpha)\%$p confidence
r = rate of allowable margin of error
N = population size
deff = design effect used in complex surveys using multistage cluster sampling

And assuming,

$\alpha = 0.005$
deff = 2
p = 0.046 (from Labour Force Survey 2010)

Variables

Variable	Definition
lnwage	natural log of weekly wage in cash and kind from both primary and secondary job
education	years of schooling up to 12 years 1 = class-I 7 = class-VII 2 = class-II 8 = class-VIII 3 = class-III 9 = class-IX 4 = class-IV 10 = class X 5 = class-V 11 = SSC 6 = class-VI 12 = HSC
experience	potential experience; (experience = [age] – [education] – [6])
experience²	squared potential experience term; (experience² = experience*experience)
computer	computer use dummy; computer = 1 if ever used computer computer = 0 if never used computer
hours	total number of hours worked per week at both primary and secondary job
assets	total amount of land owned by households, measured in acres
married	marital status dummy; married = 1 if currently married married = 0 if currently not married
children	number of children aged less than 6 years
CPI* *(Bangladesh Bank Data)	Consumer price index (CPI); CPI = 183.90 if rural, CPI = 177.71 if urban

Econometric Model

Draft; not for citation

Labour force participation model specification

$$ln(W_i^*) = \beta_0 + \beta_1 h_i + \beta_2 Ai + \beta_3 M_i + \beta_4 K_i + \beta_5 P_i + \varepsilon_i \quad (1)$$

where,

$\ln(W^*)$	=	natural log of latent wage; $W^* = 1$ if $W > 0$ (employment indicator)
h	=	hours of work
A	=	assets owned by the household
M	=	marital status dummy variable
K	=	number of children aged less than six
P	=	vector of goods prices; consumer price index

Market wage model specification

$$\ln(W_i) = b_0 + b_1 S_i + b_2 E_i + b_3 E_i^2 + b_4 C_i + u_i \quad (2)$$

where,

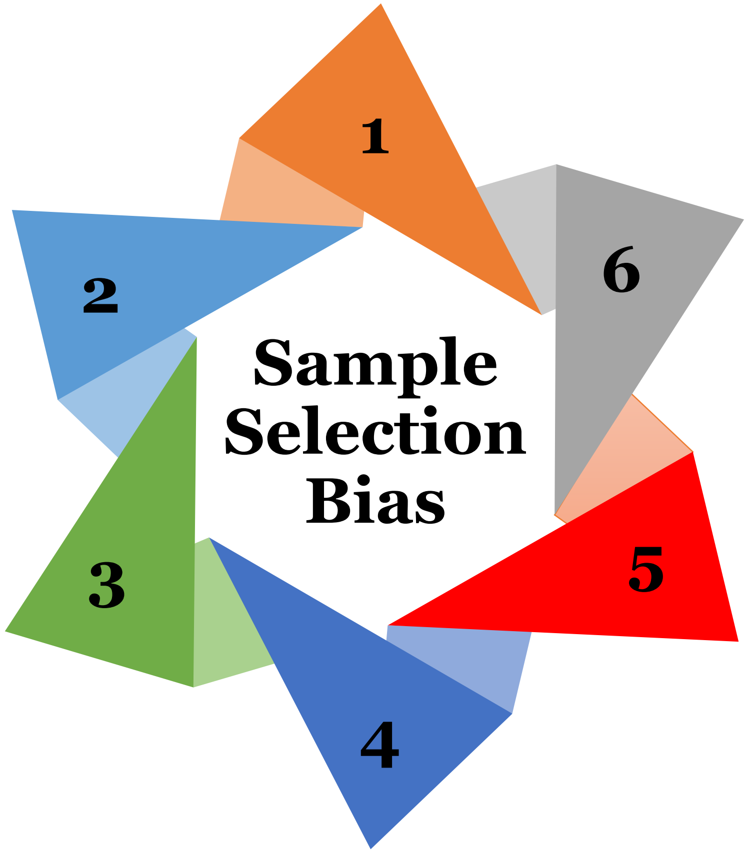
$\ln(W_i)$	=	natural logarithm of market wage rate
S	=	number of years of schooling
E	=	potential labour market experience
E^2	=	potential experience squared
C	=	computer use dummy variable

Methodology

Equation (2) suffers from unobserved heterogeneity or the problem of omitted variables.

The effect of these unobserved variables is captured through the error terms, and so the errors of the equation (4) are correlated with the independent variables.

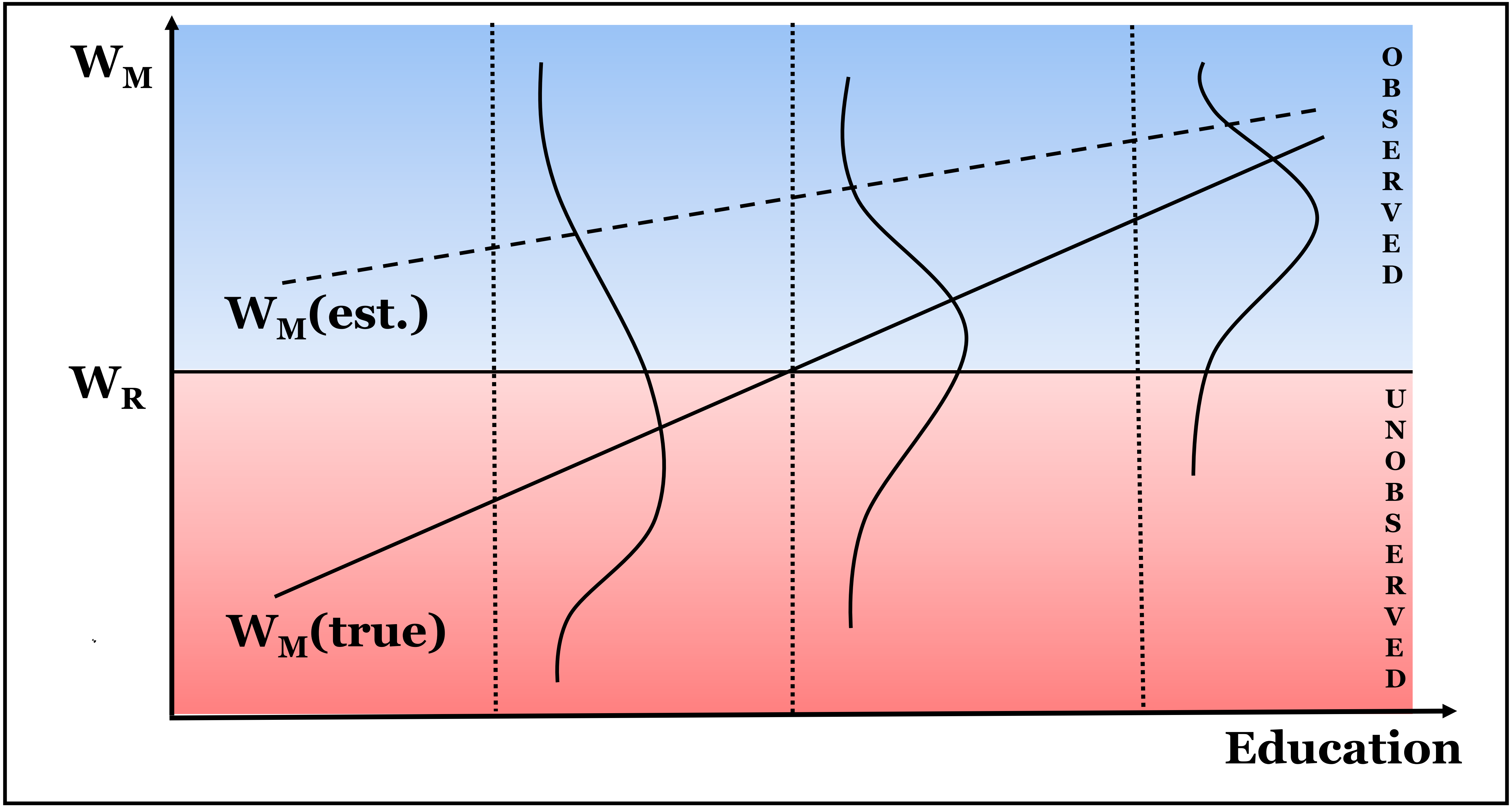
The underlying reason behind this is the fact that the samples used for estimating these equations were not randomly collected.



Market wages are only observed for individuals who are working.

However, an individual's decision to work is not a random decision, but rather a rational choice.

Hence individuals who choose to work self-select themselves into the sample.



*Author's illustration based on (Smith, 2014)

Strict exogeneity assumption of the OLS model is $E\left(\varepsilon_i|X\right) = 0, \quad \forall\, i = 1, \ldots, n$

Violation of the strict exogeneity assumption has several implications:

- $E\left(\varepsilon_i\right) \neq 0, \forall\, i = 1, \ldots, n$

(The unconditional mean of the error term (ε) is not zero.)

- $E\left(X_{jk}, \varepsilon_i\right) \neq 0, \forall\, ijk = 1, \ldots, n$

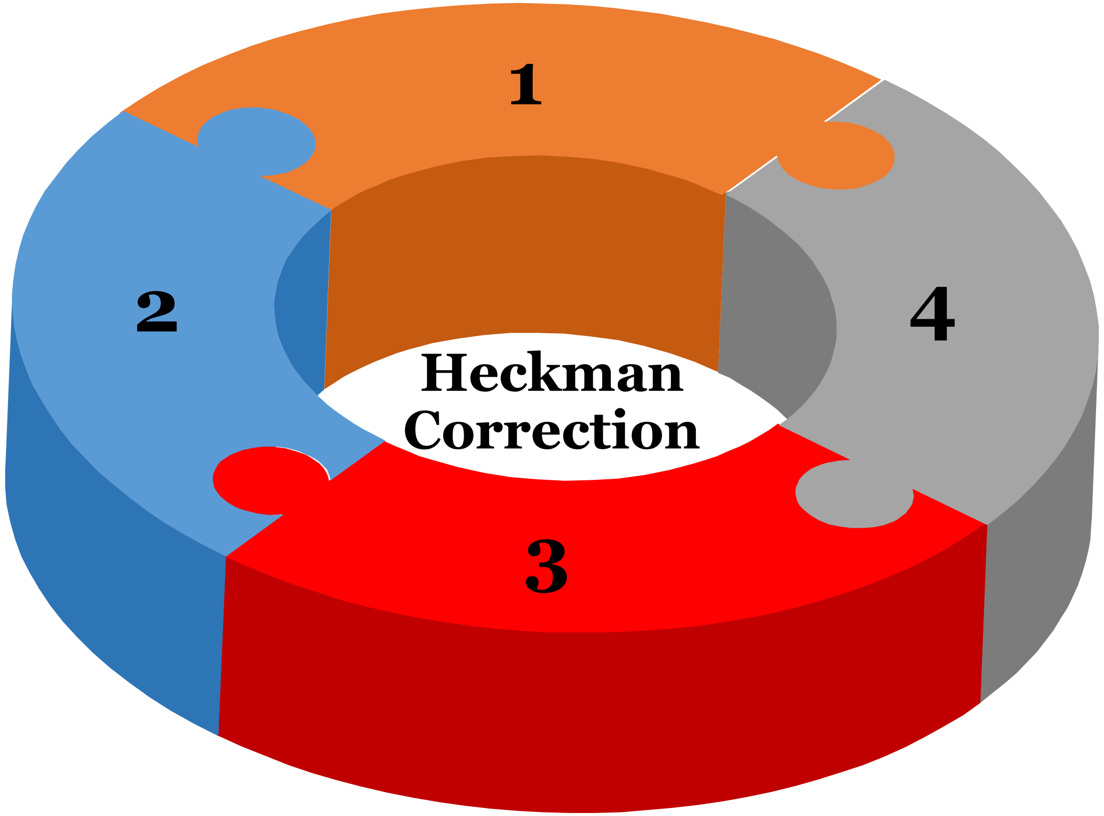
(The independent variables (X) are not orthogonal to the errors (ε) for all observations)

- $COV\left(X_{jk}, \varepsilon_i\right) \neq 0, \forall\, ijk = 1, \ldots, n$

(The independent variables (X) and errors (ε) are not uncorrelated for all observations.)

If the unobserved heterogeneity can be modelled separately, and the resulting information can be incorporated into the main model, then the problem can be resolved.

Heckman proposed that the specification of the original biased model could be improved by using the estimated values of the omitted variables as additional regressors.



Heckman outlined an ingenious two step estimation technique to correct sample selection bias (Heckman, 1979).

By doing so, the model could be estimated using ordinary least squares, without violating the strict exogeneity assumption.

The factors that influence an individual's decision to work are modelled by using a probit model. The general form of the sample likelihood function for this probit analysis is:

$$\mathcal{L} = \prod_{i=1}^{T} [F(\phi_i)]^{1-d_i} [1 - F(\phi_i)] d_i$$

where, d is a random variable, which is equal to one if the dependent variable is observed and equal to zero if the dependent variable is not observed.

Suppose there is a sample of T individuals, K of who work and T-K who do not work.

Then, the aforementioned likelihood function becomes:

$$\boldsymbol{\mathcal{L}} = \prod_{i=1}^{K} j(h_i, ln(W_i)|(W_i > W_i^*)_{h=0}).pr([W_i > W_i^*]_{h=0}) \times \prod_{i=K+1}^{T} pr([W_i < W_i^*]_{h=0})$$

$$\text{Inverse Mills Ratio} = \frac{\text{standard normal probability distribution function}}{\text{standard normal cumulative distribution function}}$$

$$\boldsymbol{\lambda_i = \frac{f(\phi_i)}{1 - F(\phi_i)}}$$

where,

$\boldsymbol{\lambda}$ = inverse Mills ratio

f = standard normal probability distribution function of the selection equation

F = standard normal cumulative distribution function of the selection equation.

The Inverse Mills Ratio can be defined as:

$$\boldsymbol{\lambda = j(h_i, \ln(W_i) | (W_i^* < W_i)_{h=0} = \frac{n(h_i, \ln(W_i))}{pr([W_i > W_i^*]_{h=0})} \because \varepsilon_i, u_i \sim N(0)}$$

Using this Inverse Mills Ratio in the original likelihood function simplifies to:

$$\boldsymbol{\mathcal{L} = \prod_{i=1}^{K} n(h_i, ln(W_i)) \prod_{i=K+1}^{T} pr([W_i < W_i^*]_{h=0})}$$

Maximizing this likelihood function with respect to the parameters of the model, including the variances and covariances of the errors in equations (1) and (2) yields **consistent**, asymptotically **unbiased**, and **efficient** parameter estimates which are asymptotically normally distributed.

Thus, the selection bias corrected now becomes:

$$\ln(W_i) = b_0 + b_1 S_i + b_2 E_i + b_3 \lambda_i + u_i (3)$$

Augmenting the basic model with the squared experience term and computer use dummy variable gives:

$$\ln(W_i) = b_0 + b_1 S_i + b_2 E_i + b_3 E_i^2 + b_4 C_i + b_5 \lambda_i + u_i (4)$$

where, λ is the inverse Mills ratio

Results

$$\ln(W_i) = b_0 + b_1 S_i + b_2 E_i + b_3 E_i^2 + b_4 C_i + u_i \quad (2)$$

	Model without computer	Model with computer
Regression	**OLS**	**OLS**
Variable	**lnwage**	**lnwage**
education	0.0345818***	0.0296701***
	(0.0008388)	(0.00087)
experience	0.020699***	0.0213602***
	(0.0006296)	(0.0006244)
experiencesq	-0.0003208***	-0.0003276***
	(0.0000124)	(0.0000123)
computer		0.1928568***
		(0.010153)
Constant	7.29877***	7.312575***
	(0.0095537)	(0.0094888)
F stat	928.38	800.22
Prob > F	0.0000	0.0000
R-squared	0.1324	0.1492
Adj R-squared	0.1322	0.1490

Note: (i) Standard errors in parentheses (ii) *** p<0.01, ** p<0.05, * p<0.1

Ramsey Regression Specification Error Test (RESET)

Null hypothesis = model is correctly specified
Alternative hypothesis = model is incorrectly specified

Decision rule: if $p < 0.05$ then the model is incorrectly specified

Statistic	Model without computer Test statistic (Probability)	Model with computer Test statistic (Probability)
F	588.80 (0.0000)	516.56 (0.0000)

Note: (i) H_0: there is no omitted variable, H_A: there is at least one omitted variable

Interpretation: The model is incorrectly specified

Link Test

- Link Test is based on the idea that if a regression is properly specified, one should not be able to find any additional independent variables that are significant except by chance.
- Link Test creates two new variables, the variable of prediction, and the variable of squared prediction.
- We wouldn't expect the squared prediction to be a significant predictor if our model is specified correctly.

Variables	Model without computer lnwage	Model with computer lnwage
Prediction	51.53402***	32.39009***
	(1.67844)	(1.341881)
Squared prediction	-3.249214***	-2.013056***
	(0.107913)	(0.086048)
Constant	-196.4245***	-122.3259***
	(6.525249)	(5.230572)
F stat.	1915.11	1922.25
Prob > F	0.0000	0.0000
R-squared	0.1734	0.1740
Adj R-squared	0.1733	0.1739

Note: (i) Standard errors in parentheses; (ii) *** p<0.01, ** p<0.05, * p<0.1

Interpretation: The model is incorrectly specified

Variance Inflation Factor

Variance inflation factor measures the linear association between an independent variable and all other independent variables.

Decision rule:

VIF > 10 : perfect multicollinearity is highly likely

5 < VIF < 10 : perfect multicollinearity is somewhat likely

0 < VIF < 5 : perfect multicollinearity is unlikely

	Model without computer		Model with computer	
Variables	**VIF**	**1/VIF**	**VIF**	**1/VIF**
education	1.02	0.985135	1.11	0.898111
experience	10.00	0.100022	10.03	0.099712
experience²	9.97	0.100328	9.98	0.100242
computer			1.11	0.897545
Mean VIF	6.99		5.56	

Note: (i) VIF > 10 : perfect multicollinearity is highly likely; 5 < VIF < 10 : perfect multicollinearity is somewhat likely; 0 < VIF < 5 : perfect multicollinearity is unlikely

Interpretation: Perfect multicollinearity is somewhat likely

Breusch-Pagan and Cook-Weisberg Test

Null hypothesis = homoskedastic,
Alternative hypothesis = heteroskedaticity

Decision rule: if $p < 0.05$ then there is heteroskedasticity.

Statistic	Model without computer Test statistic (Probability)	Model with computer Test statistic (Probability)
chi^2	107.87 (0.0000)	118.51 (0.0000)
F	55.13 (0.0000)	60.59 (0.0000)

Note: (i) H_0: errors have are homoscedastic, H_A: errors are not homoscedastic; (ii) Breusch-Pagan (1979) and Cook-Weisberg (1983) test for heteroskedasticity assumes that the heteroskedasticity is a linear function of the independent variables.

Interpretation: There is heteroskedasticity

White Test

Breusch-Pagan (1979) and Cook-Weisberg (1983) test for heteroskedasticity assumes that the heteroskedasticity is a linear function of the independent variables.

The White test allows the heteroskedasticity process to be a function of one or more independent variables. It allows the independent variable to have a non-linear and interactive effect on the error variance.

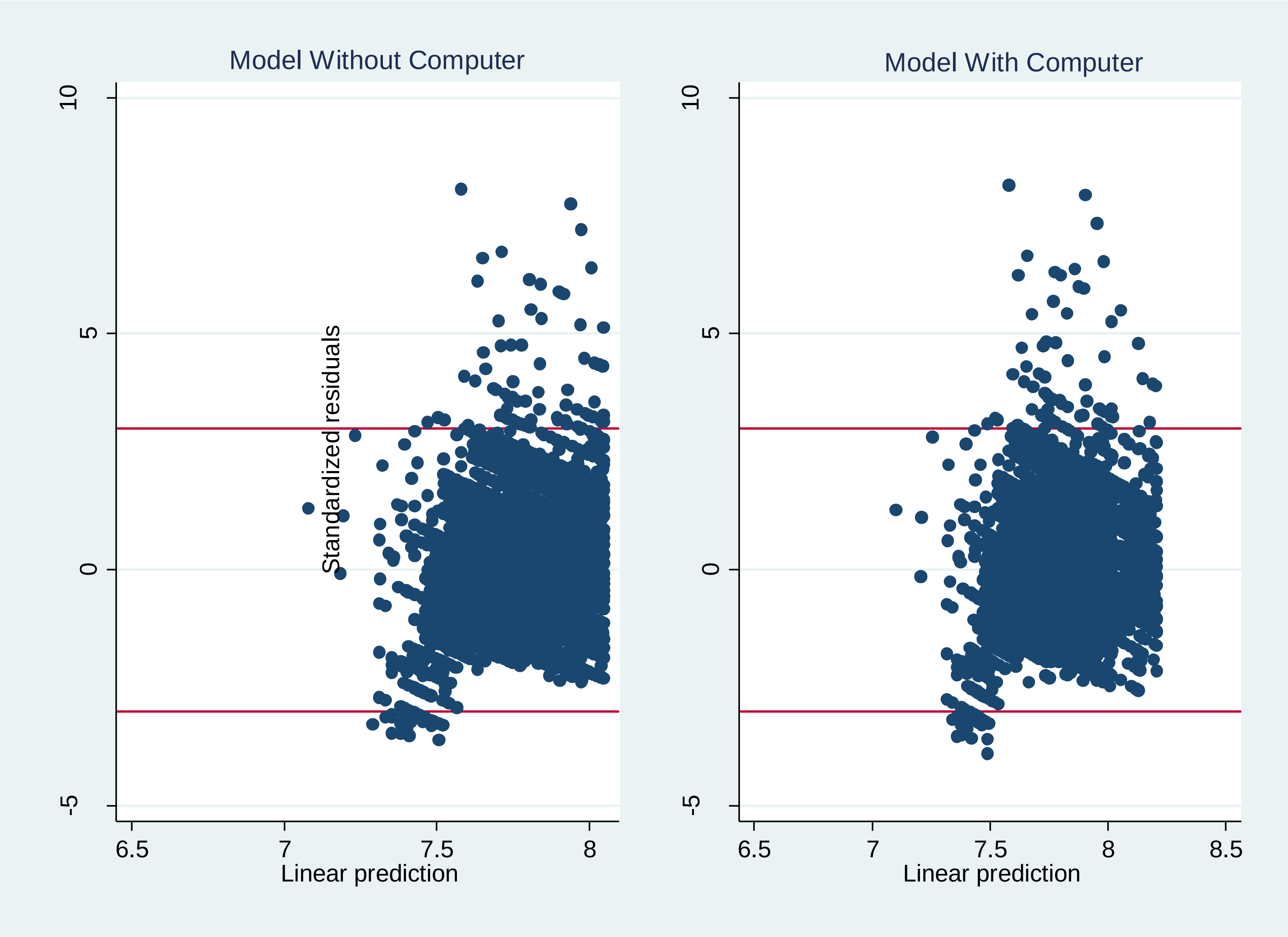
Null hypothesis = homoscedastic; Alternative hypothesis = heteroskedaticity
Decision rule: if $p < 0.05$ then there is heteroskedasticity.

Statistic	**Model without computer Test statistic (Probability)**	**Model with computer Test statistic (Probability)**
chi²	958.62 (0.0000)	962.88 (0.0000)

Note: (i) H_0: errors have are homoscedastic, H_A: errors are not homoscedastic; (ii) White test allows the heteroskedasticity process to be a function of one or more independent variables. It allows the independent variable to have a non-linear and interactive effect on the error variance.

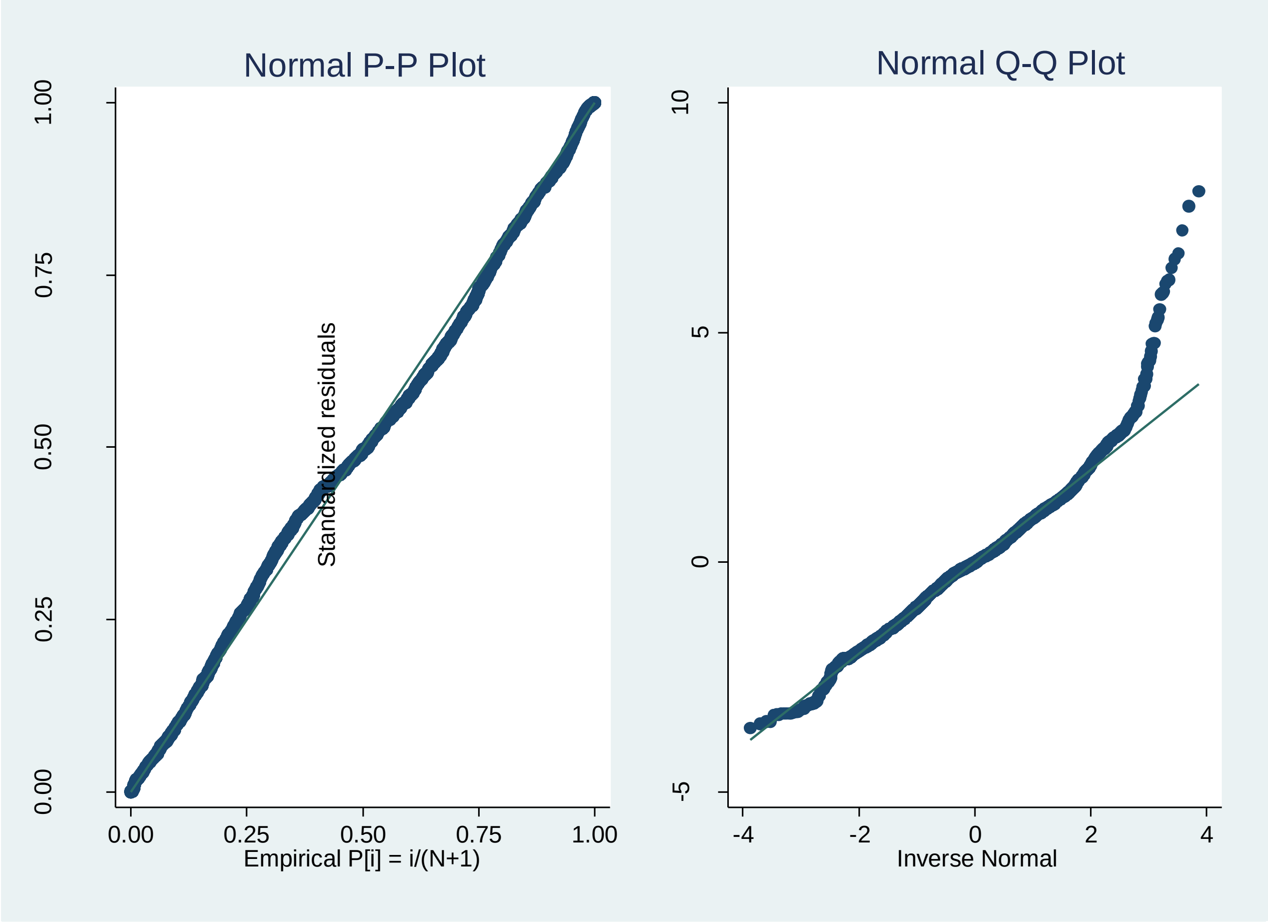
Interpretation: There is heteroskedasticity

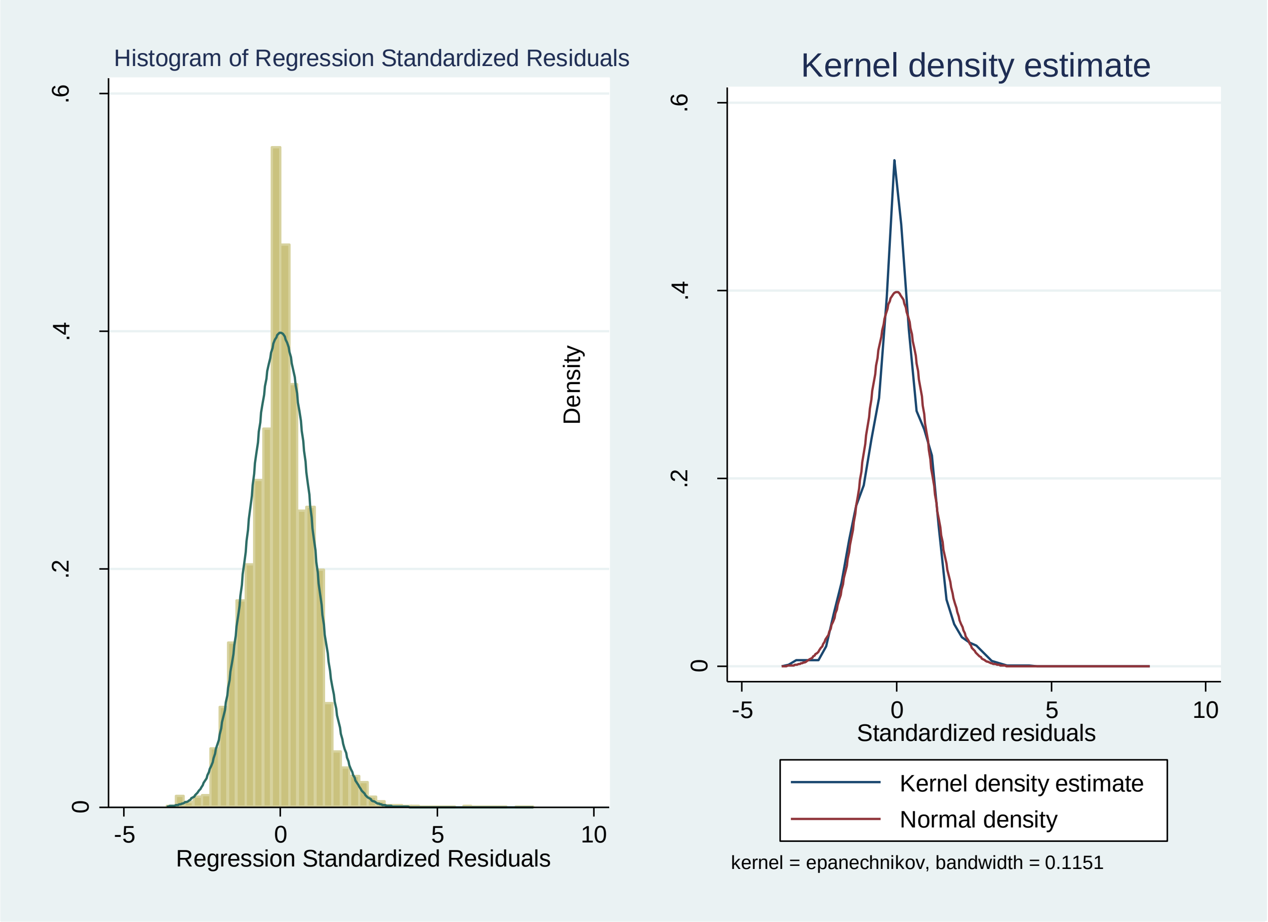
Graphical Check of Heteroskedasticity



Draft; not for citation

Normal PP Plot, Normal QQ Plot, and Distribution of Residuals for OLS Model without Computer

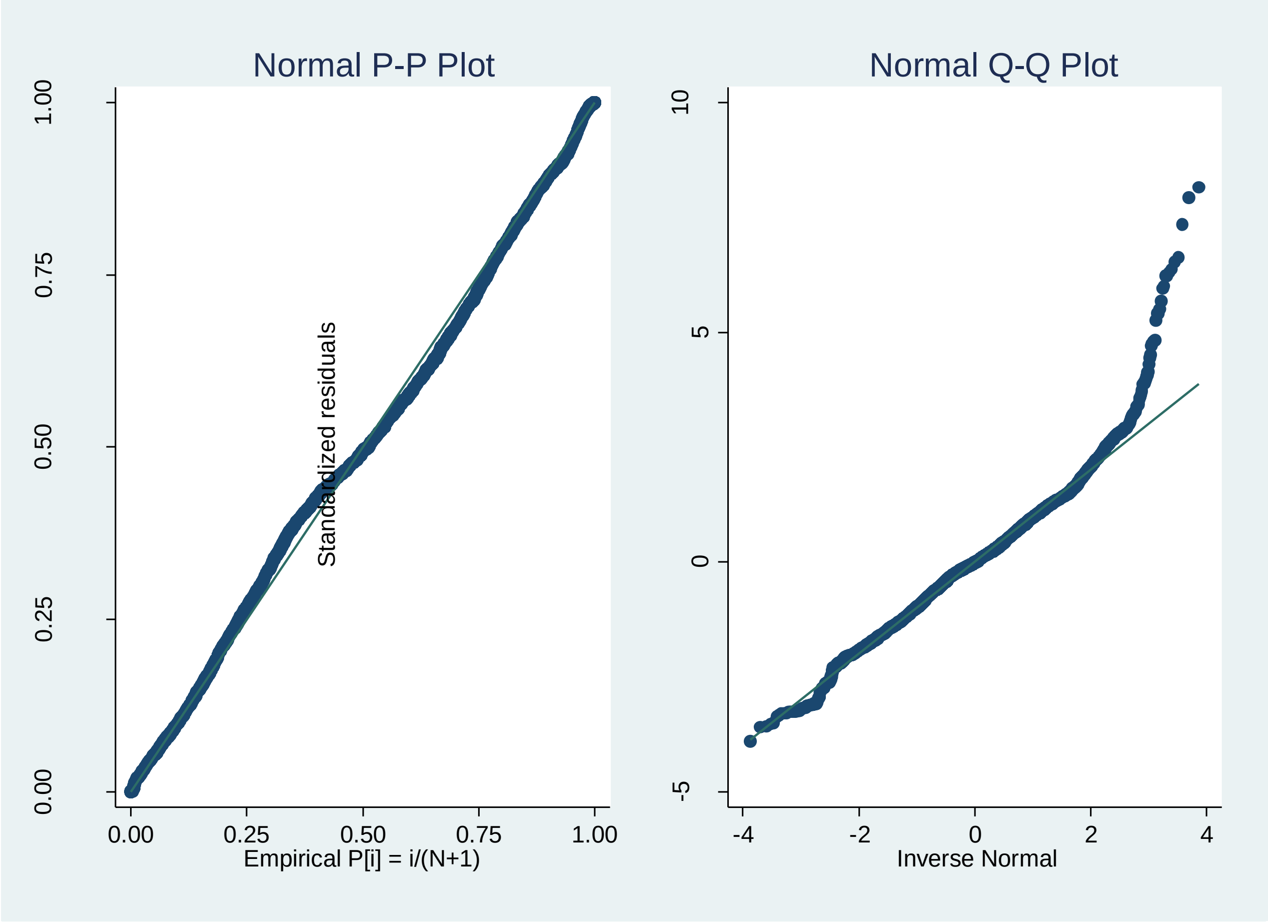


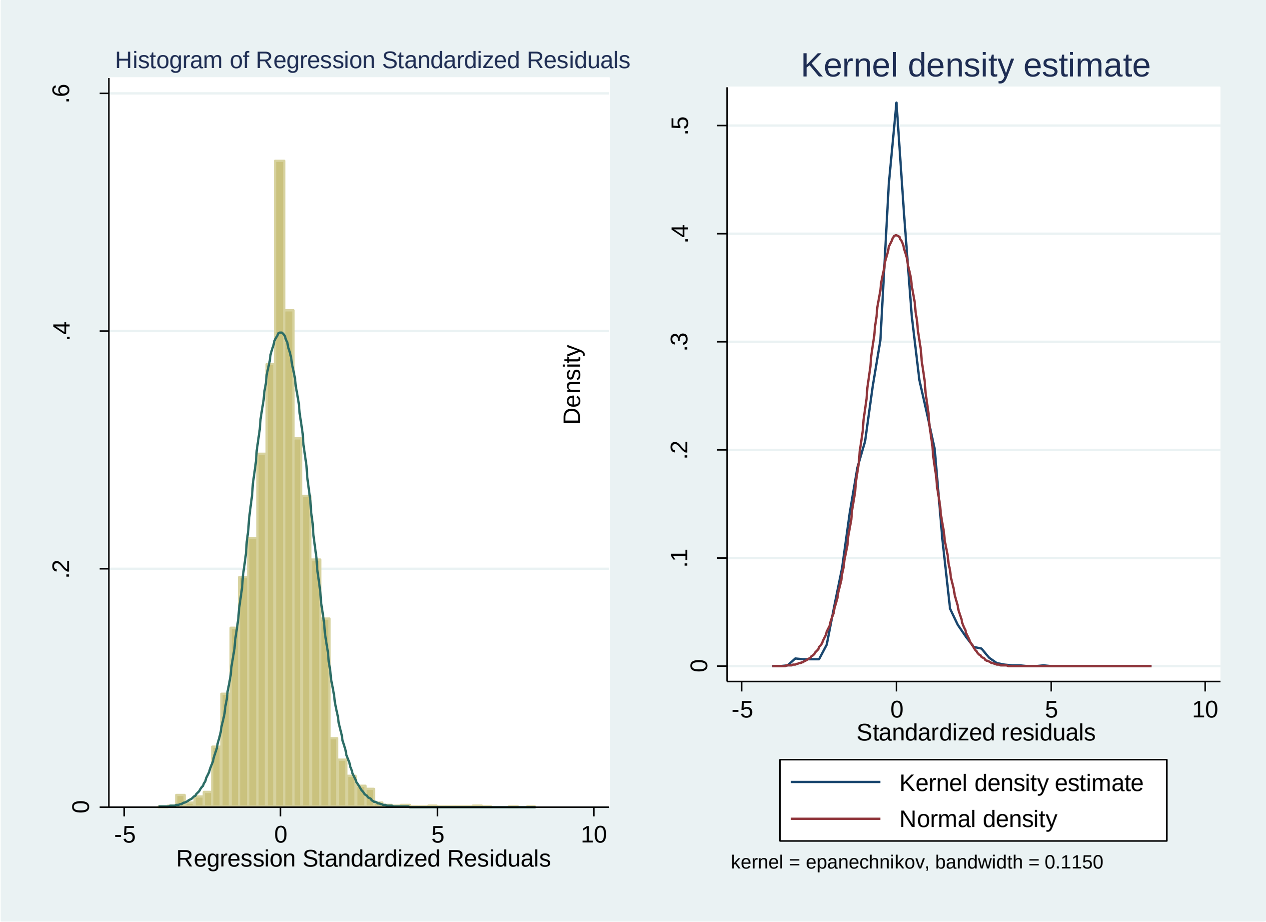


Note: (i) PP plot is more sensitive in the center; QQ plot is more sensitive at the two tails

Draft;
not for citation

Normal PP Plot, Normal QQ Plot, and Distribution of Residuals for OLS Model with Computer





Note: (i) PP plot is more sensitive in the center; QQ plot is more sensitive at the two tails

Shapiro Wilk Test

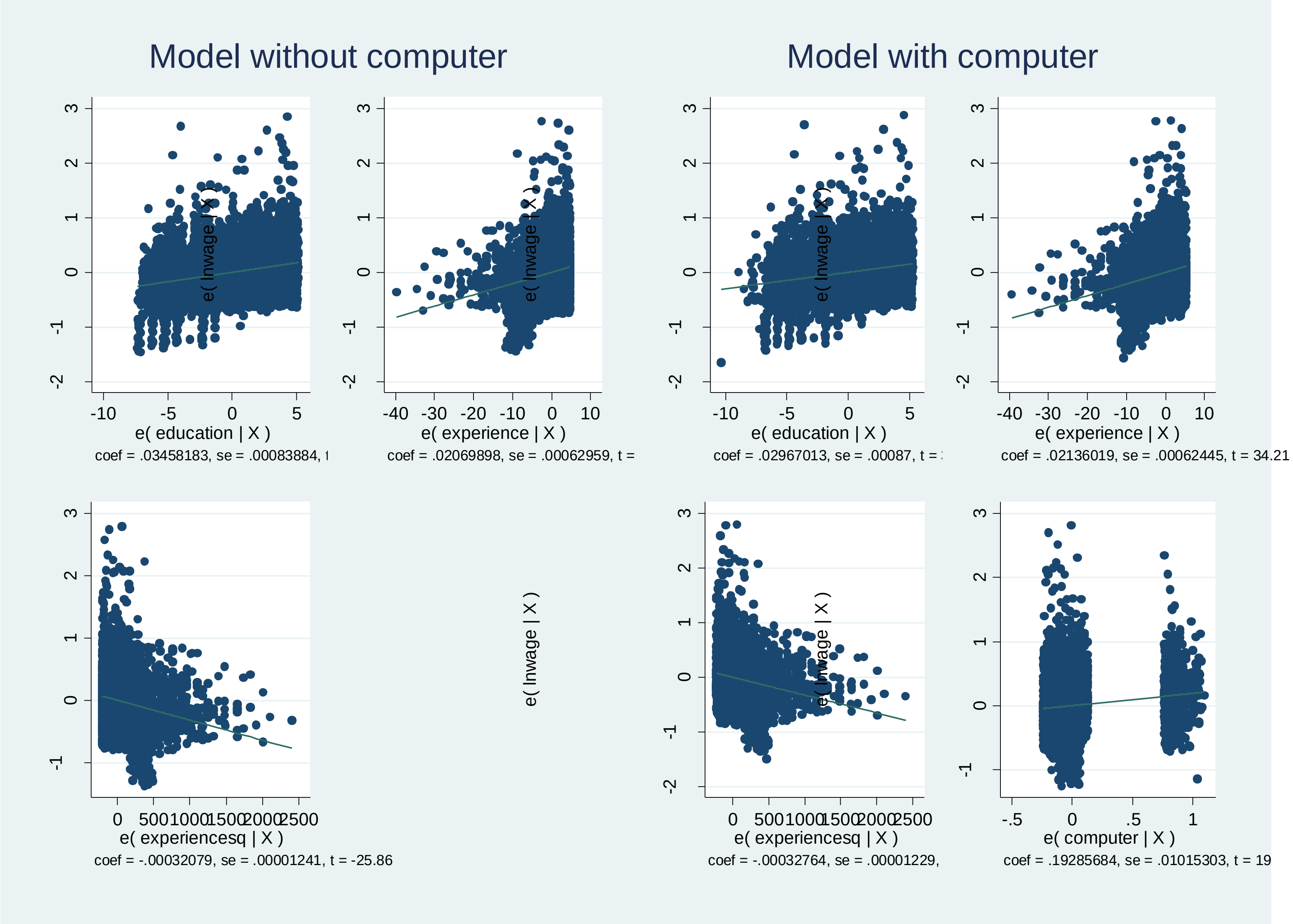
Null hypothesis = errors normal
Alternative hypothesis = errors not normal
Decision rule: If p value < 0.05 then reject null hypothesis that errors are normal.
If p value > 0.05 then cannot reject null hypothesis that errors are normal.

Variable	Model without computer Residual	Model with computer Residual
W	0.98491	0.98579
V	125.125	117.833
z	13.125	12.962
Prob > z	0.00000	0.00000

Note: (i) H_0: errors are normally distributed, H_A: errors are not normally distributed

Interpretation: The joint distribution of the errors is not normal.

Graphical Check of Outliers in OLS Models



Summary of Post-estimation Diagnostic Tests

Test	Description	Result
Ramsey Regression Specification Error Test	Test of model specification	Model is incorrectly specified; there is at least one omitted variable
Link Test	Test of model specification	Model is incorrectly specified
Variance Inflation Factor	Test of multicollinearity	Perfect multicollinearity is somewhat likely
Breusch-Pagan (1979) and Cook-Weisberg (1983) Test	Test of heteroskedasticity	There is heteroskedasticity
White Test	Test of heteroskedasticity	There is heteroskedasticity
Shapiro Wilk Test	Test of normality of errors	The errors are not normally distributed

Results from Heckman Two-step Estimation

	Model without computer		Model with computer	
Regression Variable	**Probit lnwage**	**Heckman lnwage**	**Probit lnwage**	**Heckman lnwage**
education		0.0343627*** (0.0010397)		0.0302509*** (0.0010727)
experience		0.0198216*** (0.0007809)		0.0203444*** (0.0007734)
experience²		-0.0002861*** (0.0000151)		-0.0002924*** (0.000015)
computer				0.1702658*** (0.0123785)
hours	0.0107224*** (0.0005879)		0 .0107224*** (0.0005879)	
assets	-0.0003627*** (0.0000306)		-0.0003627*** (0.0000306)	
married	-0.3196878*** (0.0171752)		-0.3196878*** (0.0171752)	
children	-0.0234653** (0.00963)		-0.0234653** (0.00963)	
CPI	-0.041563*** (0.00241)		6.955875*** (0.4395531)	
lambda		-0.1805037 (0.0128445)		-0.1558748 (0.0133838)
Constant	6.955875*** (0.4395531)	7.480404*** (0.0181366)	6.955875*** (0.4395531)	7.467818*** (0.0182261)
LR chi²	1730.34		1730.34	
Wald chi²		1794.51		2030.21
Prob > chi²	0.0000	0.0000	0.0000	0.0000

Note: (i) Standard errors in parentheses (ii) *** p<0.01, ** p<0.05, * p<0.1

Summary of Results from Heckman Two-step Estimation

Each additional year of schooling

Increases wages by

3.03%

Each additional year of experience

Increases wages by

2.03%

Each additional year of squared experience

Decreases wages by

0.03%

Ability to use computers

Increases wages by

17.02%

Draft; not for citation

Threshold Level of Experience

Let us recall that our model specification was as follows:

$$ln(W_i) = b_0 + b_1S_i + b_2E_i + b_3E_i^2 + b_4C_i + b_5\lambda_i + u_i(2)$$

Substituting the coefficients from the Heckman model with computer use, we get

$$\widehat{ln(W_i)} = 7.467818 + 0.0302509S_i + 0.0203444E_i + (-0.0002924)E_i^2 + 0.1702658C_i + (-0.1558748)\lambda_i + u_i$$

Differentiating the equation with respect to experience we get

$$\frac{\partial W}{\partial E} = 0.0203444 - 0.0005848\,E$$

At the turning point the first derivative is zero, so we get

$$0.0203444 - 0.0005848\,E = 0$$

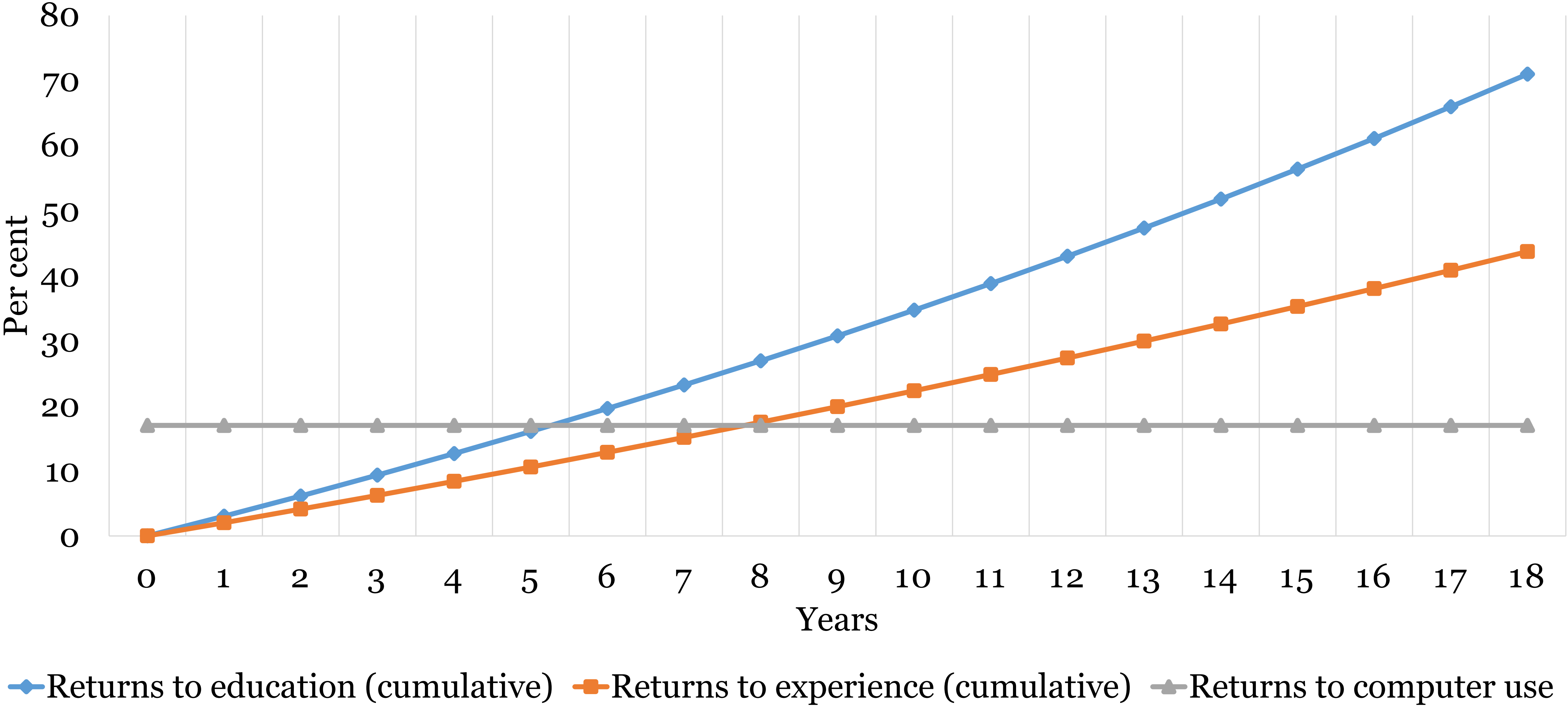
$$-0.0005848\,E = -0.0203444$$

$$E = \frac{0.0203444}{0.0005848}$$

$$E = 34.78864569$$

Thus wages are maximized at 34 years of potential experience. The second derivative is negative, further confirming the inverted U shaped nature of the relationship.

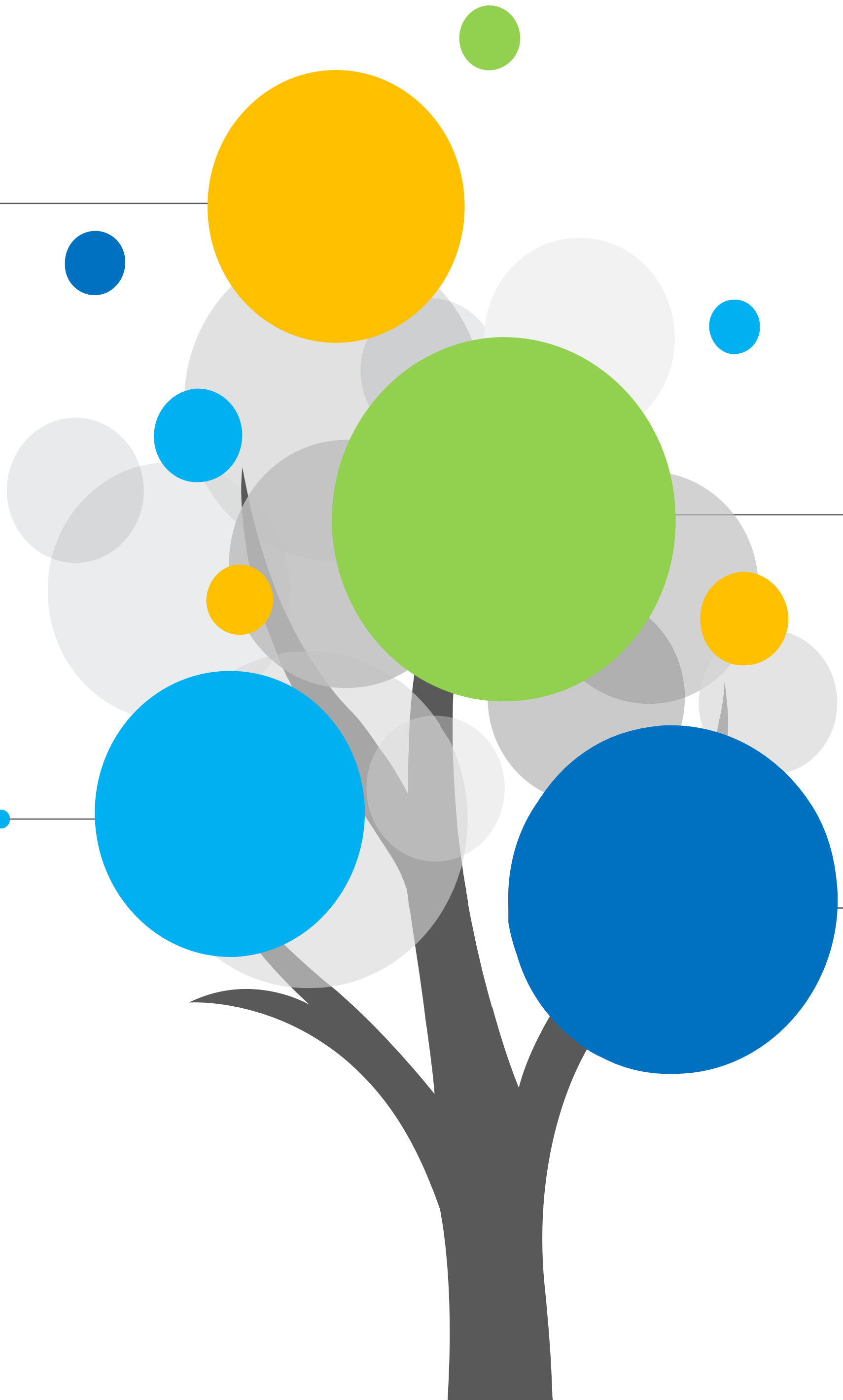
Returns to education, experience, and computers



CONCLUSION

Gaps in the labour market of Bangladesh need to be bridged urgently

Education alone cannot bridge the gaps in the labour market



Structural unemployment is now set to become the next big development challenge for Bangladesh

Computer skills are complementary to education, and can play an instrumental role in bridging the gaps in the labour market in Bangladesh

Recommendations

STUDENTS: invest time in learning computer skills

TEACHERS: increase the use of computers in the classroom

EMPLOYERS: focus on workers' computer skills for capacity building

GOVERNMENTS: allocate government resources for computer training

THANK YOU

https://cpd.org.bd

cpd.org.bd

cpdbd

CPDBangladesh

Centre for Policy Dialogue (CPD)