METHODOLOGICAL APPENDIX

Earnings inequality and the changing nature of work

Evidence from Labour Force Survey data of Bangladesh

Sayema Haque Bidisha,[1] Tanveer Mahmood,[2] and Mahir A. Rahman[2]

January 2021

WIDER Working Paper 2021/7

1.1 Methodological background of RTI analysis

In order to examine the effect of the evolution of routine task intensity (RTI) of occupations on changes in employment and earnings over time while following Sebastián (2018), we estimated the following equations (Equation 1 and 2) where RTI_i captures the time-variant routine task intensity of occupation *i*:

$$\Delta \log(E_{i,t}) = \pi_0 + \pi_1(RTI_i) + \pi_2(RTI_i)^2 \quad (1)$$

$$\Delta \log(y_{i,t}) = \rho_0 + \rho_1(RTI_i) + \rho_2(RTI_i)^2 \quad (2)$$

1.2 Regression method of job polarization

To test for polarization, we applied a regression-based test of job and earnings polarization, following Goos and Manning (2007) and Sebastián (2018). The following equations (Equation 3 and 4) have been estimated in this regard with the quadratic specification of log of mean earnings at 3 digit-occupational classification:

$$\Delta \log(E_{i,t}) = \beta_0 + \beta_1 \log(y_{i,t-1}) + \beta_2 \log(y_{i,t-1})^2 \quad (3)$$

[1] Department of Economics, University of Dhaka, Dhaka, Bangladesh, corresponding author: sayemabidisha@gmail.com

[2] Bangladesh Institute of Development Studies (BIDS), Dhaka, Bangladesh

$$\Delta \log(y_{i,t}) = \gamma_0 + \gamma_1 \log(y_{i,t-1}) + \gamma_2 \log(y_{i,t-1})^2 \quad (4)$$

Here $\Delta \log(E_{i,t})$ is the change in log of employment share of occupation *i* between survey wave *(t-1)* and *t*, $\Delta \log(y_{i,t})$is the change in log mean labour earnings in occupation *i* between survey wave *(t-1)* and *t*. As for the explanatory variables, $\log(y_{i,t-1})$ is the log of mean labour earnings in occupation *i* in survey wave *(t-1)* and $\log(y_{i,t-1})^2$ is its square. Both of these equations have been estimated by weighting each occupation *i* by its employment share at the initial survey wave to avoid any plausible biases. The sign and significance of the coefficient estimates in quadratic form can be used to test the presence of a U-shaped relationship in employment (and earnings) and broad skill classes.

2.1 Estimation method of education premium

To estimate the education premium, we followed a regression analysis, under which we regressed log weekly earnings (y_{it}) in each dataset (t) separately by sex on several regressors, e.g. dummy variables for education categories (Edu_{it}); dummy variables for age categories (Age_{it}); dummies for country-specific geographic regions (Geo_{it}), etc. and obtained the following model (Equation 5):

$$y_{it} = \alpha_t + \beta_t' Edu_{it} + \gamma_t' Age_{it} + \delta_t' Geo_{it} + \theta_t' Pop_{it} + \varepsilon_{it} \quad (5)$$

There are three versions of this model, and the final variant included all the regressors used in the other two variants and also 2-digit ISCO-88 occupation categories. This (third) version of the education premium model can therefore be considered as the most comprehensive one in terms of explaining education premium.

2.2 Methodological background of decomposition

In this connection, we used several decomposition methods, e.g. Shapley decomposition, RIF decomposition as discussed below:

Shapley decomposition

Following Chantreuil and Trannoy (1997) and Shorrocks (2013), we have decomposed earnings inequality (measured most commonly by Gini index G) into two key components: (i) changes in 'within occupation inequality' and (ii) changes in 'between occupation inequality' (as measured by two-digit ISCO-88 codes.

Let us remove within-occupation inequality and assume Y_b is the vector where earnings of individuals have been replaced by the average earnings in their occupation and Y_w is a vector where between-occupation inequality has been removed as worker earnings are re-scaled to have same earnings on an average. In this setup, $G(Y_b)$ and $[G - G(Y_w)]$ are possible estimates of the contribution of between occupation inequality and $G(Y_w)$ and $[G - G(Y_b)]$ are possible estimates for the contribution of within occupation inequality. Shapley decomposition can be given by the following equation (Equation 6) where the changes in Gini is decomposed into the contribution of each of the components:

$$G = G_B + G_W$$
$$G_B = 0.5[G(Y_b) + G - G(Y_w)]$$
$$G_W = 0.5[G(Y_w) + G - G(Y_b)] \quad (6)$$

In general, although G_B will be lower than $G(Y_b)$, it is likely to better reflect the between-group nature of inequality (Gradin and Schotte, 2020).

Changes in inequality over time can therefore be decomposed into the sum of the contribution of the following terms:

$$\Delta G = \Delta G_B + \Delta G_W$$

Gradin and Schotte (2020) explain that the contribution of inequality between occupations in the trends observed in overall inequality ΔG_B may come from two different channels. Firstly, inequality can be affected by changes in employment structure over time while earnings differences between occupations remain constant. For example, inequality may increase if middle-income occupations decrease in size relative to other groups and the earnings differences across occupations remain stable. Secondly, on the other hand, while the structure of employment remains constant, inequality can increase/decrease if the earnings gap between occupations changes. For example, if high-paying occupations experience faster income growth compared to low-paying ones, while the structure of employment does not change, overall earnings inequality may increase as well. Therefore, in order to isolate whether changes in employment structure or changes in average earnings are driving the pattern of inequality between occupations, we also perform the analysis with counterfactual distributions. In one specification, we keep the occupational shares constant while in the other one, the occupational mean earnings are kept constant.

Let ΔG_{BE} be the change in inequality between occupations in a situation where occupational employment shares are kept constant in *(t-1)* and *t* time periods—thus the only component that varies across occupations is the mean earnings. Similarly, ΔG_{BM} is the change in inequality when the employment shares are allowed to change but mean earnings of occupations are held constant. In this setup, the Shapley index can be defined in the following manner:

$$\Delta G_B = \Delta G_{BE} + \Delta G_{BM}$$

where,

$\Delta G_{BE} = 0.5[\Delta G_{be} + \Delta G_b - \Delta G_{bm}]$ and
$\Delta G_{BM} = 0.5[\Delta G_{bm} + \Delta G_b - \Delta G_{be}]$

RIF decomposition

According to Firpo et al. (2009, 2011), for the cumulative distribution of wages F_{Y}, let us assume that the distributional statistic is defined as $v(F_{\mathrm{Y}})$ (called distributional parameter in Firpo et. al. 2011). Also, let $F_{\mathrm{Y0|t=0}}$ denote the cumulative distribution observed at period 0, and for period 1 it is defined as $F_{\mathrm{Y1|t=1}}$. On the other hand, the counterfactual distribution can be denoted by $F_{\mathrm{Y0|t=1}}$ (a situation where the workers in period 1 are paid under the same wage structure of period 0).

Now, the overall change in $v(F_{\mathrm{Y}})$ between these two periods can be written in the following manner:

$$\Delta_0^v = v\left(F_{\mathrm{Y1|t=1}}\right) - v(F_{\mathrm{Y0|t=0}})$$
$$= \left[v\left(F_{\mathrm{Y1|t=1}}\right) - v(F_{\mathrm{Y0|t=1}})\right] + \left[v(F_{\mathrm{Y0|t=1}}) - v\ (F_{\mathrm{Y0|t=0}})\right]$$
$$= \Delta_w^v + \Delta_c^v$$

Here, Δ_w^v= wage structure effect and Δ_c^v = composition effect

A challenging part of this estimation is to find out the counterfactual wage distribution which uses the reweighting approach as applied by DiNardo et al. (1995). In this method, a reweighting factor is used to replace the marginal distribution of covariates X for workers in period 0 with the ones of period 1. The reweighting factor can be defined in the following manner:

$$S(X) = \frac{\Pr(X|\, d_1 = 1)}{\Pr(X|\, d_1 = 0)}$$
$$= \frac{\frac{\Pr\,(d_1=1|X)}{\Pr\,(d_1=1)}}{\frac{\Pr\,(d_1=0|X)}{\Pr\,(d_1=0)}} \qquad (7)$$

The distributional statistic $v(F_{\mathrm{Y}})$ can be calculated by *S (X)* (a value of each observation calculated while using Equation 7). This method of DiNardo et al. (1995), however, can only estimate the composition and wage structure effect but cannot decompose the contribution of every single variable. In this connection, Firpo et al. (2009) have used RIF (re-centered influence function) regression where an influence function has been used to capture the way a distributional statistic changes due to a small change in the variable(s). For each value of y, the influence function $IF(Y; v; F_Y)$ gives a value for the changes occurring in y. The re-centered influence function (RIF) can be defined as:

$$RIF(Y; v; F_Y) = \ v(F_Y) + IF(Y; v; F_Y)$$

For detailed decomposition, Firpo et al. (2009) have shown that the coefficients from RIF regression can be used to perform Oaxaca Blinder (OB) decomposition on the reweighted data. In this setup, total change can be expressed in the following manner:

Total change:

$$\hat{\Delta}^{v} = \hat{\Delta}_{w}^{v} + \hat{\Delta}_{c}^{v}$$
$$= (\bar{X}_0 - \bar{X}_1)\hat{\beta}_0 + \bar{X}_1(\hat{\beta}_0 - \hat{\beta}_1) = \text{total earning structure} + \text{total composition}$$

While incorporating specification error, total composition effect can be expressed in the following manner.

Total composition, $\hat{\Delta}_{C,R}^{v} = (\bar{X}_{01} - \bar{X}_1)\hat{\beta}_0{}^{v} + \bar{X}_1\left(\hat{\beta}_{01}{}^{v} - \hat{\beta}_1{}^{v}\right) = \hat{\Delta}_{C,P}^{v} + \hat{\Delta}_{C,Se}^{v}$

Here, $\hat{\Delta}_{C,P}^{v}$ = RIF composition effect and $\hat{\Delta}_{C,Se}^{v}$ = RIF specification error

Similarly, for the wage structure effect, we get:

Total earning structure, $\hat{\Delta}_{W,R}^{v} = \bar{X}_1\left(\hat{\beta}_1{}^{v} - \hat{\beta}_{01}{}^{v}\right) + (\bar{X}_1 - \bar{X}_{01})\hat{\beta}_{01}{}^{v} = \hat{\Delta}_{W,P}^{v} + \hat{\Delta}_{W,Re}^{v}$

Here, $\hat{\Delta}_{W,P}^{v}$ = RIF earnings structure effect and $\hat{\Delta}_{W,Re}^{v}$ = RIF reweighting error

References

Chantreuil, F., and A. Trannoy (1997). 'Inequality Decomposition Values'. Mimeo. Cergy-Pontoise: Université de Cergy-Pointoise.

DiNardo, J., N.M. Fortin, and T. Lemieux (1995). 'Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach'. NBER Working Paper 5093. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w5093

Firpo, S., N.M. Fortin, and T. Lemieux (2011). 'Occupational Tasks and Changes in the Wage Structure.

Firpo, S., N.M. Fortin, and T. Lemieux (2009). 'Unconditional Quantile Regressions'. *Econometrica*, 77(3): 953–73. https://doi.org/10.3982/ECTA6822

Gradín, C., and S. Schotte (2020). 'Implications of the Changing Nature of Work for Employment and Inequality in Ghana'. WIDER Working Paper 2020/119. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2020/876-4

Goos, M., and A. Manning (2007). 'Lousy and Lovely Jobs: The Rising Polarization of Work in Britain'. *The Review of Economics and Statistics*, 89(1): 118–33. https://doi.org/10.1162/rest.89.1.118

Sebastian, R. (2018). 'Explaining Job Polarisation in Spain from a Task Perspective'. *SERIEs*, 9(2): 215–48. https://doi.org/10.1007/s13209-018-0177-1

Shorrocks, A. F. (2013). 'Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value'. *The Journal of Economic Inequality*, 11(1): 99–126. https://doi.org/10.1007/s10888-011-9214-z