Oaxaca-Decomposition Model

To observe how differences in individual, household, business, regional and housing characteristics contributes to the gender gap in the choice of HBS, the Blinder-Oaxaca (B-O) decomposition technique for nonlinear binary outcome variable was employed. The B-O decomposition method (Blinder 1973 and Oaxaca 1973) is widely used in analyzing gender and group-based wage differences. The B-O model has been applied in health related studies (Rahimi & Seyed, 2021; Fagbamigbe, Oyewale, & Folusho, 2022), in food security ((Gebre, et al., 2021; Zingwe, Manja, & Chirwa, 2021), and in business and entrepreneurship ((Hoang, Nahm,, & Dobbie, 2021; Lemma, Gwatidzo, & Mlilo, 2022; Hewa-Wellalage, Boubaker, Hunjra, & Verhoeven, 2021).

The conventional B-O model is expressed as follows:

$$\underbrace{\bar{Y}_m - \bar{Y}_f}_{Gender\ Gap} = \underbrace{\{(\bar{X}_m - \bar{X}_f)\beta_f\}}_{Endowment\ Effects\ (E)} + \underbrace{\{(\beta_m - \beta_f)\bar{X}_i\}}_{Coefficient\ Effects\ (C)} + \underbrace{\{(\bar{X}_m - \bar{X}_f)(\beta_m - \beta_f)\}}_{Interaction\ Effects\ (I)} =$$

$$E + C + I \qquad (4)$$

where the LHS captures the average gender gap (males – females outcomes) in the outcome variables, while $\bar{X}_m$ and $\bar{X}_f$ are the average observed characteristics of males and females respectively. Similarly, β_m and β_f are the coefficients for males and females. The average gender gap is this decomposed into three components, the endowment, coefficient, and interaction effects. The endowment effects (E) captures the gender gap attributed to individual characteristics while the coefficient effects (C) captures a portion of the gap emanating from differences in returns. The interaction effects (I) also capture the joint effects of observed characteristics and coefficients in explaining the portion of the gender gap.

Considering that the outcome variable is binary, as described above, we adopted the extended version of the B-O model due to Fairlie (1999) and Fairlie (2005). This extended version follows a decomposition framework for probit and logit models as expressed below:

$$\underbrace{\overline{HBS}_m - \overline{HBS}_f}_{Gender\ Gap} = \underbrace{\left\{\sum_{i=1}^{N_m} \frac{F(X_{im}\beta_f)}{N_m} - \sum_{i=1}^{N_f} \frac{F(X_{if}\beta_f)}{N_f}\right\}}_{Endowment\ Effects\ (E)} + \underbrace{\left\{\sum_{i=1}^{N_f} \frac{F(X_{if}\beta_m)}{N_f} - \sum_{i=1}^{N_f} \frac{F(X_{if}\beta_f)}{N_f}\right\}}_{Coeffieicnt\ Effects(C)} +$$

$$\underbrace{\left\{\sum_{i=1}^{N_m} \frac{F(X_{im}\beta_m)}{N_m} - \sum_{i=1}^{N_m} \frac{F(X_{im}\beta_f)}{N_m}\right\} + \underbrace{\left\{\sum_{i=1}^{N_f} \frac{F(X_{if}\beta_m)}{N_f} - \sum_{i=1}^{N_f} \frac{F(X_{if}\beta_f)}{N_f}\right\}}_{E}}_{Interaction\ Effects\ (I)} \quad (5)$$

The LHS ie $\overline{HBS}_m - \overline{HBS}_f$ captures the average gender gap in HBS choices, while $\underline{X}_m$ and $\underline{X}_f$ are the average observed characteristics of males and females, respectively. Similarly, β_m and β_f are the coefficients for males and females. N_m and N_f are the respective sample sizes for male and female business owners. Equation 2 offers a three-decomposition approach to analyzing the gender gap in home-based business operations. The average gender gap is analyzed from the perspective of female business owners. This means that the predictors of the gender gap in HBS are weighted using the female coefficients (β_f) to measure the endowment effects (E). E, therefore, measures the expected change in HBS if female owners had male observable characteristics.

We also used the female business owners' characteristics (X_{if}) as weights in measuring the coefficient effect (C). This is the expected change in the average probability of female HBS operation if females had male characteristics. For all the components of the gender gap (E, C, and I), a negative value implies a higher probability of HBS outcomes for females. In contrast, positive values indicate a male advantage in HBS operation.

APPENDIX:

Table A: Goodness of fit test for the post selection coefficients

Sample	Deviance	Deviance ratio	No of Observations	No of covariates	No of nonzero coefficients
Training	1.123	0.178	9,079	65	38
Validation	1.126	0.172	2,317	65	38
Overall	1.122	0.178	11,396	65	38

Table B: B-O decomposition from POSTLASSO for married and non-married business owners (all years).

	Married with no child with HH		Married with at least a child in HH		Non-married	
	Log points	% of gap	Log points	% of gap	Log points	% of gap
Secondary	-0.088	42.93	-0.000	0.00	-0.014	6.76
Tertiary	0.156	-76.10	0.002	-0.66	0.001	-0.48
Age	-0.584	284.88	-0.062*	20.39	0.051	-24.64
Age2	0.548	-267.32	0.065*	-21.38	-0.079	38.16
Hours of work	0.060	-29.27	0.001	-0.33	0.004	-1.93
Institutions – All	-0.149	72.68	0.004***	-1.32	-0.001	0.48
Partner characteristics	-0.093	45.37	-0.006*	1.97		0.00
Others	-0.915	446.34	0.017*	-5.59	0.137***	-66.18
Polygamous marriage	0.016	-7.80	-0.002*	0.66		0.00
Endowments	-0.020	9.76	0.020*	-6.58	0.099***	-47.83
Coefficients	-0.181***	88.29	-0.329***	108.22	-0.140***	67.63
Interaction	-0.004	1.95	0.004	-1.32	-0.165***	79.71
Gender Gap	-0.205***	100.00	-0.304***	100.00	-0.207***	100.00
Observations	583	583	8,143	8,143	2,504	

*** p<0.01, ** p<0.05, * p<0.1