Summer 8-2024

The Global Gender Distortions Index (GGDI): An Application to Indian States

Pinelopi Goldberg
Yale University, penny.goldberg@yale.edu

Somik Lall
World Bank Group

Meet Mehta
Yale Univeristy, meet.mehta@yale.edu

Michael Peters
Yale University, michael.peters@yale.edu

Aishwarya Ratan
Yale University, aishwarya.ratan@yale.edu

The Global Gender Distortions Index (GGDI): An Application to Indian States

Penny Goldberg, Somik Lall, Meet Mehta,
Michael Peters, and Aishwarya Ratan

August 2024

Abstract

The extent to which women participate in the labor market and have access to formal employment differs greatly across Indian states. In this paper we build on the methodology developed by Hsieh, Hurst, Jones, and Klenow (2019) to estimate the productivity consequences of such differences. Using rich microdata on occupational sorting and earnings, our theory allows to separately identify labor demand distortions (e.g., discrimination in hiring for formal jobs) from labor supply distortions (e.g., frictions that discourage women's labor force participation). We find that both demand distortions and supply distortions are negatively related to state-level economic development. Equalizing distortions across Indian states could raise state-level productivity by up to 15%.

1. THEORY

We consider a simplified version of Hsieh et al. (2019) where we have two groups, men and women, and three occupational choices: employment in formal work (f), self employment (informal) (i) and home production (h). Men and women differ in three dimensions. First, women might face a *demand distortion*, which we model as an exogenous tax wedge on their labor earnings. Second women might face a *supply distortion*, whereby choosing a particular occupation reduces their utility. Third, men and women can differ in their occupation-specific human capital. Crucially, both demand and supply distortions vary across Indian states.

1.1 Labor Supply: Distortions vs Skills

Formally, we model the utility of an individual of group $g = m, w$, i.e. man or woman, in state s who chooses occupation o as

$$\log U_{og}(s) = \log C_{og}(s) + \log z_{og}(s) \tag{1}$$

where $C_{og}(s)$ denotes consumption and $z_{og}(s)$ denotes the utility of working in occupation o. Consumption is linked to individual's human capital, the prevailing wage rate (per efficiency unit), and the prevailing demand distortion by the budget constraint

$$C_{og}(s) = (1 - \tau_{og}(s))\, w_o(s)\, \bar{h}_{og}(s)\, \epsilon. \tag{2}$$

Here $w_o(s)$ is the prevailing wage rate in occupation o in state s, $\bar{h}_{og}(s)$ denotes the occupation specific human capital of an individual of group g in state s, $\tau_{og}(s)$ parametrizes the *demand distortion* and ϵ is an idiosyncratic productivity draw that allows individuals to differ in their comparative and absolute advantage in different occupations.

Substituting (2) into (1), an individual's indirect utility is given by

$$U^*_{og}(s) = \tilde{w}_{og}(s)\, \epsilon_o$$

where

$$\tilde{w}_{og}(s) \equiv w_o(s)(1 - \tau_{og}(s))\bar{h}_{og}(s) z_{og}(s). \tag{3}$$

Hence, $\tilde{w}_{og}(s)$ summarizes the systematic attractiveness of an occupation o in state s for group g. It depends on occupational skill prices $w_o(s)$, labor demand distortions $\tau_{og}(s)$, the average human capital endowment $\bar{h}_{og}(s)$, and labor supply distortions $z_{og}(s)$.

Individuals choose occupation o to maximize $U^*_{og}(s)$. For tractablity we assume that the ϵ_o are drawn from an independent Frechet distribution:

$$F(\epsilon_o) = e^{-\epsilon_o^{-\theta}}.$$

Standard arguments then allow us derive our first main result:

Proposition 1. *Let $p_{og}(s)$ denote the fraction of people from state s and group g who choose occupation o and $w\bar{a}ge_{og}(s)$ denote the geometric average of earnings in occupation o by state s of group g. Then*

$$p_{og}(s) = \frac{\tilde{w}_{og}^{\theta}}{\sum_j \tilde{w}_{jg}^{\theta}} \tag{4}$$

$$w\bar{a}ge_{og}(s) = \tilde{\Gamma}\left(\sum_j \tilde{w}_{jg}^{\theta}\right)^{\frac{1}{\theta}} z_{og}(s)^{-1}. \tag{5}$$

where $\tilde{w}_{og}$ is given in (3).

Equations (4) and (5) are at the heart of our identification strategy. Equation (4) highlights that the occupation-specific employment shares reflect an occupation's *relative attractiveness* $\tilde{w}_{og}$. This attractiveness in turn is fully determined by market prices and human capital endowments on the one hand and distortions (either from the supply or the demand side) on the other hand. Intuitively speaking: If a particular group g has a high employment share in occupation o it could be that their human capital h_{og} in this occupation is large, that skill prices w_{og} are high, that distortions τ_{og} are low, or that the utility of working in this occupation, z_{og}, is high.

Equation (5) shows that average occupation earnings provide useful different information. First of all, for a given group g, the *only* variation of the average earnings across occupations is due to differences in supply distortions $z_{og}(s)$: The lower z_{og}, the higher the average wage, because wages play the role of a compensating differential. By contrast, if z_{og} was equalized across occupations, average wages would also be equalized, i.e. neither differences in human capital, not differences in skill prices or distortions, affect average earnings across occupations. This result, which is due to the selection of individuals across occupations, is a particular property of our assumptions of ϵ_o to be Frechet distributed.

Using (5) to substitute $w\bar{a}ge_{og}(s)$ for the labor supply distortions $z_{og}(s)$ in $\tilde{w}_{og}$ (see (3)) allows us express the share of women in occupation o relative to men as

$$\frac{p_{ow}(s)}{p_{om}(s)} = (1-\tau^{w}_{ow}(s))^{\theta} \times \left[\frac{\bar{h}_{ow}(s)}{\bar{h}_{om}(s)}\right]^{\theta} \times \left[\frac{\bar{\text{wage}}_{ow}(s)}{\bar{\text{wage}}_{om}(s)}\right]^{-\theta}. \tag{6}$$

Equation (6) highlights that relative occupational shares are driven by three considerations: women can be underrepresented in a particular occupation if (i) they face a labor demand distortion τ_{ow}, (ii) they have a lower human capital endowment in this occupations ($\frac{\bar{h}_{ow}(s)}{\bar{h}_{om}(s)}$), and (iii) their average wage is relatively high. This last effect of the wage operates through the labor supply distortion. Recall that equation (5) highlights that average occupation wages *only* reflect the labor supply distortion $z_{og}(s)$. A high wage in a particular occupation thus reflects a compensating differential for an existing labor supply distortion.

Under our assumptions on labor supply, we can also derive a closed form expression for the aggregate supply of efficiency units in occupation o in state s, $H_o(s)$. In particular, $H_o(s)$ is given by

$$H_o(s) = \sum_g L_s q_g(s) p_{og}(s)\frac{\theta-1}{\theta}\bar{h}_{og}(s)\Gamma\left(1-\frac{1}{\theta}\right), \tag{7}$$

where L_s denotes the aggregate population in state s, $q_g(s) \in [0,1]$ denotes the share of the population in s that is in group g, and $\Gamma\left(1-\frac{1}{\theta}\right)$ is the gamma-function. The term $p_{og}(s)\frac{\theta-1}{\theta}$ accounts for selection: while overall human capital supply is increasing in $p_{og}(s)$, the elasticity is $\frac{\theta-1}{\theta} < 1$, reflecting the fact that average efficiency declines as more people sort into a particular occupation.

1.2 Labor Demand: Technology

To close the model in general equilibrium, we assume that each state s is populated by a representative firm that produces final output according to

$$Y(s) = \left[\sum_o (A_o(s)H_o(s))^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}. \tag{8}$$

Hence, overall output is a CES aggregator of the output produced in different occupations and occupations differ by their total factor productivity, $A_o(s)$, which can also differ across states. Aggregate GDP in India is then simply given by $Y^{IND} = \sum_s Y(s)$.

1.3 Equilibrium

Because we assume that goods are non-tradable across states and the factor markets clear locally, we can characterize the equilibrium separately for each state s. A competitive equilibrium consists of sequence of occupational choices, total efficiency units of labor in each group H_{og}, final output Y, and an efficiency wage w_o in each occupation such that

1. Each individual's occupational choice maximises utility taking as given skill prices w_o, occupation specific human capital $\bar{h}_{og}$, demand distortions τ_{og}, and supply distortions z_{og},
2. The set of skill prices w_o clear each occupational labor market
3. Total output is given by the production function in equation (8).

Given the demand distortions, the equilibrium is thus described by the occupation specific labor market clearing conditions

$$H_{om}^{supply}(s) + H_{ow}^{supply}(s) = H_o^{demand}(s),$$

where $H_{og}^{supply}(s)$ is given in (7) and labor demand is consistent with firms' profit maximization

$$H_{og}^{demand}(s) = \left(\frac{A_o(s)^{\frac{\sigma-1}{\sigma}}}{w_o(s)} \right)^{\sigma} Y(s) \tag{9}$$

Together with the production function (8), these equations fully characterize the set of equilibrium wages $w_o(s)$.

2. The Gender Distortion Index

Given this modeling environment we can now define the *Gender Distortion Index* formally. From above we can compute overall output as a function of supply and demand distortions in each state s:

$$Y(s)^D \equiv Y(s; [\tau_o, z_o]_o),$$

where $Y(s; [\tau_o, z_o]_o)$ denotes equilibrium output in the presence of distortions, hence the superscript D. By contrast, output in the absence of distortions is given by

$$Y(s)^E \equiv Y(s; [\tau_o = 1, z_o = 1]_o),$$

where we use the mnemonic EQ to indicate that this is equilibrium output under *gender equality*. The *Gender Distortion Index* is then given by

$$GDI \equiv \ln\Big(Y(s)^D / Y(s)^{EQ} \Big). \tag{10}$$

In words, our index is given by the percentage loss in output due to gender-based distortions: the higher GDI, the higher the extent of gender-based talent misallocation.

3. IDENTIFICATION

To compute the GDI in (10), we need to (i) measure the prevailing demand and supply distortions, (ii) compute $Y(s)^D$, and then (iii) compute the counterfactual output $Y(s)^{EQ}$ if these distortions were absent.

3.1 Measuring Market Distortions

To measure the extent of misallocation we assume that we have empirical measures of occupational employment shares $p_{og}(s)$, average wages $\bar{\text{wage}}_{og}(s)$, and average human capital terms $\bar{h}_{og}(s)$. Furthermore, we assume that the parameters θ is known. In Section (4) we describe in detail how we measure these objects empirically.

We use the assumption on θ to calibrate the parameters $\tau_{ow}(s)$ and $z_{og}(s)$.

Demand Distortions $\tau_{og}(s)$ The occupation-specific demand distortions $\tau_{ow}(s)$ can be directly recovered from equation (6):

$$\frac{1}{1-\tau^{w}_{ow}(s)} = \left(\frac{p_{ow}(s)}{p_{om}(s)}\right)^{-1/\theta} \times \left(\frac{\bar{h}_{ow}(s)}{\bar{h}_{om}(s)}\right) \times \left(\frac{\bar{\text{wage}}_{ow}(s)}{\bar{\text{wage}}_{om}(s)}\right)^{-1}. \tag{11}$$

Supply Distortions $z_{og}(s)$ For $z_{og}(s)$, we use the relative wages from equation 5. Notice that we have the normalisation $z_{hg} = 1$ for both the groups. This gives us the following equation:

$$\frac{w\bar{a}ge_{og}(s)}{w\bar{a}ge_{hg}(s)} = z_{og}^{-1}(s) \tag{12}$$

This normalisation can also be used to estimate $m_g(s)$. Rearranging equation 5 for the home sector, we get:

$$\hat{m}_g(s) = \bar{\text{wage}}_{hg}(s)^{\theta}\tilde{\Gamma}^{-\theta}$$

3.2 Other Parameters

Scale parameter θ for distribution of ϵ Hsieh et al. (2019) estimates $\theta = 1.52$ after adjusting for the elasticity of human capital w.r.t. human capital expenditure. We assume $\theta = 1.5$ as of now, which can be further updated by fitting the Frechet distribution on the wage data.

Imputing home sector earnings Note that equation 12 requires us to know the earnings for the home sector for both men and women. For men, we assume that whenever they decide not to work for formal or informal sector, they can always go back and work in the agriculture sector. Therefore, we set the earnings in the home sector for men to be equal to average earnings of men in the agriculture sector in that state.

For earnings of women, we the exploit the assumption that $\tau_{hw} = 0$. We use equation 11 for the home sector h. Given that earnings of men in the home sector is known using the assumption above, the only remaining unobservation in this is equation is earnings for women in the home sector.

$A_o(s), w_o(s), Y(s)$ To estimate $A_o(s)$, we use the labor demand equation 9. To use that equation, we additionally require the estimates of $H_{og}^{demand}(s)$, $Y(s)$ and $w_o(s)$. For $w_o(s)$, we use equation 4, which can be re-written as follows for the male group (given $\tau_{om}^{w} = 0$):

$$w_o(s) = \frac{\left(p_{om}(s)\hat{m}_m(s)\right)^{\frac{1}{\theta}}}{\tilde{z}_{om}(s)\bar{h}_{om}(s)}$$

where has used the estimate of $\hat{m}_m(s)$ derived while estimating $z_{og}(s)$.

Given the estimate of $w_o(s)$, we can find labor supply:

$$H_o^{supply}(s) = \sum_g q_g(s)p_{og}(s)\mathbf{E}[h_{og}(s)\epsilon_{og}(s)|\text{choose o}]$$

$$H_o^{supply}(s) = \sum_g q_g(s)p_{og}(s)^{\frac{\theta-1}{\theta}}\bar{h}_{og}(s)\Gamma(1-\frac{1}{\theta})$$

We use the equilibrium relationship $H_o^{supply}(s) = H_o^{demand}(s) = H_o(s)$ to plug back into the labor demand equation and finally, we can estimate Y as a sum of total wage payments and taxes:

$$Y(s) = \sum_o w_o(s)H_o(s)$$

Lastly, following the paper, we pick $\sigma = 3$ arbitrarily.

Arithmetic and Geometric Mean Lastly, we observe arithmetic average of earnings in the data where equation 11 and 12 requires geometric average of wages. We convert arithmetic average wage to geometric average wages using the following formula.

$$\bar{\text{wage}}_{og} = \text{Aritmetic Average Wage}_{og} \frac{\tilde{\Gamma}}{\Gamma(1-\frac{1}{\theta})}$$

where $\tilde{\Gamma} = e^{\frac{\gamma_{em}}{\theta}}$.

3.3 Solving for Equilibrium

Given the parameters, we can now move to solve the model. The way to solve for equilibrium here is to guess the values of $Y(s)$, $m_m(s)$ and $m_w(s)$. These 3 guess, along with the model parameters help us in calculating $H_o^{demand}(s)$ and $H_o^{supply}(s)$. We find $w_o(s)$ such that the two are equal and re-estimate the values of $Y(s)$, $m_m(s)$ and $m_w(s)$. We iterate this process until the guess and the estimated values of $Y(s)$, $m_m(s)$ and $m_w(s)$ are equal.

4. Quantitative Application: The *GDI* Across Indian States

4.1 Data and Measurement

We apply this model to India using the Periodic Labour Force Survey (PLFS) of 2018-19. It contains detailed data on labor force participation of a representative sample of (approx.) 100,000 households. We restrict the sample to working age population: age 25-60 years. This leaves us with 202,696 individual level observations. Table 1 reports the descriptive statistics.

PLFS classifies the employment status of an individual into following categories: Regular salaried employee (code 31, 71, 72), Self employed (code 11, 12, 21, 61, 62), Casual wage labour (code 41, 51), Unemployed (code 81) and Out of labour force (code 91-97). We classify regular salaried employee into the formal sector (f), self employed and casual wage labour into informal sector (i), and unemployed and out of labour force in the home sector (h). As shown in table 1, roughly $\frac{4}{5}^{th}$ of women are not in labor force, where as this number is less than 10% for men. Majority of people are self-employed, followed by regular wage and casual labor.

We use average monthly earnings at the state-group level to estimate $\bar{\text{wage}}_{og}$. PLFS reports monthly earnings in the month preceding the survey for regular wage employees and for self employed. For people working as a casual labor, they report the daily wages for each of the day in the preceding week. We multiply their total weekly earnings by 4 to calculate monthly earnings. The earnings are reported for almost everyone employed as either regular wage employee or as casual labor. For individuals in the self employed category, earnings is not reported for those who work as unpaid

TABLE 1: Descriptive Statistics

	Men	Women	Total
Count	100501	102195	202696
Employment Proportions			
Regular Wage	23.18%	6.67%	14.83%
Self Employed	46.54%	13.48%	29.82%
Casual Labor	20.37%	6.27%	13.24%
Not in LF	9.92%	73.58%	42.12%
Total	100.00%	100.00%	100.00%
Average Monthly Earnings (in rupees)			
Regular Wage	18540	13836	17468
Self Employed	11640	5223	10772
Casual Labor	6684	3898	6015
Not in LF	8515	1656	2459
Total	11943	3090	7635
Agriculture			
Employment Proportion	32.2%	13.4%	22.7%
Missing Observations for Earnings			
Regular Wage	0.1%	0.1%	0.1%
Self Employed	9.6%	50.0%	19.9%
Casual Labor	0.0%	0.0%	0.0%
Not in LF	100.0%	100.0%	100.0%
Total	35.0%	86.9%	61.3%

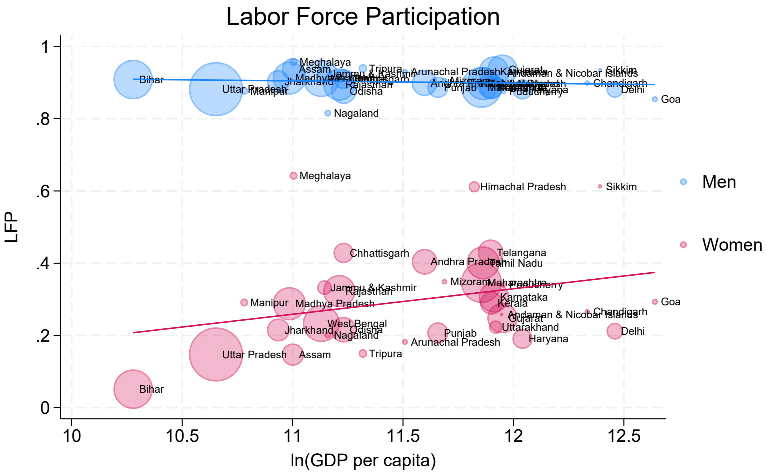
Notes: Descriptive statistics are calculated from Periodic Labor Force Survey of 2018-19. The values of monthly earnings are winsorized at 5^{th} and 95^{th} percentile for every state-sector cell. Monthly earnings of agriculture sector is used to calculate the average monthly earnings of men in the Not in LF (home) sector. Monthly average earnings of women in that sector is then calculated using the equation 11. Survey weights are used for all the calculations.

workers in the family enterprise. They constitute 50% of self employed women and 20% of self employed men. We assume that the earnings for them is equal to the average earnings of paid self employed individuals in their respective state-group cell.

Lastly, for the individuals who are not in the labor force, details for imputation of home sector earnings is provided in section 3.2. In section B in the Appendix we describe in detail how we clean the data.

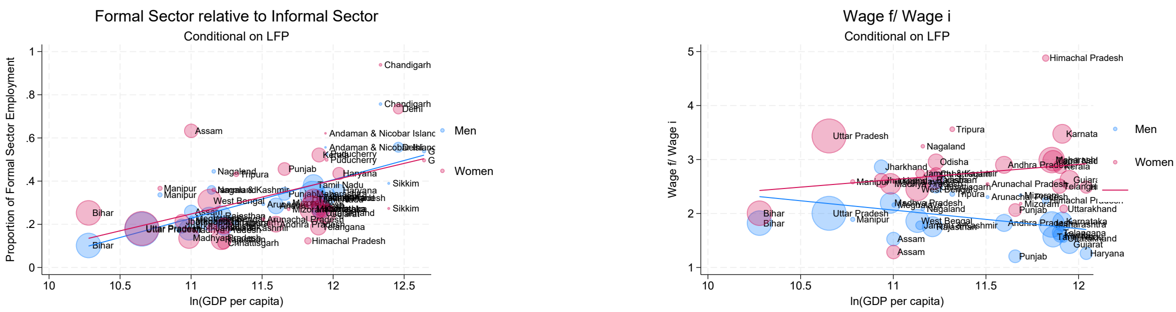
Figure 1 plots the labor force participation for men and women across states against the log of per capita GDP. It shows a robust positive correlation between female labor force participation and state per capita GDP. For men, the labor force participation is high and not related to state per capita GDP. Figure 2a plots the proportion of employment in formal sector relative to informal sector for men and women across states. We see that for both men and women, formal sector employment increases as state per capita

FIGURE 1: Labor force participation across states



Notes: The figure plots labor force participation rate of men and women in each state again log of GDP per capita of the state.

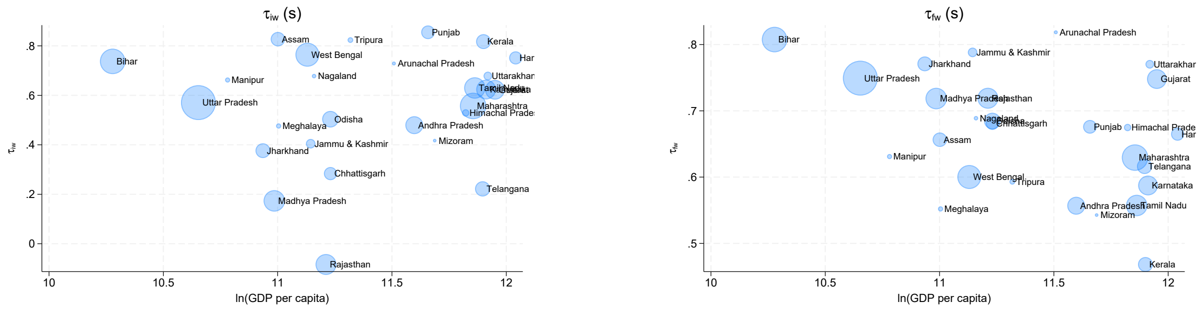
FIGURE 2: Sectoral Employment and Wage Premium



(A) (B)

Notes: Panel A plots the ratio of formal sector and informal sector employment of every state against the log of GDP per capita of the state, for men and women separately. Panel B plots the ratio of average earnings in formal sector to that in informal sector for every state against the log of GDP per capita of the state, for men and women separately.

FIGURE 3: Labor Demand Distortions



(A) (B)

Notes: Panel A plots labor demand distortions for women in the informal sector $\tau_{iw}(s)$ for every state against the log of state's GDP per capital. Panel B plots labor demand distortions for women in the formal sector $\tau_{fw}(s)$ for every state against the log of state's GDP per capital. They are calculated using equation 11.

income increases. Lastly, we look at the earnings in formal sector relative to informal sector for men and women in figure 2b. We notice that earnings premium for the formal sector is high for women (compared to men) for most of the states, however, this premium is not systematically related to state per capital GDP for both men and women. In figures A.1a, A.1b and A.1c we plot the average years of schooling of women relative to men ($\frac{h_{ow}}{h_{om}}$) for the home sector, informal sector and formal sector respectively. This ratio is increasing with the state GDP per capita on average for the home sector and the informal sector, where as it is unrelated to state GDP per capita for the formal sector.

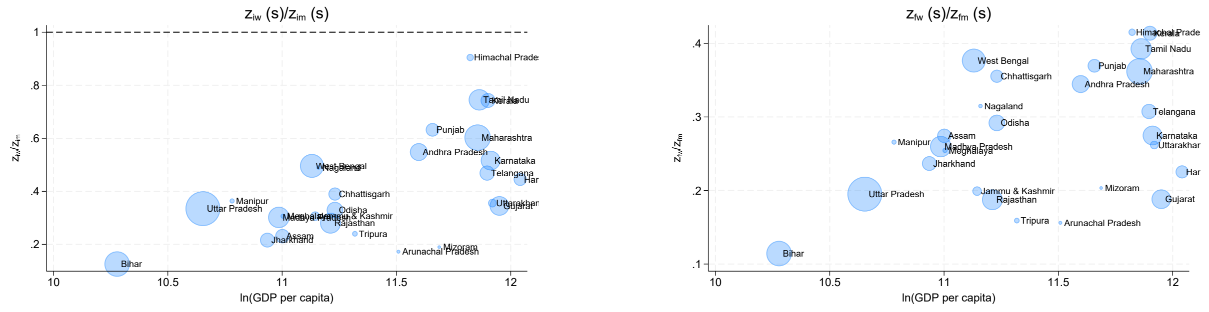
4.2 Estimated Distortions

From PLFS, we get the data on $p_{og}(s)$ and $\bar{\text{wage}}_{og}(s)$. We plug these into equations 11 and 12 to estimate the distribution of labor taxes $\tau_{ow}(s)$ and supply distortions z_{ow}. We assume $\theta = 1.5$[1].

Figure 3a plots the labor demand distortions for the informal (i) and figure 3b plots the labor demand distortions for the formal (f) sector against state GDP per capita (in logs). The higher the value of $\tau_{ow}(s)$, the higher the level of distortions. The plots highlight 3 points: 1. The level of labor demand distortions decline on average with increase in state GDP per capital, especially for the formal sector 2. For most of the states, the level of demand distortions in the formal sector is higher compared to the informal sector and 3. The variation in the level of demand distortions increase with the increase in state GDP per capita, particularly for the formal sector - increase in per capita income doesn't necessarily bring down the distortions for all the states.

[1]Hsieh et al. (2019) estimates $\theta = 1.52$ after adjusting for the elasticity of human capital w.r.t. human capital expenditure. We assume $\theta = 1.5$ as of now, which can be further updated by fitting the Frechet distribution on the wage data.

FIGURE 4: Labor Supply Distortions



(A) (B)

Notes: $\frac{z_{ow}}{z_{om}}$ denote supply distortion for women in sector o. $\frac{z_{ow}}{z_{om}} < 1$, it implies that women derive lesser utility than men by working in sector o. Panel A plots this ratio for informal sector and panel B plots this ratio for the formal sector. They are calculated using equation 12.

Figure 4a and 4b plot the ratio $\frac{z_{ow}(s)}{z_{om}(s)}$ for the informal (i) and formal (f) sector respectively. Notice that for both the sectors this ratio is below 1 for most of the states. This implies that the estimated supply distortions is relatively more severe for women compared to men in all the states. The lower the ratio, the higher is the level of labor supply distortions. Again, the three observations highlighted for the demand distortions also hold for the labor supply distortions.

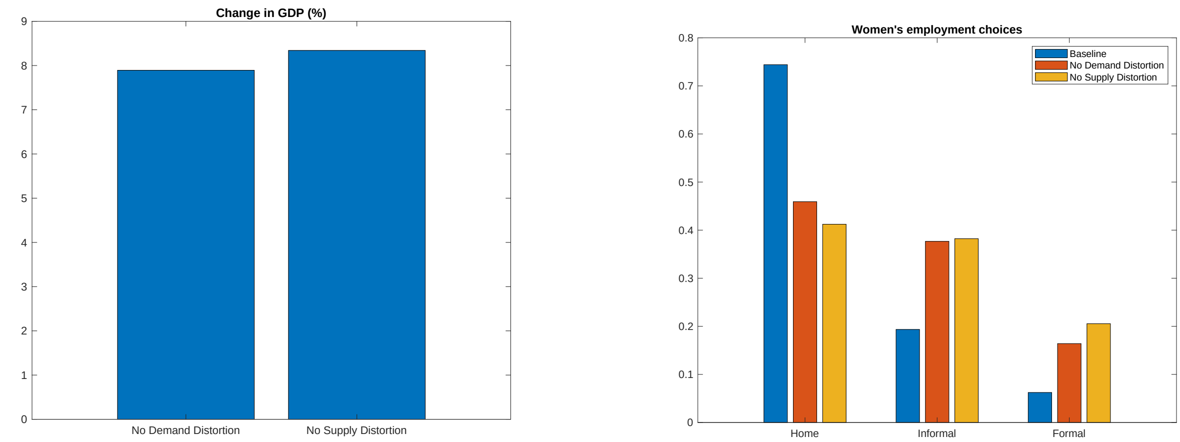
4.3 Implications for Productivity

Distortions on either supply or demand reduce productivity through a misallocation of talent - more productive women end up working in less productive sector. We now quantify the economic costs of our estimates from Section 4.2 for productivity both at the aggregate (i.e. India-wide) level and across states.

We study two counterfactual scenario: 1. Women face no labor demand distortions relative to men ($\tau_{iw} = \tau_{fw} = 0$) 2. Women face no labor supply distortion relative to men ($z_{ow} = z_{om}$ for $o \in \{h, i, f\}$).

In Figure 5a we focus on the aggregate level. Specifically, we report the changes in aggregate GDP (figure 5a) and women's employment choices (figure 5b) in the two counterfactual cases. The productivity consequence of both the distortions is similar in magnitude. Removal in labor demand distortions lead to 7.90 % increase in GDP whereas removal of labor supply distortions lead to 8.34 % increase in GDP per capita. These positive effects on aggregate productivity are due to changes in occupational sorting. Removal of either distortions lead to significant reallocation of women from the home sector to the informal and formal sector. Female labor force participation rate increases more than 28 p.p. in case of no demand distortions and more than 33 p.p. in case of no supply distortions. Informal sector employment of women increase

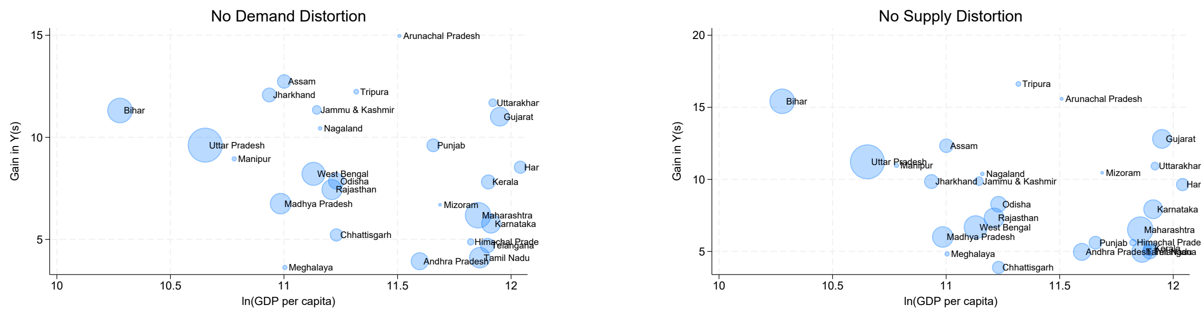
FIGURE 5: Counterfactual GDP and Employment Choices of Women



(A) (B)

Notes: Panel A shows the % increase in aggregate GDP and panel B shows women's employment choices in the two counterfactual cases. Under no demand distortion, we set $\tau_{iw} = \tau_{fw} = 0$. Under no supply distortion, we set $z_{ow} = z_{om}$ for $o \in \{h, i, f\}$.

FIGURE 6: Counterfactual GDP across states



(A) (B)

Notes: The figure plots % increase in state GDP in the two counterfactual cases again the log of GDP per capita of states. In panel A, we set $\tau_{iw} = \tau_{fw} = 0$. In panel B, we set $z_{ow} = z_{om}$ for $o \in \{h, i, f\}$.

by more than 18 p.p. by removal of either distortions. For the formal sector, removal of labor supply distortion leads to higher increase (14.3 p.p.) compared to removal of labor demand distortion (10.2 p.p.).

The analysis in Section 4.2 suggests that distortions are much more prevalent in some states relative to others. In Figures 6a and 6b, we therefore estimate such counterfactual gains for each Indian state. We find that there is substantial heterogeneity across states. Relatively poor states such as Bihar and Uttar Pradesh could increase their GDP by up to 11.3 % and 9.6 % respectively if women would not face labor demand distortions which increases their marginal product above the ones of men. Similarly, they could increase their GDP by 15.4 % and 11.2 % respectively if women faced no labor supply distortions. By contrast, such growth protential is much lower in rich states such as

Kerela and Tamil Nadu, where differences in implicit ”labor taxes” and occupational preferences are less pronounced. We also note that not all rich states have less to gain by removal of labor market distortions for women. States like Gujarat and Haryana stand to gain equally high from the removal labor demand and supply distortions as like Bihar and Uttar Pradesh. The variation in gains from removal of distortions increase with increase in state GDP per capita. This highlights that it is not guaranteed that the labor market distortions for women will go down for all states as economic development takes place.

5. ROBUSTNESS

In the section, we explore the senstivity of our estimates to the values of σ and θ. σ is the elasticity of substitution between the outputs of the three sectors in the aggregate production function. We chose $\sigma = 3$ for our model arbitrarily. In figure A.2, we plot the gain in aggregate GDP from removal of labor demand and labor supply by varying the values of σ from 2 to 10. We note that our results are robust to the values of σ. The estimates remain within 1% range of the baseline estimates.

θ is the scale parameter of Frechet distribution of idiosyncratic occupational productivity. Our current choice of $\theta = 1.5$ is close to the estimate of 1.52 of Hsieh et al. (2019). In figure A.3, we plot the aggregate gains for by varying θ from 1.1 to 3.3. We note that the magnitude of aggregate gains is similar to the baseline results for $\theta > 1.5$. However, for smaller values of θ, the aggregate results is sensitive to the assumption of θ. θ is an important parameter of our model. It is not used for solving the equilibrium, but also used for estimation of our distortion parameters and for imputing the home sector earnings of women. As a part of future exercise, we aim to estimate the value of θ by fitting the Frechet distribution on the distribution of residuals of earnings obtained after regressing them on state-sector-group-education fixed effects.

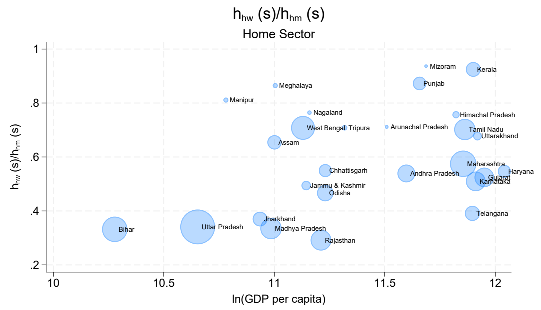
References

HSIEH, C.-T., E. HURST, C. I. JONES, AND P. J. KLENOW (2019): "The allocation of talent and us economic growth," *Econometrica*, 87, 1439–1474.

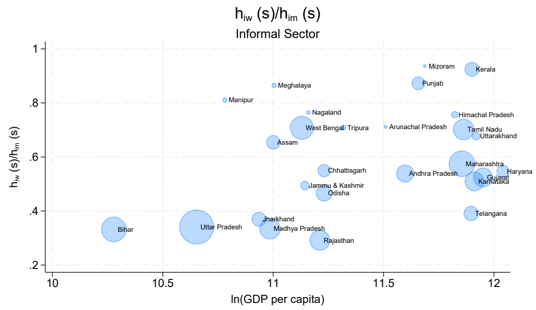
APPENDIX

A. ADDITIONAL EMPIRICAL RESULTS

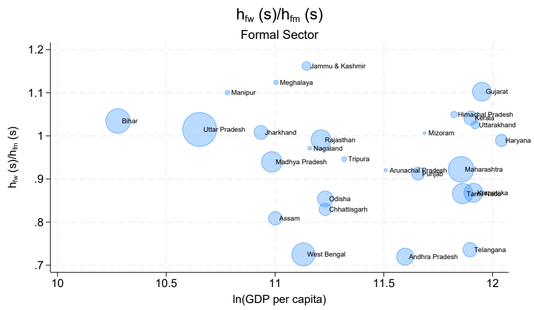
FIGURE A.1: Relative human capital across states



(A)



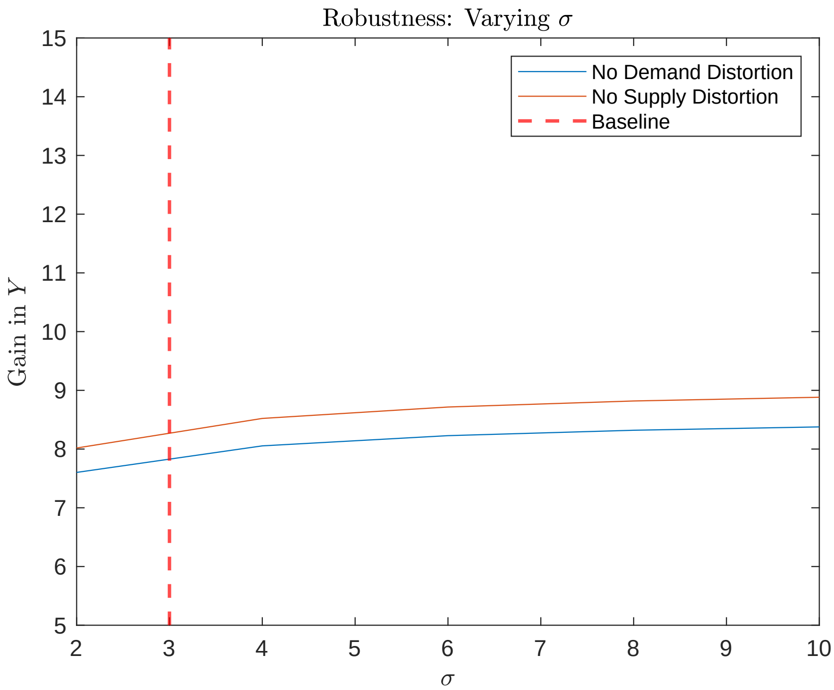
(B)



(C)

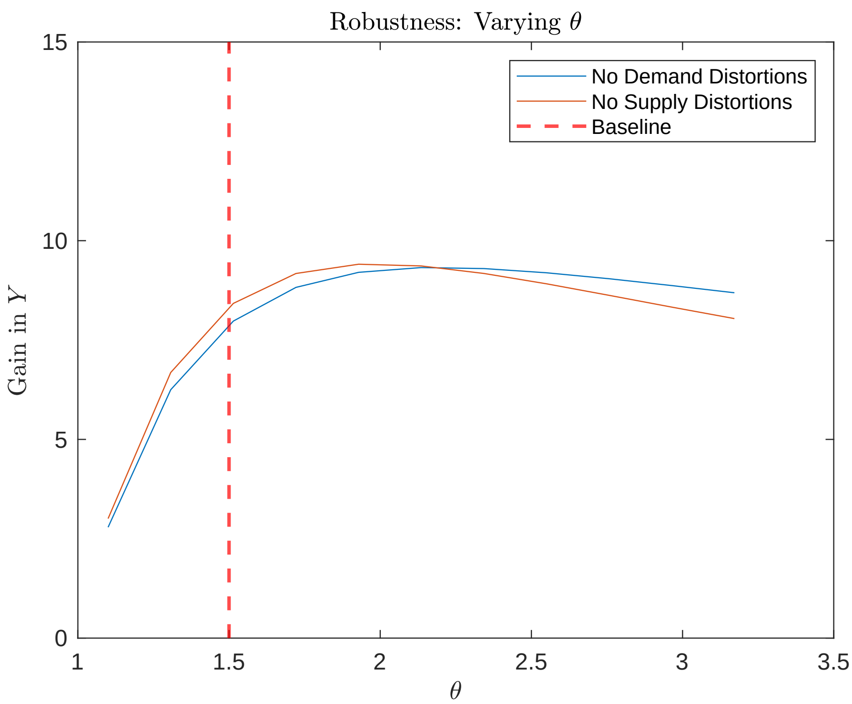
Notes: The figure plots the average years of school of women relative to men for each state for relevant sectors, against the log of state GDP per capita. Panel A plots for the home sector, panel B for the informal sector and panel C for the formal sector.

FIGURE A.2: Sensitivity to value of σ



Notes: The figure plots % increase in GDP when there are no labor demand or labor supply distortion, for different values of σ. The red bar denotes the value of estimates for our baseline choice $\sigma = 3$.

FIGURE A.3: Sensitivity to value of θ



Notes: The figure plots % increase in GDP when there are no labor demand or labor supply distortion, for different values of θ. The red bar denotes the value of estimates for our baseline choice $\theta = 1.5$.

B. Data Cleaning

We apply standard data cleaning procedures. We drop states where any of the state-sector-group cells have less than 50 observations. We winsorize the wage data in every state-sector cell at the 5^{th} and 95^{th} percentile. Lastly, for individuals who work as unpaid labour in the informal sector, we assign their wage equal to the average wage of other informal sector employee in their relevant state-group cell. We use survey weights to calculate state level aggregates.