# Aggregate Implications Of Barriers To Female Entrepreneurship*

Gaurav Chiplunkar
University of Virginia

Pinelopi Koujianou Goldberg
Yale University, BREAD, CEPR and NBER

February 2021

**Abstract**

We develop a framework for identifying and quantifying barriers to entry and operation faced by female entrepreneurs in developing countries, and apply it to the Indian economy. We find that despite considerable progress over time, female entrepreneurs still face substantial entry and business registration costs (almost twice their male counterparts'). The costs of expanding a business, conditional on entry, are also substantially higher for women. However, there is one area in which female entrepreneurs have an advantage: hiring female workers is easier for them. We show that this pattern is not driven by the sectoral composition of female employment. Counterfactual simulations indicate that removing all excess barriers faced by women entrepreneurs would: (a) increase the fraction of female-owned firms significantly (nine times); (b) increase the real wages of female relative to male workers; and (c) generate substantial aggregate productivity and welfare gains (ca. 7% and 18% respectively). These large gains are due to reallocation: low productivity male-owned firms previously sheltered from female competition are replaced by higher productivity female-owned firms previously excluded from the economy.

**JEL Classification:** J16, J70, O17, O40
**Keywords:** Female entrepreneurship, gender discrimination, misallocation, economic development

*We would like to thank Rafael Dix-Carneiro, Rohini Pande, and Gabriel Ulyssea for many helpful comments. Chiplunkar would like to acknowledge research support from the Insititute for Business in Society, Darden School of Business. Chiplunkar: Darden School of Business, University of Virginia, 100, Darden Blvd., Charlottesville, VA 22905. Email: ChiplunkarG@darden.virginia.edu. Goldberg *(corresponding author)*: Dept. of Economics, Yale University, 37, Hillhouse Ave., New Haven, CT 06511. Email: penny.goldberg@yale.edu.

# 1 Introduction

Low female labor force participation coupled with a sustained lack of female entrepreneurs have been a policy concern in many developing countries, especially in South Asia. Figure 1(a) plots the fraction of female-owned firms across 25 sectors using a sample of around 140k firms, surveyed under the Enterprise Surveys (World Bank, 2020), which covers 141 countries across 13 years (2006-2018).[1] The lack of business ownership by women is striking. On average, less than a quarter (22.5% to be exact) of businesses across the world are owned by women, with women's share of ownership ranging from 3-6% in petroleum, leather and wood products to at most 35% in textiles, services and garments. Using the same sample, Figure 1(b) plots the fraction of female workers in male-owned versus female-owned firms, as well as the probability that the top manager in the firm is a woman. While 25% of employees in male-owned firms are women, the share of female employees is 43% in female-owned firms. More strikingly, while only 6.2% of male-owned firms have a woman as their top manager, the probability of a top manager being a woman is over 50% in women-owned firms. These patterns suggest that female entrepreneurship may have important implications for women's employment patterns.

Taking the above observations as a starting point, this paper develops a framework for examining potentially differential barriers to entry and operation faced by female-owned as opposed to male-owned firms in developing countries, as well as their aggregate implications. Earlier work has shown that eliminating distortions in the allocation of talent can result in sizeable productivity and welfare gains in advanced economies[2]. Such gains could be even more important in settings characterized by misallocation of resources, low productivity, and low per capita income levels, as in many developing economies (Hsieh and Klenow, 2009; Restuccia and Rogerson, 2017). While there are many sources of identity-based distortions, gender-based distortions are a common theme in developing countries[3]. With around half of the world's population women, such distortions are likely to have important aggregate implications. If it were possible to improve aggregate productivity

[1] The Enterprise Surveys are firm-level surveys of a representative sample of the economy's private sector. More details on the methodology and data can be found in: https://www.enterprisesurveys.org.

[2] Hsieh, Hurst, Jones and Klenow (2019) estimate large such gains for the U.S. between 1960 and 2010. Their study focuses on race- and gender-based distortions.

[3] See Jayachandran (2020), Quinn and Woodruff (2019), and Cuberes and Teignier (2014) for reviews.

and welfare in developing countries by allocating the talent available in such economies efficiently, irrespective of gender, then policies promoting gender equality would be more than human rights initiatives, they would be effective development policies.

In the vein of this proposition, this paper aims to identify and analyze a particular type of distortion, namely gender-based distortions that affect female entrepreneurship. The focus of our analysis is India, a country in which female labor participation and entrepreneurship are particularly low (Fletcher, Pande and Moore, 2019; Lahoti and Swaminathan, 2016). While total female labor force participation has remained stagnant in India in the past three decades (Fletcher, Pande and Moore (2019), Figure 1), female entrepreneurship, has shown signs of progress, as we show in this paper. Moreover, female entrepreneurs tend to hire more female than male workers. Therefore, the advancement of female entrepreneurship could offer a way to promote general participation of women in the labor market. We utilize data from two waves of the Economic Census, which–in contrast to the World Enterprise Surveys–are nationally representative, and include the informal sector. The latter feature of the Census offers an important advantage relative to other data sets given that the majority of female-owned businesses are informal. Using this data and a model-based approach, we identify entry and operation frictions faced by female-owned firms and use counterfactual simulations to assess the productivity and welfare implications of various policy interventions.

Our analysis is guided by a simple, stylized model that nevertheless captures some important features of developing economies. The model features an economy with multiple industries. Within each industry, there are two sectors, a formal and an informal sector. Accounting for the informal sector is important, as it commands a large share of economic activity in developing countries (La-Porta and Shleifer, 2008, 2014; Ulyssea, 2018, 2020). Moreover, women, while under-represented among entrepreneurs, are over-represented in the informal sector (World Bank, 2012). Firms (entrepreneurs) need to pay an entry cost to operate in either sector and an additional registration cost to formalize[4]. Firms in the informal sector avoid paying the registration cost as well as taxes,

[4]The importance of these fixed entry and registration costs has been emphasized across many contexts. See Djankov, La Porta, Lopez-de Silanes and Shleifer (2002), and comprehensive reviews by Jayachandran (2020) and Quinn and Woodruff (2019).

but face a size-dependent penalty. This penalty captures both the cost of the actual penalty firms may have to pay if they are caught evading taxes and the implicit cost informal firms face by being denied access to formal finance, for which they have to be registered with a government agency[5]. There is only one input in production: labor. Conditional on entry (either in the formal or informal sector), firms make hiring decisions. We assume perfect competition in both product and labor markets.

Gender enters the model in four ways: First, we allow for male and female workers to be imperfect substitutes in the production function. Second, we allow for the productivity distributions of men and women entrepreneurs to be different. Third, we allow men and women entrepreneurs to face different entry and registration costs. Fourth, we assume that there are hiring frictions in the labor market that prevent firms from expanding, and allow these frictions to differ both by the gender of the firm owner and by the gender of the worker, i.e., we allow for women entrepreneurs to face different hiring frictions than men, and we also allow frictions to be different depending on whether the (male or female) entrepreneur hires a man versus a woman. This formulation is general and covers many of the factors that the literature has offered as potential explanations for gender inequality (e.g., legal barriers, cultural norms and attitudes, comparative advantage)[6]. We do not attempt to measure such factors or relate them to observables[7]. Rather, we model them as "wedges", and use the structure of the model in conjunction with the rich data of the Census to back them out and examine their implications for aggregate outcomes.

We have three key findings. First, even though the excess costs faced by female entrepreneurs (both on the intensive and extensive margins) decrease between the two rounds of the Census, they nevertheless remain substantial. For example, despite a significant decline, entry and formalization costs are 1.5-2 times larger for women entrepreneurs than for men. Similarly, on the intensive margin, women entrepreneurs face two to three times the cost of expanding their business through hiring

[5]See Beck and Hoseini (2014) and Nikaido, Pais and Sarma (2015).

[6]For comprehensive surveys of this literature, see Altonji and Blank (1999), Bertrand (2011), Blau et al. (2014).

[7]Some of the most important drivers of gender inequality in developing countries, i.e., norms and culture, may be difficult to measure. For the importance of such factors, see the work of Fernández (2013), Fernández and Fogli (2009), Fernández, Fogli and Olivetti (2004), and Ashraf, Bau, Nunn and Voena (2020) among others.

(both in the informal and formal sectors), as compared to their male counterparts. Second, average numbers mask substantial heterogeneity across industries and regions. For example, the excess fixed costs (both entry and formalization), though punitive across all sectors, are concentrated in manufacturing and services. Furthermore, excess entry barriers faced by women entrepreneurs are concentrated geographically in the Northern and Central states of India, consistent with what is reported in Evans (2020). On the intensive margin, the hiring frictions faced by women entrepreneurs are pervasive across all regions and sectors, but particularly high the services sector. Third, the only area where female entrepreneurs seem to have a significant advantage over their male counterparts is in hiring female workers (both in the informal and formal sectors). We show that this advantage is not driven by sectoral effects – it holds even within narrowly defined industries (at the 4-digit National Industry Classification level). This is especially important in a context like India, where female labor force participation is low and women workers are scarce. Put together, our results suggest that while there has been progress over time, women entrepreneurs face substantially larger costs to operate both on the extensive (entry and registration costs) and intensive (hiring workers) margins.

Given these results, we investigate the potential gains to the economy in a series of counterfactual scenarios where these barriers are eliminated. Specifically, we examine the impact of four affirmative action policies that aim to sequentially reduce the various excess costs faced by women entrepreneurs. We label these scenarios "affirmative action" policies because in all industry-regions where women entrepreneurs face higher costs than men, we equalize costs across women and men; however, in the one case where women have an advantage over men (i.e., attracting female workers), we do not eliminate this advantage. The first scenario we consider is a policy that eliminates excess formalization costs. In the second scenario, we eliminate excess entry costs as well, thus eliminating all excess fixed costs. The third scenario leaves fixed costs unchanged, but eliminates all intensive margin hiring frictions. Finally, the last scenario considers the elimination of all excess costs (both on the intensive and extensive margins).

The counterfactual simulations lead to four policy-relevant insights. First, removing the excess barriers has quantitatively meaningful impacts on the gender composition of entrepreneurs. Elimi-

nating the excess costs on the extensive (intensive) margin increases the fraction of women-owned firms in the economy by around two (six) times, from around 6% to 12% (38%). With no excess barriers (both on the intensive and extensive margins), around 56% of the firms in the economy would be owned by women, a nine-fold increase. Second, eliminating these barriers disproportionately helps female workers as compared to male workers. Under all counterfactual scenarios, the increase in real wages for women workers is greater than for male workers. This is intuitive given the low female labor force participation in India coupled with the fact that female workers are more likely to be hired by women entrepreneurs. Third, the counterfactual scenarios highlight the presence of low-productivity male entrepreneurs, who operate in the economy only because they do not face competition from more productive female-owned firms facing higher entry and operation barriers. Removing these barriers allows the marginal, higher-productivity woman entrepreneur to enter, thus reducing the misallocation of talent and resources in the economy. Lastly, this more efficient reallocation results in substantial gains in aggregate productivity and welfare (as measured by real income). The median increase in aggregate productivity (welfare) across regions is 2.2% (0.5%) if all excess fixed costs of entry and registration were to be eliminated. Removing hiring frictions results in substantially larger gains, namely an increase of aggregate productivity by 8.9% and of welfare by 12.3%. The larger gains under this scenario are consistent with the observation that since women entrepreneurs face substantive barriers on the intensive margin, removal of excess fixed costs alone does not result in significant gains. Lastly, removing both types of barriers increases aggregate productivity by 7% and welfare by 18%. These gains are large and suggest that promoting gender equality in entrepreneurship contributes to economic development.

Our paper speaks to a nascent literature focusing on the aggregate implications of eliminating gender-based distortions. While the literature on gender-based disparities is voluminous, studies focusing on the macroeconomic implications of such disparities are relatively scarce. The two studies that are closest in spirit to our work are the U.S.- focused paper by Hsieh, Hurst, Jones and Klenow (2019) and the cross-country analysis of Cuberes and Teignier (2016). However, our model differs from the models used in the aforementioned papers in several respects as it is geared towards capturing key features of developing economies, most importantly the prevalence of informality and its significance for women entrepreneurs.

The rest of the paper is organized as follows. Section 2 outlines the theoretical model. Section 3 discusses the data and provides descriptive evidence on the entrepreneurial landscape of India. Section 4 discusses the quantification of the model. Section 5 discusses the results, and in particular, the nature and extent of the barriers faced by women entrepreneurs. Section 6 examines the impacts of counterfactual affirmative action policies that eliminate these excess barriers. Section 7 concludes.

# 2 Theory

## 2.1 Setup

The economy consists of $J$ industries. Each industry $j$ has two sectors (denoted by $s$), the informal ($I$) and formal sector ($F$). Firms in both sectors produce a homogeneous product that is sold in a competitive market at price $p$. Hence, we do not allow for product differentiation across the formal and informal sectors. The only difference between firms in the formal and informal sectors is in their compliance with regulations.

There is a mass of potential male ($m$) and female ($f$) entrepreneurs $M_g$, $g \in \{m, f\}$ in each industry ($j$ will be dropped for notational convenience), and each entrepreneur is indexed by her/his individual productivity $z \sim G(z)$. Entrepreneurs (or firms - we will use the two terms interchangeably in the rest of the paper) make two decisions: (a) whether to pay a sector- and gender-specific fixed cost and enter sector $s \in \{I, F\}$, or not enter at all; and (b) conditional on entry, how many male and female workers to hire. The setup is static so that after entry, firms stay active forever[8]. Labor is the only input in production. We allow for men and women to be imperfect substitutes in the production function. A worker of gender $g \in \{m, f\}$ can be hired in a competitive labor market at a wage $\widetilde{w}^g$.

For notational consistency, we will henceforth use $x_{g,s}^{g'}$ to denote a variable $x$ (e.g., wages, labor, etc.) that refers to an entrepreneur of gender $g$, in sector $s$, and a worker of gender $g'$ (that is, the

[8] In fact, as reported by Hsieh and Klenow (2009), most firms in India are born small, never grow, and never die.

subscripts in our notation will refer to the gender of entrepreneurs and the superscripts to the gender of workers). We present the components of the model as follows. Section 2.2 presents the problem facing incumbent firms in the informal and formal sectors respectively. Section 2.3 discusses entry, and section 2.4 characterizes the model equilibrium.

## 2.2 Incumbents

Output $y$ of a firm with productivity $z$ is given by:

$$l = \left[ (l^m)^\gamma + (l^f)^\gamma \right]^{\frac{1}{\gamma}}$$

$$y = zl^\rho$$

where: the elasticity of substitution between male and female workers is given by $1/(1-\gamma)$, and $0 < \rho < 1$. The production decision in the formal and informal sector differs as follows. Firms in the formal sector have to pay a per-unit sales tax $t$.[9] Firms in the informal sector do not pay any taxes, but face a size-dependant penalty of being informal.

**Informal sector incumbents:** In the informal sector, firms do not pay taxes, but face a size-dependent penalty of operating in the informal sector. The profit maximization problem of a firm in the informal sector, owned by an entrepreneur $g$ with productivity $z$, is given by:

$$\pi_{gI} = \max_{\{l^m_{gI}, l^f_{gI}\}} pz\widetilde{l^\rho_{gI}} - w^m_{gI} l^m_{gI} - w^f_{gI} l^f_{gI}$$

where $\widetilde{\rho} = \rho/\theta$ and $\theta > 1$. The parameter $\theta$ captures a size-based penalty faced by the firm for operating in the informal sector.[10] This implies that it is less desirable for larger firms to remain informal. This is plausible given that according to the law, firms with more than 10 employees

[9] In reality, firms in the formal sector face many regulations in addition to sales taxes. We do not model these regulations in this paper, but use the per-unit sales tax as a shorthand for all measures that effectively reduce the net revenues of formal firms.

[10] An alternative way to model the size-based penalty is as a convex cost (as in Ulyssea (2018), for example). However, without separate data on revenues and costs, these two will be isomorphic in the model.

must be registered and pay taxes, and larger firms have a higher probability of being detected and penalized.[11]

The terms $w_{gI}^m$ and $w_{gI}^f$ denote the *effective* wages facing entrepreneurs in the informal sector. Entrepreneurs, especially women, may face frictions in hiring workers. We capture these in a reduced form way, as "wedges", i.e., additional costs over and above the nominal wages paid to workers. We assume that an entrepreneur with gender $g$, may face an additional per-unit cost $\tau_{gI}$ for hiring a worker in the informal sector, and a further cost $\tau_{gI}^f$ for hiring a female (relative to male) worker. These additional costs serve as a shorthand for many factors that may affect the hiring experience of women, on both sides of the labor market. For example, cultural norms may make it hard for some men to work for women, so that women entrepreneurs may have a harder time recruiting employees. Conversely, in some environments, cultural norms may inhibit women from working outside the home. But outside work may be considered more acceptable if the employer is a woman, making it easier for female entrepreneurs to recruit female workers. While such "cultural" factors and norms are considered important for employment decisions, they are hard, if not impossible, to quantify based on existing data. Accordingly, we do not attempt to measure them in this paper, but model them in a reduced form way as distortions that increase the effective cost of labor. It is important to note that since these additional costs will be estimated in the empirical part of the paper, in principle, they could also be zero or negative. While the model structure allows for them, it does not impose them.

The *effective* wages paid by an entrepreneur $g$ in the informal sector are given by $\boldsymbol{w_{gI}} \equiv \{w_{gI}^m, w_{gI}^f\} = (1+\tau_{gI})\{\widetilde{w}^m, (1+\tau_{gI}^f)\widetilde{w}^f\}$. The first order conditions imply that demand for male and female workers, optimal firm size, and profits are given by:

[11] In appendix C.2, we show that this size-based penalty can be re-written as a per-unit tax. As we explain in the Data Section, firms with fewer than 10 workers or fewer than 20 workers and no electricity do not have to pay taxes in India. Hence, failing to register is not illegal for such small firms. Nevertheless, such firms face an economic penalty in that they do not have access to formal credit channels. The parameter $\theta$ captures both the actual penalty larger firms may have to pay if they are caught evading taxes and the implicit penalty smaller informal firms may face because of financing constraints.

$$l_{gI}^{g'} = \left(\frac{w_{gI}^{g'}}{w_{gI}}\right)^{-\frac{1}{1-\gamma}} \times l_{gI}(z) \tag{1}$$

$$l_{gI}(z) = \left[\widetilde{\rho}\frac{p}{w_{gI}} \times z\right]^{\frac{1}{1-\widetilde{\rho}}} \tag{2}$$

$$\pi_{gI} = (1-\widetilde{\rho}) \times pzl_{gI}^{\widetilde{\rho}}(z) \tag{3}$$

$$\text{where: } w_{gI} = \left[(w_{gI}^{m})^{-\frac{\gamma}{1-\gamma}} + (w_{gI}^{f})^{-\frac{\gamma}{1-\gamma}}\right]^{-\frac{1-\gamma}{\gamma}}$$

See derivation in appendix C.1.

**Formal sector incumbents:** A firm in the formal sector, owned by an entrepreneur $g$ with productivity $z$, chooses labor to maximize variable profits given by:

$$\pi_{gF} = \max_{\{l_{gF}^{m}, l_{gF}^{f}\}} (1-t)pzl_{gF}^{\rho} - w_{gF}^{m}l_{gF}^{m} - w_{gF}^{f}l_{gF}^{f}$$

As with the informal sector, we assume that an entrepreneur $g$ faces hiring frictions, modeled as an additional cost $\tau_{gF}$ and $\tau_{gF}^{f}$ of hiring a worker and female worker respectively in the formal sector. Therefore, the *effective* wage is given by $\boldsymbol{w_{gF}} \equiv \{w_{gF}^{m}, w_{gF}^{f}\} = (1+\tau_{gF})\{\widetilde{w}^{m}, (1+\tau_{gF}^{f})\widetilde{w}^{f}\}$.

The first order conditions imply that demand for workers of gender $g'$, optimal firm size, and profits are given by:

$$l_{gF}^{g'} = \left(\frac{w_{gF}^{g'}}{w_{gF}}\right)^{-\frac{1}{1-\gamma}} \times l_{gF}(z) \tag{4}$$

$$l_{gF}(z) = \left[\rho\frac{(1-t)p}{w_{gF}} \times z\right]^{\frac{1}{1-\rho}} \tag{5}$$

$$\pi_{gF} = (1-\rho) \times (1-t)pzl_{gF}^{\rho}(z) \tag{6}$$

$$\text{where: } w_{gF} = \left[(w_{gF}^{m})^{-\frac{\gamma}{1-\gamma}} + (w_{gF}^{f})^{-\frac{\gamma}{1-\gamma}}\right]^{-\frac{1-\gamma}{\gamma}}$$

See derivation in appendix C.1.

## 2.3 Firm Entry

We now turn to the entry decision of firms. There is a mass of potential entrepreneurs ($M_g$) of a gender $g \in \{m, f\}$ in each industry $j$ (we remind the reader that we suppress the index $j$ for notational convenience, but that in the quantification of the model we allow all parameters to vary by industry and, when appropriate, by region). These entrepreneurs observe a pre-entry signal of productivity, $x \sim F(x)$, of their post-entry productivity $z$. We assume that $F(x)$ is continuous with support $(0, \infty)$, has finite moments, and is the same and independent for all firms within a gender and industry, but can vary across these. An entrepreneur $g$ pays a fixed sunk cost of entry (denominated in units of output) $E_{gI}$ to enter the informal sector, and $E_{gF} = E_{gI} + E_{gR} > E_{gI}$ to enter the formal sector, where $E_{gR}$ is a fixed cost of registration, i.e., formalization. As the notation suggests, we allow entry and formalization costs to differ by gender to accommodate the possibility that women face higher costs of bureaucracy, and more difficulty getting access to credit, electricity, and other services. The descriptive results based on the World Enterprise Surveys presented in Appendix B, as well as comprehensive reviews by Jayachandran (2020) and Quinn and Woodruff (2019) suggest that this possibility is borne out by the data.

After entry occurs, the entrepreneur draws her/his actual productivity from a conditional cumulative distribution $G(z|x)$, which is assumed to be continuous in both $z$ and $x$ but is strictly decreasing in $x$. This implies that a higher pre-entry signal $x$ is indicative of a better chance of getting a high post-entry draw $z$ as well. The entry structure is similar to Ulyssea (2018) and offers the advantage of allowing for an overlap of firm size distributions of the informal and formal sectors, which is a salient feature of the data. Figure A1 plots the fraction of 'formal' firms by firm size using the 1998 and 2005 rounds of the Economic Census of India (see section 3 for a detailed discussion on the data). As shown in figures A1(a) and A1(b), across both rounds, less than 10 percent of firms with fewer than 20 workers operate in the formal sector while over 70 percent of the larger firms (more than 60 workers) are formal[12]. While on average, larger firms are more likely to be formal, i.e., registered, we also see a substantial overlap in the firm size distribution of formal and informal firms in the data. Our entry structure allows for this overlap whereas entry decisions similar to Melitz

[12]Firms with less than 10 workers do not pay taxes anyway and we therefore categorize them as informal.

(2003) for example, would result in a perfect sorting of firms into either the formal or informal sector based on firm size, which is inconsistent with the data.

Therefore, an entrepreneur with pre-entry productivity $x$ (dropping $g$ for notational convenience) has an expected profit $V_s(x)$, $\forall s = \{I, F\}$ given by:

$$V_s(x, \boldsymbol{w}) = \int \pi_F(z, \boldsymbol{w}) dG(z|x)$$

where $\boldsymbol{w}$ denotes the vector of effective wages. Accordingly, (s)he will enter the formal sector as long as: $V_F(x) - pE_F \geq \max\{V_I(x) - pE_I, 0\}$, and the informal sector as long as: $V_I(x) - pE_I \geq \max\{V_F(x) - pE_F, 0\}$. Since entry into both sectors is positive, this implies that there is a threshold level of productivity in each sector $x_s^*$ such that:

$$\begin{aligned} V_I(x_I^*, \boldsymbol{w}) &= pE_I \\ V_F(x_F^*, \boldsymbol{w}) - V_I(x_F^*, \boldsymbol{w}) &= pE_R \end{aligned} \tag{7}$$

Entrepreneurs with $x < x_I^*$ will not enter at all, entrepreneurs with $x \in [x_I^*, x_F^*]$ will enter the informal sector, and entrepreneurs with $x > x_F^*$ will enter the formal sector.

## 2.4 Equilibrium

To close the model, we specify the demand side. We assume that representative households inelastically supply $\bar{\boldsymbol{L}} = \{\bar{L}^m, \bar{L}^f\}$ units of male and female workers at wages $\widetilde{\boldsymbol{w}} = \{\widetilde{w}^m, \widetilde{w}^f\}$ respectively. Consumers do not derive any disutility from work, cannot save, and hence consume their income. Total household income is accordingly given by $I = \widetilde{\boldsymbol{w}}\bar{\boldsymbol{L}} + \Pi + T$. $\Pi$ denotes the total profits of the economy net of the total entry costs $\sum_j p_j \sum_s \sum_g N_{gjs} E_{gjs}$, where $N_{gjs}$ is the mass of entrepreneurs of gender $g$ that enter sector $s$ of industry $j$. $T$ denotes the total taxes collected in the formal sector that are redistributed lump-sum to households. Lastly, we assume a simple demand system where the representative household consumes a composite bundle $C$ whose price is normalized to 1, and a constant share $\kappa_j$ of the household's income $I$ is spent on industry $j$.

The equilibrium in this economy is defined by the following conditions:

(i) the labor markets clear for both genders, i.e., $\bar{L}^g = \sum_j \sum_s L_j^g$, $\forall g = \{m, f\}$.

(ii) the zero-profit condition in (7) holds in both sectors with equality.

(iii) the goods market clears for each industry so that $\sum_s Y_{js} = \kappa_j I/p_j + \sum_s \sum_g N_{gjs} E_{gjs}$, where $Y_j$ is the total output produced in industry $j$ and $E_{gjs}$ are the total fixed costs paid by an entrepreneur $g$.

# 3 Data

Our primary data comes from two rounds of the Economic Census of India (EC) for 1998 and 2005.[13] The EC is meant to be a complete enumeration of all (formal and informal) non-farm business establishments in India in a given year. It is the only database in India that measures the unconditional distribution of establishment size. Other databases such as CMIE's Prowress Database, the Annual Survey of Industries (ASI) or the National Sample Surveys (NSS) only cover certain parts of the distribution and hence are unsuitable for our analysis.

Though it has uniform coverage, the EC has information only on a handful of variables, such as total number of workers, workers by gender, registration status, identity of the firm owner, 4-digit NIC industry code, and the source of finance for each establishment. It does not have information on output, capital, or profits, and the data are cross-sectional. We use the 2000 and 2005 rounds of the ASI and NSS to complement the EC when necessary. Formality in the model relates to firms paying taxes to the government. Accordingly, we define as "informal", those firms who have either not registered with the government or do not have to pay taxes (i.e., firms with fewer than 10 workers or fewer than 20 workers and no electricity). We omit from our analysis: (a) public-sector firms and co-operatives, since they do not have information on gender-ownership; (b) firms that

[13] We do not use the 2013 round of the Economic Census since it does not report whether a firm has registered or not. Hence in the 2013 data, we cannot measure informality, which is an important feature of India as well as most developing countries (LaPorta and Shleifer (2008), LaPorta and Shleifer (2014), Ulyssea (2018), Ulyssea (2020)).

do not hire any workers (owner-only firms). Lastly, we restrict our sample to the 18 major states of India[14], which cover 94.6 (97.25) percent of firms and 96 (97.5) percent of female-owned firms in 1998 (2005). Our final sample consists of 12.48 million firms in 1998 and 38.75 million firms in 2005.[15]

Table 1 presents summary statistics from the Economic Census data. We classify each firm into four categories based on gender (Male or Female) and formality (Formal or Informal). Columns (1), (3) and (5) report on the 1998 round of the EC, while columns (2), (4) and (6) report on the 2005 round. Four stylized facts stand out. First, more than 99 percent of firms (both male and female) operate in the informal sector, and the fraction of informal firms increased slightly between the two rounds of the EC from 99.33% in 1998 to 99.54% in 2005. Second, female-owned firms account for less than 10% of the total firms (6.59% in 1998 and 8.86% in 2005). Third, as reported in columns (3) and (4), female-owned firms are smaller than male-owned firms in the informal sector, but larger than male-owned firms in the formal sector. Lastly, from columns (5) and (6), female-owned firms employ more female workers compared to male-owned firms, and more so in the informal sector.

A comparison between 1998 and 2005 reveals further interesting patterns. The average number of workers (columns (3) and (4)) decreased for all categories between 1998 and 2005 suggesting a decline in entry costs; but the decline is particularly pronounced for formal firms (both male- and female-owned) suggesting a decline in the costs of formalization. This is consistent with a package of policy reforms (fiscal, financial, technology and infrastructural support) implemented in the early 2000s primarily for the micro, small and medium firms (Govt. of India, 2000). The fraction of female employees (columns (5) and (6)) remained relatively stable for male-owned firms, but increased substantially for female-owned firms, both in the formal and informal sectors. The latter pattern provides further support for the idea the female ownership matters for women's labor force participation.

To explore whether these patterns are driven by firm sorting either across space (districts in India),

[14]These states are Andhra Pradesh, Assam, Bihar, Chhattisgarh, Gujarat, Haryana, Jharkhand, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Odisha, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh (including Uttarakhand) and West Bengal.

[15]The change in the total number of firms over time comes from growth in the total number of firms over time, as well as a decline in the number of single-owned firms, which are dropped from our sample.

or across industries, we run regressions of the form:

$$y_{fjd} = \alpha_d + \beta_1 \text{Female}_f + \beta_2 \text{Formal}_f + \beta_3 \text{Female}_f \times \text{Formal}_f + \delta X_{fjd} + \alpha_j + \varepsilon_{fjd} \tag{8}$$

where $y_{fjd}$ is an outcome variable (either log-labor or fraction of female employees) for a firm $f$ that operates in industry $j$ and district $d$. "Female" and "Formal" are dummy variables that take the value 1 if the firm is female-owned and operates in the formal sector respectively, and 0 otherwise. Industry $j$ is the 4-digit National Industry Classification (NIC) code, and $X_{fjd}$ are a set of firm controls, such as access to electricity, dummy variables for different forms of financial access (formal, informal, government etc.), a dummy for whether the firm is primarily agriculture-based, and a dummy for whether the firm operates in a rural or urban area. We cluster standard errors at the district level.

Table A1 in the Appendix reports the results. Panel A of table A1 reports the regressions with district fixed effects ($\alpha_d$), but without industry fixed effects ($\alpha_j$), whereas Panel B adds industry fixed effects. Columns (1) and (3) report the results for the 1998 round of the EC while columns (2) and (4) report results for the 2005 round. The findings are consistent with the simple descriptive patterns discussed earlier. For example, as we can see from Panel B, in 2005, within each district and industry, female-owned informal firms are approximately 11.3 log-points (11%) smaller in size than male-owned informal firms, but 16.6 log-points (18%) larger than male-owned formal firms. In both the formal and informal sectors, female-owned firms employ more female workers than male-owned firms; in 2005, this difference is 52.4 pp in the informal sector, and 31.7 pp in the formal sector. Interestingly, a comparison of the estimates in Panel A to those in Panel B shows that the magnitude of these differences is hardly affected by the inclusion of industry fixed effects. This indicates that the advantage that female entrepreneurs have in hiring female workers is not driven by sectoral composition effects.

# 4 Model quantification

The purpose of quantifying the model is twofold. First, we estimate the hiring wedges and fixed costs of entry and registration. Second, we evaluate the impact of counterfactual policies that eliminate the excess entry, registration and hiring barriers faced by female entrepreneurs. Table 2 lists the model parameters. Given data limitations, we use a combination of calibration and estimation to set their values. Section 4.1 discusses the parameterization of the model and section 4.2 provides the implementation details.

## 4.1 Parameterization

We treat every state in India as a separate closed economy (or region $r$) and aggregate all four-digit industries into three broad industries (denoted by $j$), namely (i) agriculture and mining; (ii) manufacturing and (iii) services. As noted earlier, we use the 1998 and 2005 rounds of the Economic Census and allow for different parameters for each round.

We classify our parameters into two sets:

(a) Fundamental parameters $\{\Gamma, \Psi\} = \left\{\{\kappa_j, \rho_j, \gamma_j, t_{jr}\}, \{\theta_j, \{\sigma^2_{x,gj}, \sigma^2_{\varepsilon,gj},\}_{\forall g,j,r}\}\right\}$ and

(b) "Barriers" faced by entrepreneurs, such as fixed costs $\Upsilon = \{E_I, E_R\}_{\forall g,j,r}$ and hiring wedges $\Theta = \{\tau_{fI}, \tau_{fF}, \tau^f_{mF}, \tau^f_{fI}, \tau^f_{fF}\}_{\forall j,r}$.

The parameters in $\Gamma$ are determined based on statutory values or taken from the literature. The parameters in $\Psi$ and all barriers faced by entrepreneurs ($\Upsilon$,$\Theta$) are estimated.

We assume that the pre-entry productivity distribution for an entrepreneur $g$ in an industry $j$ (in the following, we drop $g$ and $j$ for notational ease) follows a log-normal distribution with mean 0 and variance $\sigma^2_x$, i.e., $x \sim \log N(0, \sigma^2_x)$. Further, we assume that post-entry productivity in the informal sector does not change, but in the formal sector, it differs from the pre-entry signal by a log-additive i.i.d shock $\varepsilon_F$, so that $z_F = x\varepsilon_F$, where $\varepsilon_F \sim \log N(0, \sigma^2_\varepsilon)$. The product of the

log-normal distributions therefore produces a gender-specific log-normal distribution of post-entry firm productivity $z$ in the formal sector.

Lastly, we normalize $\tau_{mI} = \tau_{mF} = 0$, i.e., we set the hiring barriers faced by male entrepreneurs equal to zero. This normalization is harmless, but implies that the interpretation of the hiring barriers faced by female entrepreneurs (i.e., $\tau_{fI}$ and $\tau_{fF}$) is *relative to* their male counterparts. Finally, we set any additional barriers faced by male entrepreneurs to hire female workers in the informal sector equal to zero, i.e., $\tau^f_{mI} = 0$. Accordingly, the additional barriers to hire female workers in the formal sector are $(1 + \tau^f_{mF})$ and $(1 + \tau^f_{mF}) \times (1 + \tau^f_{fF})$ for male and female entrepreneurs respectively. The parameters $\tau^f_{fI}$ and $\tau^f_{fF}$ therefore capture the additional barriers faced by female entrepreneurs to hire female workers in the informal and formal sectors respectively, *relative* to their male counterparts.

## 4.2 Implementation

This section discusses how the values of the various parameters are determined and provides some heuristic identification arguments. A more detailed discussion is provided in Appendix D. The parameters in $\Gamma$ are displayed in the first three rows of Table 3. They are determined based on statutory values or values taken from the literature as follows.

With regard to the demand structure, we fix $\kappa$ to be the total sales across all firms (as reported in the ASI and NSS) in a particular industry as a fraction of the total sales in the economy. This yields values of 0.216, 0.357, and 0.427 for agriculture and mining, manufacturing, and service industries respectively. The parameter $\rho = 0.738$, capturing (decreasing) returns to scale in the production function, is calibrated as the average labor share across firms in the ASI and NSS. The term $1/(1-\gamma)$ measures the elasticity of substitution between male and female workers in production. We set $\gamma = 0.68$ for manufacturing based on Acemoglu, Autor and Lyle (2004), and use the fraction of female workers in male-owned informal firms to identify $\gamma$ for the other two industries. The implied values (as reported in Table 3) are 0.42 for agriculture and 0.69 for services. These imply an elasticity of substitution of 1.72, 3.12 and 3.23 between male and female workers

for agriculture, manufacturing and services respectively.[16] Lastly, sales tax ($t$) for each industry $j$ in region $r$ is taken to be the average tax paid by a formal firm in that industry-region as reported in the ASI, which is a representative dataset for formal firms in India. As reported in Table 3, the tax rates we use are between 5-8%, consistent with the sales tax on most products in India during that period.

Since we normalize $\tau_{mI}$ and $\tau_{mF}$ to be equal to zero, from equations (2) and (5), the difference in the average firm size in the formal and informal sectors (conditional on entry) for male-owned firms help us identify $\widetilde{\rho}$, and hence $\theta$. Similarly, the variance in the log-firm size (conditional on entry) in the informal and formal sectors help us identify $\sigma_x^2$ and $\sigma_\varepsilon^2$ respectively. We accordingly estimate $\theta$, and the parameters of the productivity distribution for entrepreneurs in each industry ($\Psi$) from their counterparts in the data using: (i) average firm size; and (ii) variance in firm size of male- and female-owned firms in the formal and informal sectors. We employ a Simulated Minimum Distance (SMD) estimator, to minimize the distance between the simulated and actual moments in the data.

Having estimated the above parameters, we back out the fixed costs of entry and registration ($\Upsilon$) from the zero-profit conditions in (7).

The hiring distortions ($\Theta$) consist of two sets. The first set includes the distortions facing female entrepreneurs (relative to male) in hiring workers in the informal and formal sectors, i.e., $\tau_{fI}$ and $\tau_{fF}$ respectively. To identify these distortions, we compare the average firm size of male-owned to the average size of female-owned firms in the informal and formal sectors, conditional on entry (equations (2) and (5)). The second set of distortions includes the distortions entrepreneurs face in hiring female workers, i.e., $\{\tau_{mF}^f, \tau_{fI}^f, \tau_{fF}^f\}$. Let $R_{gs} = (w_{gs}^f/w_{gs}^m)^{-1/(1-\gamma)}$ be the ratio of female to male workers hired by an entrepreneur $g$ in sector $s$. From equations (1) and (4), this ratio of female to male workers hired by a male- and female-owned firm in the informal and formal sector helps us identify $\tau_{gs}^f$. Specifically, $R_{mF}/R_{mI} = (1+\tau_{mF}^f)^{-1/(1-\gamma)}$, $R_{fI}/R_{mI} = (1+\tau_{fI}^f)^{-1/(1-\gamma)}$, and $R_{fF}/R_{mF} = (1+\tau_{fF}^f)^{-1/(1-\gamma)}$.

Appendix D provides the precise equations used and a more detailed discussion of which moments

[16] As an independent validation of our approach, the $\gamma$ estimates for agriculture are consistent with Udry (1996).

in the data identify which parameters.

# 5 Results

**Productivity distribution and monitoring costs:** Table 3 reports the estimates for the productivity distribution $\{\sigma_x^2, \sigma_\varepsilon^2\}$ and informality penalty ($\theta$), which are estimated separately for each industry $j$. Column 1-3 report results across industries for 1998 while columns 4-6 report results for 2005. The size-based penalty of operating in the informal sector ($\theta$) ranges from 1.05 to 1.99 in 1998, and from 1.01 to 1.05 in 2005. The decrease in 2005 (among other things) is consistent with the increase in the share of informal firms between the two rounds, as discussed earlier. In Appendix C.2, we discuss how these estimates relate to size-based penalties (such as the probability of detection, etc.).

**Fixed costs of entry and formalization:** We now turn to the fixed costs of entry and formalization, reported in Table 4. Column 1 reports the values for 1998, column 2 reports the values for 2005, and column 3 reports the difference between the first two columns. We estimate these fixed costs separately for male and female entrepreneurs, and separately for each industry ($j$), region ($r$), and year ($t$). To make meaningful comparisons, we normalize the entry costs for a male entrepreneur in 1998 to have mean 1, so that relative comparisons across gender and over time can then be easily made.

As reported in the table, in 1998, female-owned firms faced on average 2.4 times the entry costs of male-owned firms. We find a sizeable reduction in entry costs for both genders over time, with those for female-owned firms declining almost twice as much as for male-owned firms. Specifically, entry costs declined by 23% for male entrepreneurs and by 44% for female entrepreneurs. However, despite this significant decline, entry-costs for female entrepreneurs were still 1.75 times those of male-owned firms in 2005 (column 2). Formalization costs were almost twice as large for female relative to male entrepreneurs in 1998. Despite the drastic reduction of formalization costs for both genders (23% for male and 39% for female entrepreneurs according to column 3), they remained

almost 1.5 times higher for female relative to male entrepreneurs in 2005 (column 2). Overall, these patterns suggest that while significant progress has been made over time, a wide gender gap remains as high fixed costs of entry and formalization continue to excessively penalize women. This resonates with the findings of Hyland, Djankov and Goldberg (2020), who report a similar pattern in their investigation of a broad set of gendered laws across countries.

The average numbers reported in Table 4 mask considerable heterogeneity across industries and Indian states, as well as over time. Using the 2005 estimates, we plot for each industry-region pair, the ratio of female to male entry costs, i.e., $E_{fI}/E_{mI}$. A ratio that is less than 1 indicates that female entrepreneurs face lower entry costs than their male counterparts, while a ratio greater than 1 would indicate the opposite. Figures 2(a) and 2(b) reveal substantial heterogeneity across industries as well as across states within an industry. Figure 2(a) shows that for almost all regions and industries, the ratio of female to male entry costs is greater than 1, with women facing more than three times the entry costs of men in some cases. Figure 2(b) breaks down the distribution across industries (i.e., shows the ratio of female to male entry costs across states within each industry). We see that women face barriers across all industries. For formalization costs, women face around around 1.5 times the formalization costs of men in 2005 on average, as discussed in Table 4 earlier. However, this ratio is greater than 3 in some industry-regions (Figure 2(c)), and seems particularly high in manufacturing (Figure 2(d)).

Figures 3(a) and 3(b) show the variation in women's relative entry and formalization costs (described earlier) across states in 2005. We see a stark pattern: relative to the southern states, the central and northern states have higher entry, and in many cases, formalization costs for women. These patterns are consistent with the geographical differences highlighted by Evans (2020).

Lastly, Figure 3(c), shows how the fixed costs have changed for male and female entrepreneurs between 1998 and 2005. Figure 3(c) shows a box plot with the 25-75th percentile and the median reported within the box, and with the intervals indicating the range. The substantial reduction in entry costs between the two years for both male and female entrepreneurs is evident. For women, the reductions are largest in agriculture (around 70%) and smallest in services. Median formalization costs decreased for both men and women across all industries as well, except possibly for services,

where there is large variation for women.

**Frictions in hiring workers:** For each industry $j$, region $r$, and year $t$, we quantify two types of barriers that distort firms' hiring decisions. First, $\tau_{fs}$ is the additional cost of hiring a worker for a female entrepreneur in sector $s$ relative to a male entrepreneur. We remind the reader that we have normalized $\tau_{ms} = 0$. Accordingly, the marginal cost faced by female entrepreneurs (relative to male entrepreneurs) is expressed in relative terms as $1 + \tau_{fs}$. $1 + \tau^f_{mF}$ is the additional marginal cost incurred by a male entrepreneur in hiring a female worker in the formal sector *relative* to the informal sector (again, we remind the reader that we have normalized $\tau^f_{mI} = 0$). Similarly, $1 + \tau^f_{fs}$ is the additional marginal cost incurred by women entrepreneurs in hiring female workers (relative to male entrepreneurs) in sector $s$.

As shown in Table 5, women entrepreneurs in the informal sector faced a 2.46 times higher cost per worker relative to their male counterparts in 1998. Although this cost was reduced by 24% by 2005 (column 3), the cost of hiring a worker for informal women entrepreneurs was still approximately double that for male entrepreneurs in 2005. In the formal sector, the marginal cost of hiring a worker was 2.9 times higher for female entrepreneurs as compared to their male counterparts in 1998. Despite a 28% reduction (column 3), female entrepreneurs still paid more than twice to hire a worker in the formal sector compared to male-owned firms in 2005.

As with the fixed costs, hiring frictions facing women exhibit substantial heterogeneity across industries, Indian states, and time. Figures 4(a) and 4(b) show the variation in the excess hiring barriers facing female-owned firms in the informal sector, while figures 4(c) and 4(d) show the variation in the formal sector. A ratio greater than 1 implies female-owned firms face a higher marginal cost in that sector, compared to male-owned firms. From figures 4(a) and 4(c), we see that this is true almost across all industry-regions (ratio is greater than 1), and especially pronounced in informal and formal services sector as well as the formal agricultural sector.

Turning to the gender composition of hired workers, the estimates indicate that this is the only area in which female entrepreneurs have an advantage, and more so in the informal sector. According to Table 5, female entrepreneurs in the informal sector incurred 4% (9%) lower costs to hire a female

(relative to male) worker in 1998 (2005), relative to male entrepreneurs. This advantage is still pervasive, but muted in the formal sector, where though the average is 1 in 2005, the median is less than 1. Figures 5(a)-5(c) further examine the heterogeneity in this hiring barrier, i.e., in the $1 + \tau_{fs}^{f}$ across industries and states. Of special note is that the advantage for female entrepreneurs in hiring women (relative to men) in the informal sector is quite substantial, over 15-20% in some industry-regions. This advantage is also present in all cases in the formal sector, except for the agricultural sector in a few states.

The comparative advantage that female entrepreneurs have in hiring female workers may reflect social norms and attitudes. For example, women workers may feel more comfortable working for other women; or, to the extent that women face resistance from male members of their household if they seek work outside the home, such resistance may be less pronounced in cases where they work for other women.

Finally, in Figure 5(d), we show the change of $1 + \tau_{fs}$ across the three industries for both the informal and formal sectors. The relative barriers facing women in both the formal and informal sectors have decreased in agriculture and manufacturing, but not as much in services. Turning to the relative barriers women entrepreneurs face in hiring women workers, i.e., $1 + \tau_{fs}^{f}$, we do not see any substantial changes at the median (see Figure 5(e)), which is intuitive, since women-owned firms already had an advantage in hiring women in the informal sector, as discussed previously.

To summarize the above discussion, these results suggest that while the excess barriers faced by female entrepreneurs have been reduced over time, there nevertheless remains a substantial gender gap across all industries and regions. The only exception is in the hiring of female workers, where female entrepreneurs appear to have an advantage.

**Model fit:** Tables 6 and 7 show the fit of the model for the 2005 data[17]. We start by discussing the model fit for the four types of firms, namely male-owned and female-owned firms in the formal and informal sectors. We focus on three types of moments, namely the fraction of firms, the fraction

[17]We show the fit only for 2005 since this is the data that we use to evaluate counterfactual policies in the next section.

of workers, and the average firm size (number of hired workers) in each group. We accordingly have a total of 12 moments. Since these moments are generated at the industry-region level for each firm type, we report the average and standard deviation of each of these moments across industry-regions in Table 6. The model fits the data moments across all categories very well. In Table 7, we aggregate the data and report for same moments, the average across sectors or gender. Again, the table indicates that the model fits the data well. Figure A2 in the appendix suggests a good fit across all disaggregate moments, which are at the industry-region for each firm type.

# 6 Impact of Affirmative Action Policies

Apart from quantifying the various types of barriers faced by female entrepreneurs, the advantage of the outlined theoretical framework is that it allows us to evaluate the aggregate effects of counterfactual affirmative action policies (in general equilibrium). In particular, we evaluate the impact of four such policies that sequentially eliminate the excess barriers faced by female entrepreneurs on both the extensive (i.e., entry and formalization) and intensive (i.e., expansion through hiring) margins. This exercise allows us to identify the barriers that are most consequential for productivity and welfare. We consider the following scenarios that eliminate:

(i) **Excess formalization costs:** In this scenario, we eliminate excess formalization costs faced by women entrepreneurs, i.e., we set $E_{fR} = \min\{E_{mR}, E_{fR}\}$.
(ii) **Excess fixed costs:** We eliminate all excess fixed costs (both entry and formalization) for female entrepreneurs, i.e., in addition to the previous policy, we also set $E_{fI} = \min\{E_{mI}, E_{fI}\}$.
(iii) **Excess hiring barriers:** Under this policy, we eliminate excess hiring barriers on the intensive margin, but not the fixed costs, i.e, $\{E_{gI}, E_{gR}\}$ are set to their baseline values, but we set $\tau_{fs} = \min\{\tau_{fs}, 0\}$ and $\tau^f_{fs} = \min\{\tau^f_{fs}, 0\}$, for $s = \{I, F\}$.
(iv) **All excess barriers:** Lastly, we eliminate all excess fixed costs as well as all hiring barriers.

We examine the effects of these policies on various outcomes, such as the fraction of female-owned firms (in both the formal and informal sectors), the entry thresholds (and hence average produc-

tivity), and real wages for male and female workers. Finally, we examine their impact on aggregate productivity and real income. For each region, aggregate productivity is measured as the average productivity of its firms across industries. Given our preference structure, real income is a natural candidate for measuring welfare. Aggregate consumption (or utility) is given by: $C = \prod_j c_j^{\alpha}{}_j = I/P$, where $P = \prod_j (\alpha_j/p_j)^{\alpha_j}$ is normalized to 1. Therefore, compared to the baseline, $C_0$, the percentage change in utility in any counterfactual scenario will be equivalent to the percentage change in real income.

**Removing excess formalization costs:** We begin by discussing the impact of removing excess formalization costs for female-owned firms. From Figure 6(a), this policy increases the median fraction of female-owned firms in the formal sector from less than 1% to 2%. However, the median fraction of female-owned firms in the informal sector or in the economy as a whole does not change. This policy has little effect on the median real wage (Figure 6(b)), productivity (Figures 6(c), 6(d), and 7(a)) or welfare (Figure 7(b)).

**Removing all excess fixed costs (entry and formalization):** From section 5, we know that entry costs are substantially higher for female- than for male-owned firms. We therefore examine the impact of removing all excess fixed costs (both entry and formalization costs) for female-owned firms. From Figure 6(a), this policy increases the median fraction of female-owned firms in the informal (formal) sector to 13% (2.1%). The median fraction of female-owned firms in the economy also doubles to 12%. Real wages for female workers also increase by almost 1.5 times relative to those of male workers (Figure 6(b)).

Turning to productivity, while the average productivity of male entrepreneurs changes very little (compared to the baseline), the average productivity of female entrepreneurs *decreases* by 13.2% percent. Though counter-intuitive at first, these effects can be perfectly rationalized by Figure 6(d), which shows, for the baseline as well as for all counterfactual scenarios, the productivity of the marginal entrepreneur, i.e., the entrepreneur who makes zero profits in expectation, after paying the fixed cost of entry and formalization. To make the comparison easier, we normalize the productivity

of the marginal male entrepreneur to be 1 at baseline. It is interesting to note that at baseline, the marginal woman entrepreneur has to be almost twice as productive as her male counterpart. The removal of the excess fixed costs allows more women to enter, presenting male entrepreneurs with more competition. Accordingly, the productivity of the marginal female (male) entrepreneur decreases (increases). This implies that the average female (male) entrepreneur is now less (more) productive. Moreover, the set of female entrepreneurs who now enter are more productive than their male counterparts, which translates into aggregate productivity gains, as shown in Figure 7. The median increase in the average productivity is 2.2% across all Indian states (with a 25th-75th percentile increase of 2.1-2.4%), and median welfare increases by 0.5% (with a 25th-75th percentile increase by 0.4-0.6%).

**Removing excess hiring frictions:** As discussed in the previous section, hiring frictions for female-owned firms are substantial and quantitatively important, in both the formal and informal sectors. Therefore, we now construct a counterfactual where we do not change the baseline fixed costs, but remove the excess hiring costs for female-owned firms. One of the most notable impacts of this policy is that it helps female-owned firms enter the formal sector (Figure 6(a)). The fraction of informal (formal) firms that are female-owned increases to 34% (44%). The share of female-owned firms in the economy increases to 38%. Further, real wages for both male and female workers increase substantially by 11% and 27% respectively (Figure 6(b)), suggesting that female workers gain relatively more than male workers. Finally, the average productivity of male-owned (female-owned) firms increases (decreases) by 8.1% (26%). This translates into a 8.9% median increase across regions in the average productivity (with the 25th-75th percentile increasing by 8.5-10.15%), and a 12.25% increase in aggregate welfare (with the 25th-75th percentile increasing by 9.5-15.8%). Put together, these results suggest that removing hiring barriers implies substantial gains in both productivity and welfare. These gains are higher than those realized by removing excess fixed costs.

**Removing all excess barriers:** The last counterfactual we consider is one where we remove all excess barriers for female entrepreneurs, both on the extensive and intensive margins (though we do not change anything in cases where female entrepreneurs have an advantage over their male

counterparts). Removing all excess costs implies that female-owned firms now form a substantial share of firms in both the informal and formal sectors, with 56% of firms in the economy now owned by women (Figure 6(a)). Further, real wages for both male and female workers increase substantially by 15.4% and 29.8% respectively (Figure 6(b)), which implies that female workers gain almost twice as much as male workers.

The large wage gains for female relative to male workers arise in our model because the supply of female workers is fixed. Hence, as demand for them increases (due to the increased entry of female entrepreneurs), female wages have to increase by more than male wages to match demand with supply. Our primary motivation for this modeling choice was the observation that while the number of female-owned firms increased during this time period (as discussed in section 3), female labor force participation in India has not changed (and if anything, has slightly declined). A more general model would allow for the supply of female workers to increase in response to higher wages. While this supply response would mitigate the wage gains we currently estimate, we note that it would result in an increase in female labor force participation which would imply additional welfare gains for women and the economy as a whole. Our stylized model misses this positive labor force participation effect while exaggerating the wage response. Since these two biases go in opposite directions, it is therefore possible that the welfare gains in a more general model would not be substantially different from the ones we currently estimate. Of course, we cannot assess the net effect without developing a more general framework, an endeavor we leave to future research.

The average productivity of male-owned (female-owned) firms increases (decreases) by 10.4% (33%) Figure 6(c). However, the marginal female entrepreneurs who now enter the economy are on average *more* productive than the male entrepreneurs who exit (Figure 6(d)). This implies that overall, there is an increase in the average productivity of the economy (Figure 7(a)). The median increase across regions in the average productivity of a firm is 7% (with the 25th-75th percentile increasing by 6.6-7.9%), and aggregate welfare increases by 17.7% (with the 25th-75th percentile increasing by 15.2-22.1%).

**Discussion:** The counterfactual scenarios considered above lead to a few key policy-relevant insights. First, the barriers faced by women entrepreneurs are substantial, both on the intensive and extensive margins. Their removal has quantitatively meaningful impacts on aggregate productivity and welfare. On the extensive margin, policies targeting removal of excess entry costs, rather than formalization costs, have larger effects. But policies that target the intensive margin have a far greater impact than those targeting the extensive margin. Second, removing excess barriers not only helps female entrepreneurs, it also benefits female workers more relative to male workers. This can have potentially important implications in the Indian setting, where female labor force participation is low. Third, the counterfactual scenarios highlight the presence of low productivity male-owned firms, who operate in the economy only because they do not face competition from female-owned firms. The latter cannot enter or grow post-entry because they face excessive barriers. Removing these barriers results in the marginal (low-productivity) male entrepreneurs exiting the market, allowing for the marginal (higher-productivity) female entrepreneurs to enter. Put together, affirmative action policies that eliminate all excess barriers result in substantial aggregate gains in productivity and real income.

# 7 Conclusion

We conclude with some final thoughts. Our results demonstrate that eliminating excess barriers to entrepreneurship facing women is beneficial not only to women, but to the entire economy. However, they do not speak to the question of which specific policies would lead to elimination of such barriers. Barriers at both the extensive and intensive margins are modeled as "wedges" in our framework, and are identified based on the data patterns in the Economic Census related to entrepreneurship. Further research needs to relate the estimated wedges to actual policies to assess which interventions are most effective in reducing them. The main challenge is that several of the barriers women face are not due to legal constraints, but to norms and attitudes that are more difficult to measure. Combining case studies of specific interventions to empower women with our framework can be a fruitful approach towards assessing not only whether such interventions were

successful, but also their aggregate impacts.

# References


**Acemoglu, Daron, David H Autor, and David Lyle**, "Women, War, and Wages: The Effect of Female Labor Supply on the Wage Structure at Mid-century," *Journal of Political Economy*, 2004, *112* (3), 497–551.

**Altonji, Joseph G and Rebecca M Blank**, "Race and Gender in the Labor Market," *Handbook of Labor Economics*, 1999, *3*, 3143–3259.

**Ashraf, Nava, Natalie Bau, Nathan Nunn, and Alessandra Voena**, "Bride Price and Female Education," *Journal of Political Economy*, 2020, *128* (2), 591–641.

**Beck, Thorsten and Mohammad Hoseini**, "Informality and Access to Finance: Evidence from India," *Center Discussion Paper Series*, 2014.

**Bertrand, Marianne**, "New Perspectives on Gender," *Handbook of Labor Economics*, 2011, *4*, 1543–1590.

**Blau, Francine D., Marianne A. Ferber, and Anne E. Winkler**, "The Economics of Women, Men and Work," *Upper Saddle River, New Jersey: Pearson, 7th Edition*, 2014.

**Cuberes, David and Marc Teignier**, "Gender Inequality and Economic Growth: A Critical Review," *Journal of International Development*, 2014, *26* (2), 260–276.

**_ and _**, "Aggregate Effects of Gender Gaps in the Labor Market: A Quantitative Estimate," *Journal of Human Capital*, 2016, *10* (1), 1–32.

**Djankov, Simeon, Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer**, "The regulation of entry," *The Quarterly Journal of Economics*, 2002, *117* (1), 1–37.

**Evans, Alice**, "Why are North and South India so Different on Gender?," 2020. https://www.draliceevans.com/post/why-are-southern-north-eastern-indian-states-more-gender-equal.

**Fernández, Raquel**, "Cultural Change as Learning: The Evolution of Female Labor Force Participation over a Century," *American Economic Review*, 2013, *103* (1), 472–500.

_ , **Alessandra Fogli, and Claudia Olivetti**, "Mothers and Sons: Preference Formation and Female Labor Force Dynamics," *The Quarterly Journal of Economics*, 2004, *119* (4), 1249–1299.

_ **and** _ , "Culture: An Empirical Investigation of Beliefs, Work, and Fertility," *American Economic Journal: Macroeconomics*, 2009, *1* (1), 146–77.

**Fletcher, Erin, Rohini Pande, and Charity Maria Troyer Moore**, "Women and Work in India: Descriptive Evidence and a Review of Potential Policies," 2019.

**Govt. of India**, "Comprehensive Policy Package for Small Scale Industries," 2000. http://www.dcmsme.gov.in/MSME-DO/policypkg.htm.

**Hsieh, Chang-Tai and Peter J Klenow**, "Misallocation and manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 2009, *124* (4), 1403–1448.

_ , **Erik Hurst, Charles I Jones, and Peter J Klenow**, "The Allocation of Talent and US Economic Growth," *Econometrica*, 2019, *87* (5), 1439–1474.

**Hyland, Marie, Simeon Djankov, and Pinelopi Koujianou Goldberg**, "Gendered Laws and Women in the Workforce," *American Economic Review: Insights*, 2020.

**Jayachandran, Seema**, "Microentrepreneurship in developing countries," 2020.

**Lahoti, Rahul and Hema Swaminathan**, "Economic Development and Women's Labor Force Participation in India," *Feminist Economics*, 2016, *22* (2), 168–195.

**LaPorta, R. and A. Shleifer**, "The Unofficial Economy and Economic Development," *Brooking Papers on Economic Activity*, 2008, *105* (3), 473–522.

_ **and** _ , "Informality and Development," *Journal of Economic Perspectives*, 2014, *28* (3), 109–126.

**Melitz, Marc J**, "The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 2003, *71* (6), 1695–1725.

**Nikaido, Yuko, Jesim Pais, and Mandira Sarma**, "What Hinders and What Enhances Small Enterprises' Access to Formal Credit in India?," *Review of Development Finance*, 2015, *5* (1), 43–52.

**Quinn, Simon and Christopher Woodruff**, "Experiments and Entrepreneurship in Developing Countries," *Annual Review of Economics*, 2019, *11*, 225–248.

**Restuccia, Diego and Richard Rogerson**, "The causes and costs of misallocation," *Journal of Economic Perspectives*, 2017, *31* (3), 151–74.

**Udry, Christopher**, "Gender, Agricultural Production, and the Theory of the Household," *Journal of Political Economy*, 1996, *104* (5), 1010–1046.

**Ulyssea, Gabriel**, "Firms, Informality, and Development: Theory and Evidence from Brazil," *American Economic Review*, 2018, *108* (8), 2015–47.

_ , "Informality: Causes and Consequences for Development," *Annual Review of Economics*, 2020, *12* (1), 525–546.

**World Bank**, *World Development Report 2012 : Gender Equality and Development* 2012. Washington, DC.

**World Bank**, "World Bank Enterprise Surveys," 2020. Data from http://www.enterprisesurveys.org.

# Tables and Figures

Table 1: Summary statistics

| Firm type | Total firms | | Firm size | | Frac. Female Emp. | |
|---|---|---|---|---|---|---|
| | 1998 | 2005 | 1998 | 2005 | 1998 | 2005 |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Male, Informal | 11.58 | 35.15 | 3.29 | 3.13 | 0.19 | 0.10 |
| | (92.75%) | (90.71%) | (3.68) | (2.35) | (0.25) | (0.21) |
| Male, Formal | 0.08 | 0.17 | 77.47 | 60.83 | 0.21 | 0.24 |
| | (0.65%) | (0.43%) | (438.82) | (153.52) | (0.25) | (0.29) |
| Female, Informal | 0.82 | 3.42 | 2.96 | 2.92 | 0.57 | 0.75 |
| | (6.57%) | (8.83%) | (2.98) | (2.25) | (0.33) | (0.37) |
| Female, Formal | 0.003 | 0.01 | 97.87 | 62.44 | 0.45 | 0.59 |
| | (0.02%) | (0.03%) | (1118.2) | (109.24) | (0.37) | (0.41) |
| Total | 12.48 | 38.75 | | | | |

*Notes:* A firm is classified as "informal" if it is either not registered with the govt. or does not have to pay taxes (fewer than 10 workers or fewer than 20 workers without electricity), and "formal" otherwise. Numbers in columns (1)-(2) are reported in millions. Percentage of the total are reported in parentheses below. Firm size in columns (3) and (4) report the average employees within a firm. Frac. of Female Emp. in columns (5) and (6) are the fraction of female employees within a firm. Standard deviations are reported in parentheses below.

Table 2: List of parameters

| Parameters | Level | Details |
|---|---|---|
| $\kappa$ | Industry | Share of each industry in consumer demand. |
| $\rho$ | Economy | Returns to scale in production |
| $\gamma$ | Industry | $\frac{1}{1-\gamma}$ is elasticity of substitution b/w male and female workers |
| $\theta$ | Industry | Size-based penalty for being in the informal sector |
| $t$ | Industry-State | Average sales tax |
| $\{\sigma_x^2, \sigma_\varepsilon^2\}$ | Industry-Gender | Productivity distribution |
| $\{E_I, E_R\}$ | Industry-Gender-State | Fixed costs of entry and formalization |
| $\{\tau_{gs}, \tau_{gs}^f\}$ | Industry-Gender-State | Hiring barriers |

Table 3: Parameter values

| Year: | | 1998 | | | 2005 | |
|---|---|---|---|---|---|---|
| Industry: | Agri. | Manfc. | Services | Agri. | Manfc. | Services |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $\kappa$ | 0.216 | 0.357 | 0.427 | 0.216 | 0.357 | 0.427 |
| $t$ | 0.05 | 0.07 | 0.07 | 0.05 | 0.07 | 0.08 |
| $\gamma$ | 0.42 | 0.68 | 0.69 | 0.42 | 0.68 | 0.69 |
| $\theta$ | 1.99 | 1.05 | 1.07 | 1.01 | 1.03 | 1.05 |
| $\sigma_{x,m}$ | 0.01 | 0.37 | 0.26 | 0.13 | 0.33 | 0.25 |
| $\sigma_{x,f}$ | 0.46 | 0.47 | 0.37 | 0.26 | 0.38 | 0.36 |
| $\sigma_{\varepsilon,m}$ | 0.23 | 0.25 | 0.26 | 0.22 | 0.23 | 0.21 |
| $\sigma_{\varepsilon,f}$ | 0.18 | 0.26 | 0.19 | 0.19 | 0.22 | 0.16 |

*Notes:* Each row reports the average value for the parameter across regions. Columns (1)-(3) reports the parameter value for each industry in 1998 while columns (4)-(6) report the values for each industry in 2005.

Table 4: Estimates for fixed costs

| | 1998 | 2005 | (2)-(1) |
|---|---|---|---|
| | (1) | (2) | (3) |
| *Normalized Entry Costs ($E_I$):* | | | |
| Male | 1.00 | 0.77 | -0.23*** |
| | (0.35) | (0.21) | |
| Female | 2.43 | 1.35 | -1.08*** |
| | (1.13) | (0.32) | |
| *Normalized Formalization Costs ($E_R$):* | | | |
| Male | 1.00 | 0.77 | -0.23*** |
| | (0.37) | (0.34) | |
| Female | 1.84 | 1.13 | -0.71*** |
| | (1.24) | (0.80) | |

*Notes:* Each row reports the average (across industries and regions) value of each parameter with standard deviations in parentheses below. Entry and registration costs across all firms have been normalized so that male-owned firms in 1998 have mean 1. Columns (1) and (2) report the value for 1998 and 2005 respectively. Column (3) reports the difference between columns (2) and (1). * for p<0.1, ** for p<0.05 and *** for p<0.01.

Table 5: Estimates for hiring wedges

| | 1998 | 2005 | (2)-(1) |
|---|---|---|---|
| | (1) | (2) | (3) |
| $1+\tau_{fI}$ | 2.46 | 1.87 | -0.59*** |
| | (0.61) | (0.22) | |
| $1+\tau_{fF}$ | 2.88 | 2.07 | -0.81*** |
| | (1.28) | (0.63) | |
| $1+\tau^{f}_{mF}$ | 1.00 | 0.80 | -0.21*** |
| | (0.35) | (0.28) | |
| $1+\tau^{f}_{fI}$ | 0.96 | 0.91 | -0.05*** |
| | (0.04) | (0.04) | |
| $1+\tau^{f}_{fF}$ | 1.07 | 1.00 | -0.07 |
| | (0.5) | (0.77) | |

*Notes:* Each row reports the average (across industries and regions) value of each parameter with standard deviations in parentheses below. Columns (1) and (2) report the values for 1998 and 2005 respectively. Column (3) reports the difference between columns (2) and (1). * for p<0.1, ** for p<0.05 and *** for p<0.01.

Table 6: Model Fit I

| | Male entrepreneurs | | Female entrepreneurs | |
|---|---|---|---|---|
| | Data | Model | Data | Model |
| | (1) | (2) | (3) | (4) |
| *Panel A: Fraction of firms* | | | | |
| Informal | 0.90 | 0.88 | 0.09 | 0.10 |
| | (0.11) | (0.16) | (0.11) | (0.13) |
| Formal | 0.01 | 0.02 | 0.00 | 0.00 |
| | (0.01) | (0.06) | (0.00) | (0.01) |
| *Panel B: Fraction of workers* | | | | |
| Informal | 0.84 | 0.80 | 0.08 | 0.05 |
| | (0.12) | (0.25) | (0.09) | (0.04) |
| Formal | 0.08 | 0.13 | 0.00 | 0.01 |
| | (0.09) | (0.22) | (0.01) | (0.03) |
| *Panel C: Average firm size* | | | | |
| Informal | 3.37 | 2.95 | 3.17 | 2.47 |
| | (0.48) | (0.84) | (0.52) | (0.64) |
| Formal | 56.39 | 54.89 | 58.97 | 59.25 |
| | (17.10) | (29.30) | (36.44) | (40.53) |

*Notes:* Each row reports the average value for the parameter with the standard deviation in parentheses.

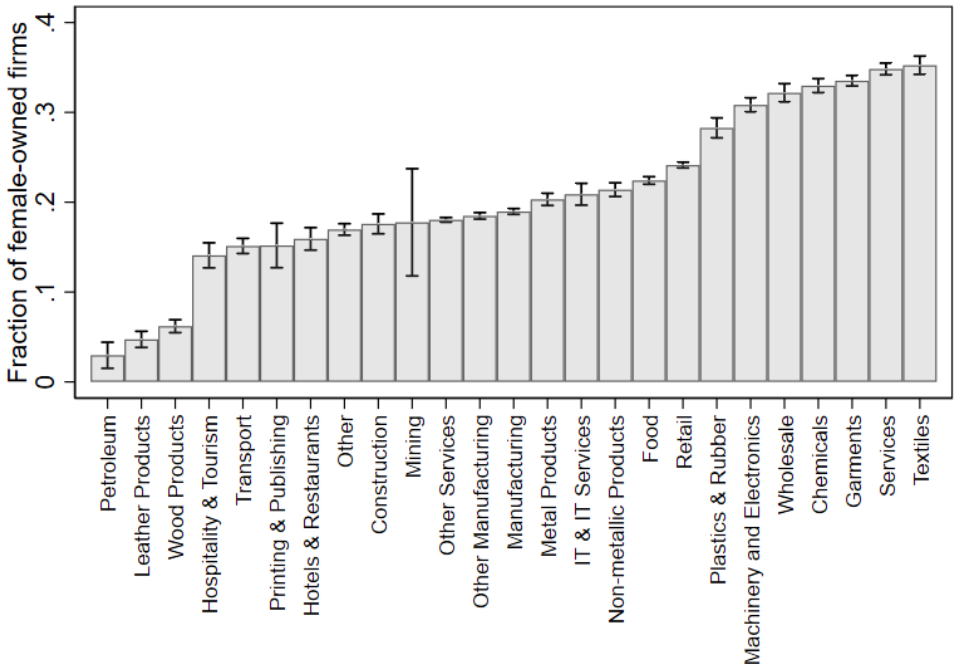
Table 7: Model Fit II

| | Frac. of firms | | Frac. of workers | | Average firm size | | | |
|---|---|---|---|---|---|---|---|---|
| | Informal | Female owned | Informal | Female owned | Informal | Formal | Male owned | Female owned |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Data | 0.99 | 0.09 | 0.91 | 0.08 | 3.32 | 56.64 | 3.74 | 3.43 |
| | (0.01) | (0.11) | (0.09) | (0.1) | (0.45) | (17.68) | (0.81) | (0.63) |
| Model | 0.98 | 0.10 | 0.86 | 0.06 | 2.89 | 59.35 | 4.52 | 3.02 |
| | (0.06) | (0.13) | (0.24) | (0.06) | (0.84) | (56.07) | (5.05) | (2.43) |

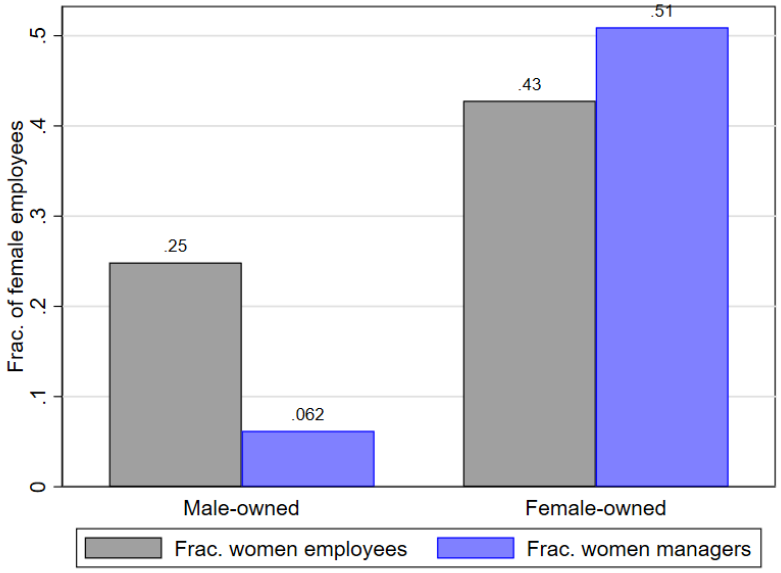
*Notes:* Each row reports the average value with the standard deviation in parentheses.

Figure 1: Share of women entrepreneurs, employees and managers

(a) Fraction of female entrepreneurs across industries

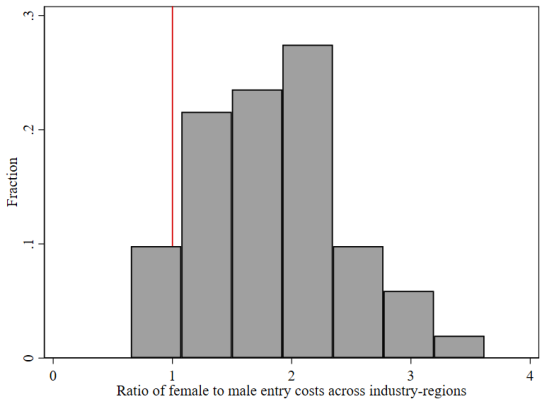



(b) Fraction of women employees and managers

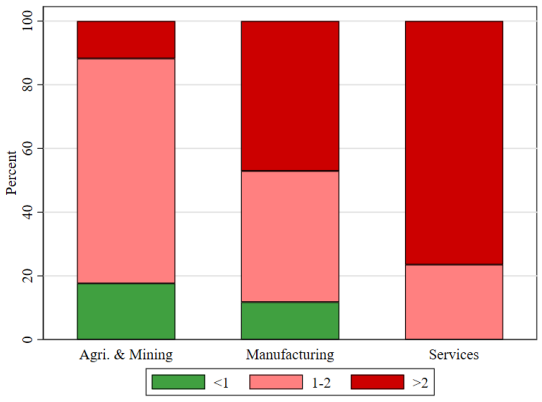



*Notes:* Both figures use the World Bank Enterprise Surveys. Figure 1(a) plots the average fraction of female-owned firms across 25 sectors. Figure 1(b) plots the fraction of women employees and the probability that the top manager in a firm is a female.

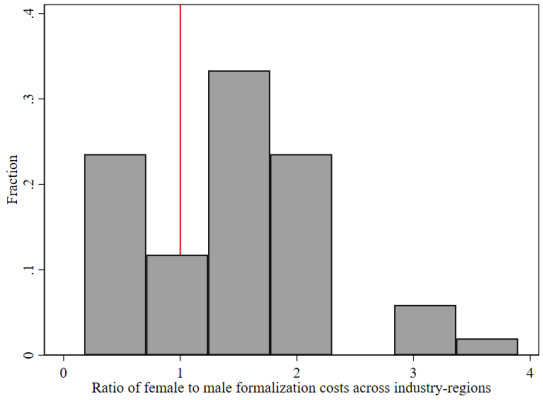
Figure 2: Excess entry barriers for female entrepreneurs across industries



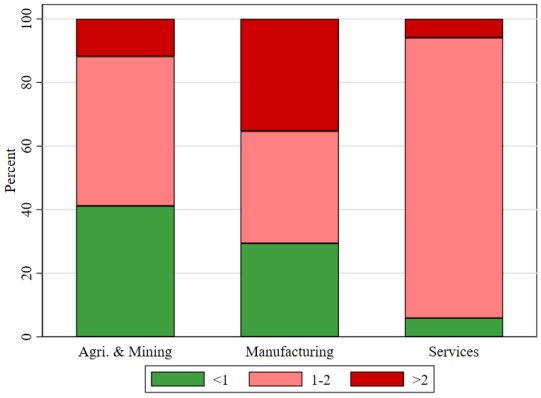

(a) Ratio of female-male entry costs




(b) Ratio of female-male entry costs within industries



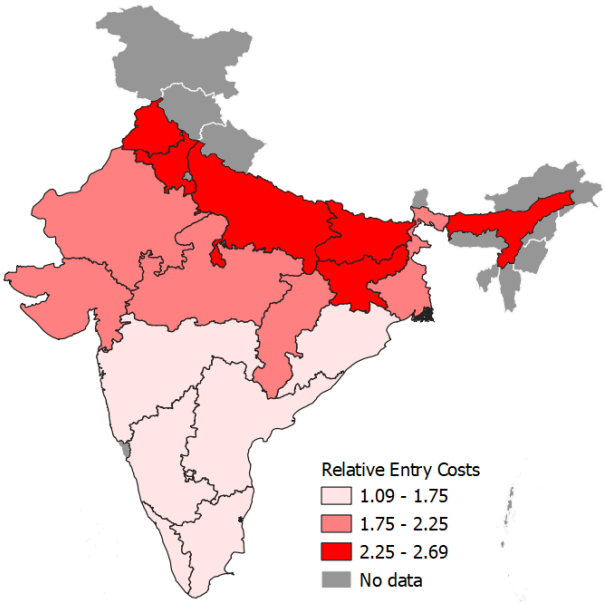

(c) Ratio of female-male formalization costs



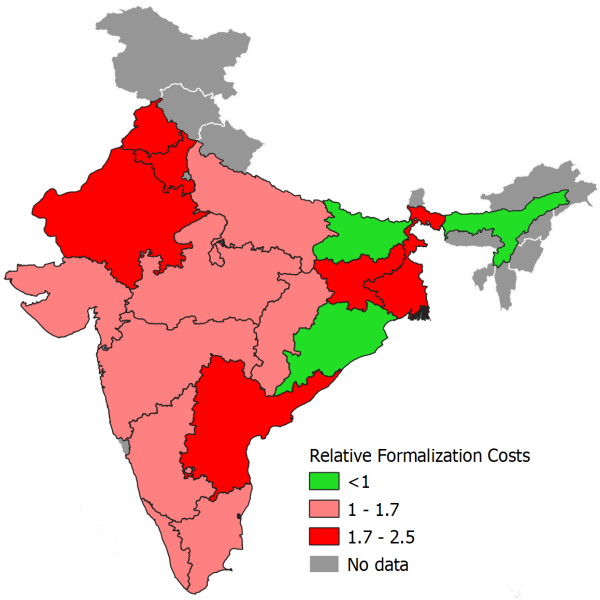

(d) Ratio of female-male form. costs within industries

*Notes:* Figures 2(a) and 2(c) show the distribution of the female to male ratios of entry and formalization costs using the 2005 data. Figures 2(b) and 2(d) then show the breakdown within industry across states. Green corresponds to a ratio less than 1 (favorable for females), pink represents a ratio between 1-2, and red represents a ratio greater than 2.

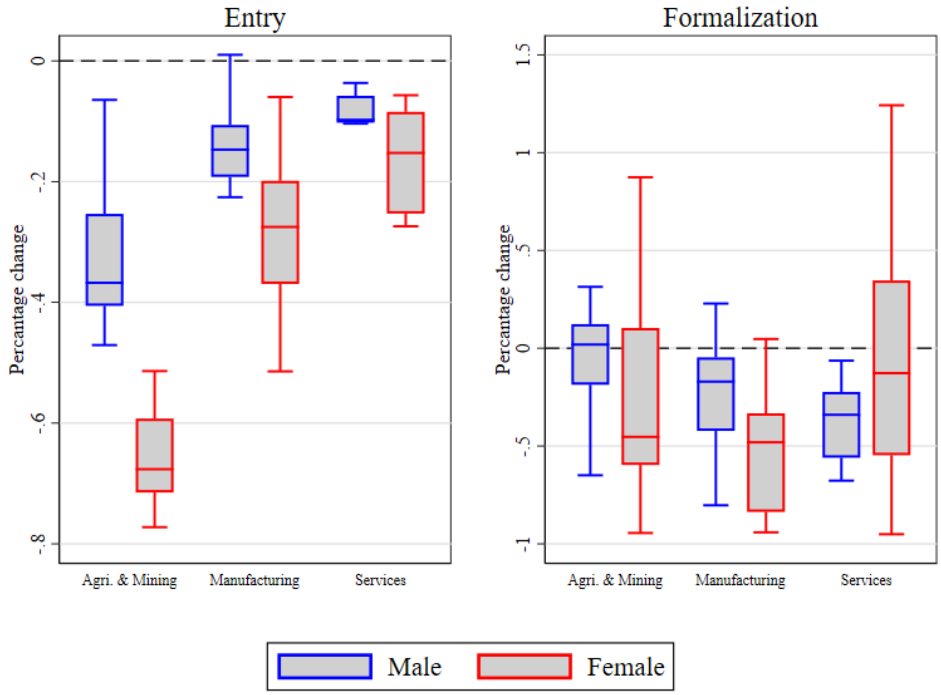
Figure 3: Excessive entry barriers for women entrepreneurs across states




(a) Relative entry costs



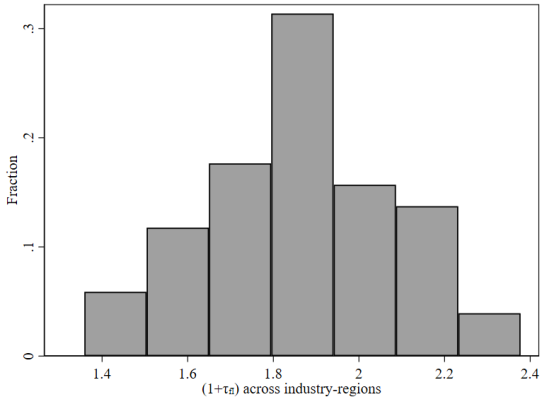

(b) Relative formalization costs



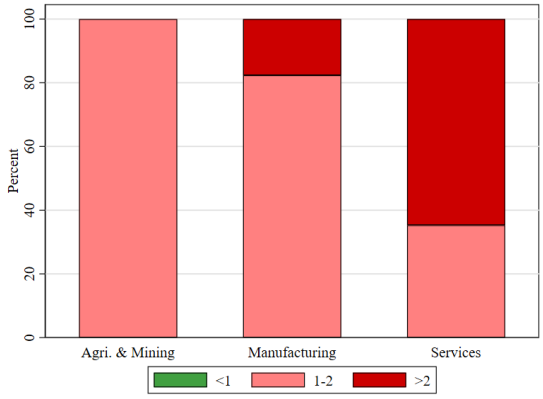

(c) Changes in fixed costs between 1998 and 2005

*Notes:* Figures 3(a)-3(b) shows the distribution of excess entry and formalization costs faced by female entrepreneurs, i.e., $E_{fI}/E_{mI}$ and $E_{fR}/E_{mR}$ across Indian states in 2005. Figure 3(c) plots the 25-75th percentile with the median reported inside the box. Blue shows the changes for male-owned firms, while red shows the changes for female-owned firms.

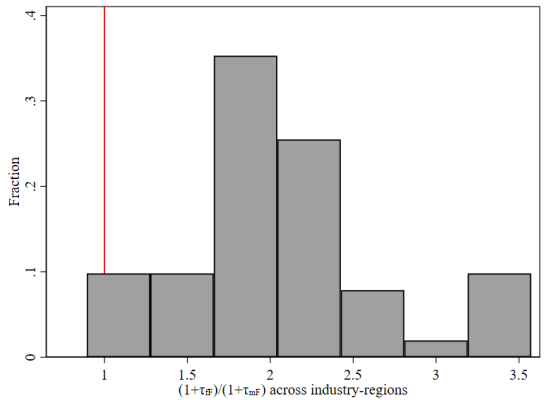
Figure 4: Hiring barriers in the formal and informal sectors



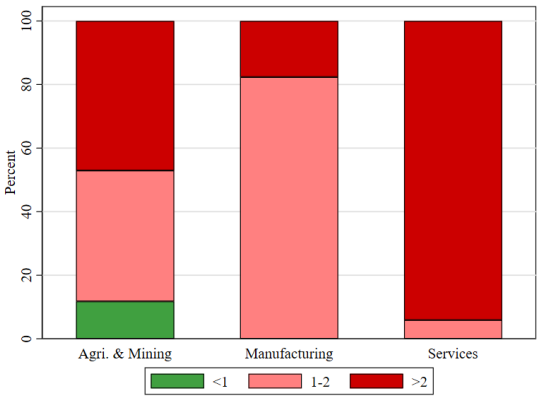

(a) $1 + \tau_{fI}$




(b) $(1 + \tau_{fI})$ across industries



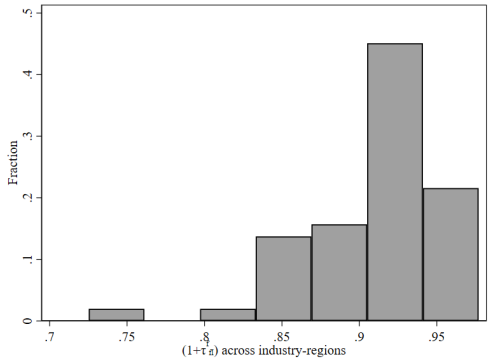

(c) $1 + \tau_{fF}$



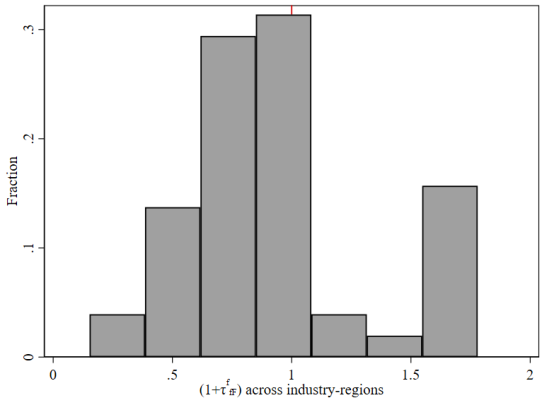

(d) $(1 + \tau_{fF})$ across industries

*Notes:* Figures 4(a) and 4(c) show the distribution of hiring barriers faced by women in the informal and formal sector relative to men in 2005. Figures 4(b) and 4(d) show the breakdown within industry across states. Green corresponds to a ratio less than 1 (favorable for females), pink to a ratio between 1-2, and red to a ratio>2.

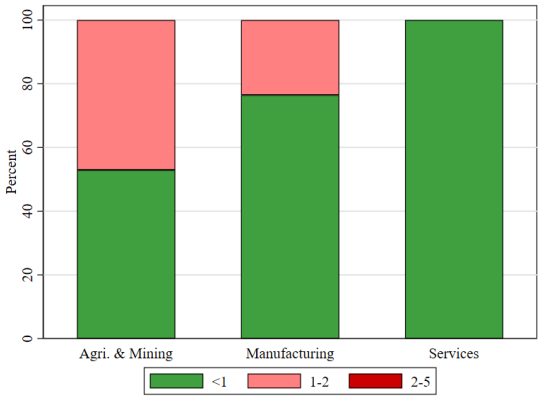
Figure 5: Hiring barriers of hiring female (relative to male) workers



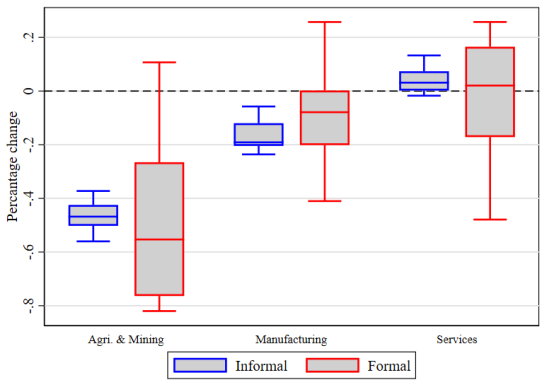

(a) $1 + \tau_{fI}^{f}$



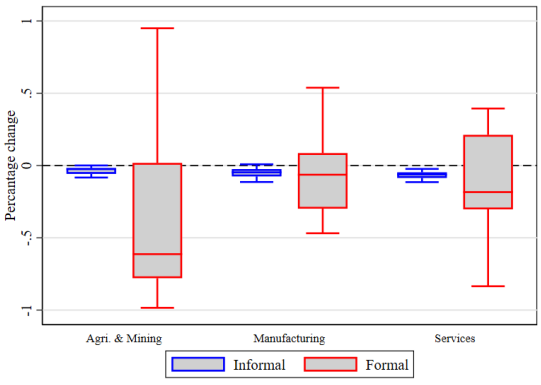

(b) $1 + \tau_{fF}^{f}$




(c) $(1 + \tau_{fF}^{f})$ across industries



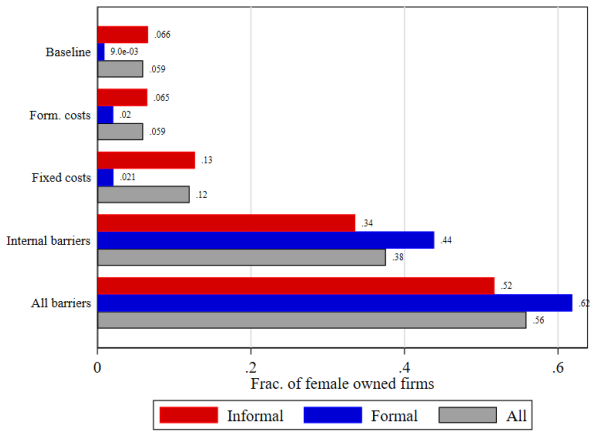

(d) Change in $(1 + \tau_{fs})$



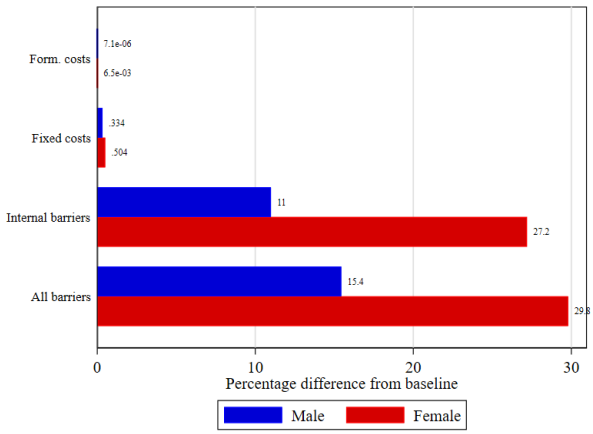

(e) Change in $(1 + \tau_{fs}^{f})$

*Notes:* Figures 5(a) and 5(b) show the distribution of hiring barriers faced by women-owned firms in hiring women workers in the informal and formal sectors respectively in 2005. Figure 5(c) shows the breakdown within industries across states and figures 5(d)-5(e) show box plots of the changes over time with the 25-75th percentile shown in the box, and the median reported inside the box.

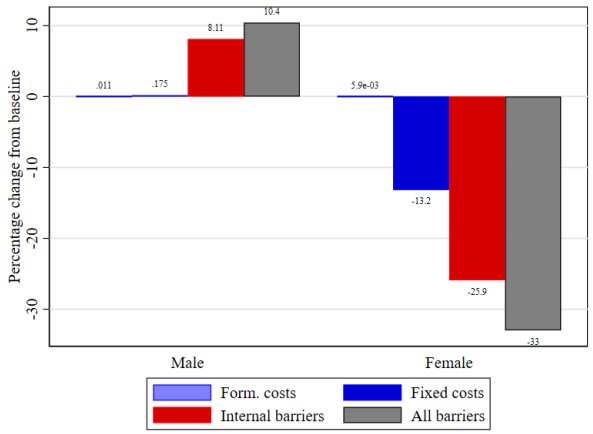
Figure 6: Impact of affirmative action policies on female entrepreneurship and prices



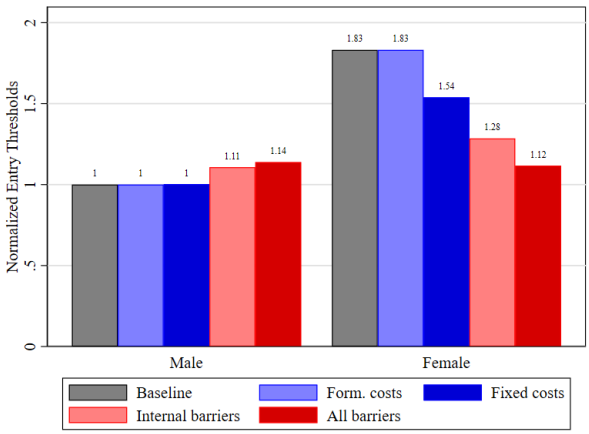

(a) Fraction of female firms




(b) Real wages



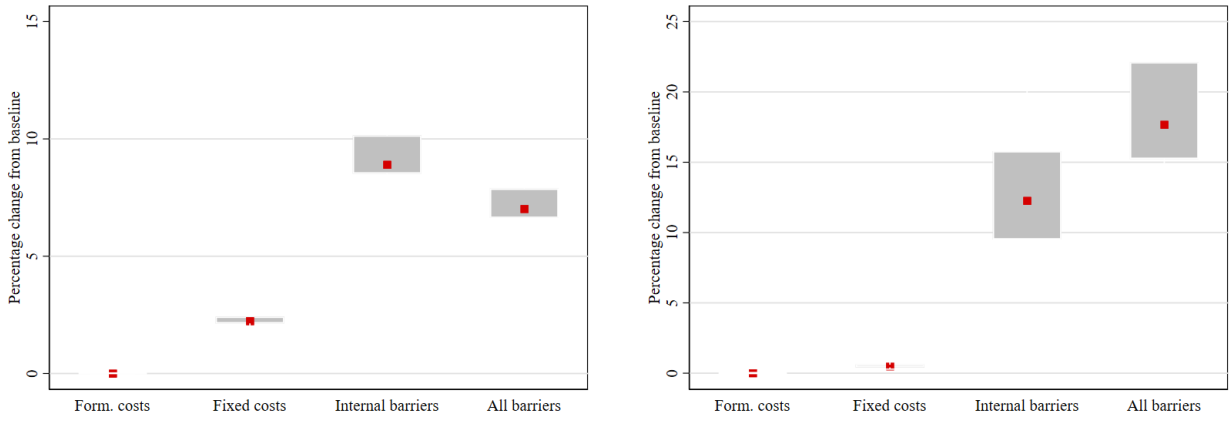

(c) Average productivity changes




(d) Normalized entry thresholds

*Notes:* The above figures report the median impact of four affirmative action policies on male-owned and female-owned firms. Figure 6(a) plots the number of female firms as a fraction of informal, formal, and total firms. Figure 6(c) reports the percentage change in the average productivity of a firm of gender $g$ for each of the counterfactuals as compared to the baseline. Figure 6(b) shows the percentage change in real wages for male and female workers as compared to the baseline.

Figure 7: Impact of affirmative action policies on productivity and welfare




(a) Aggregate productivity changes

(b) Aggregate welfare changes

*Notes:* Figures 7(a) and 7(b) report the percentage changes in aggregate productivity and welfare (measured by real income) for each counterfactual policy as compared to the baseline. We report the median change along with the 25-75th percentile range in the shaded area around the median.

# A Appendix Tables and Figures

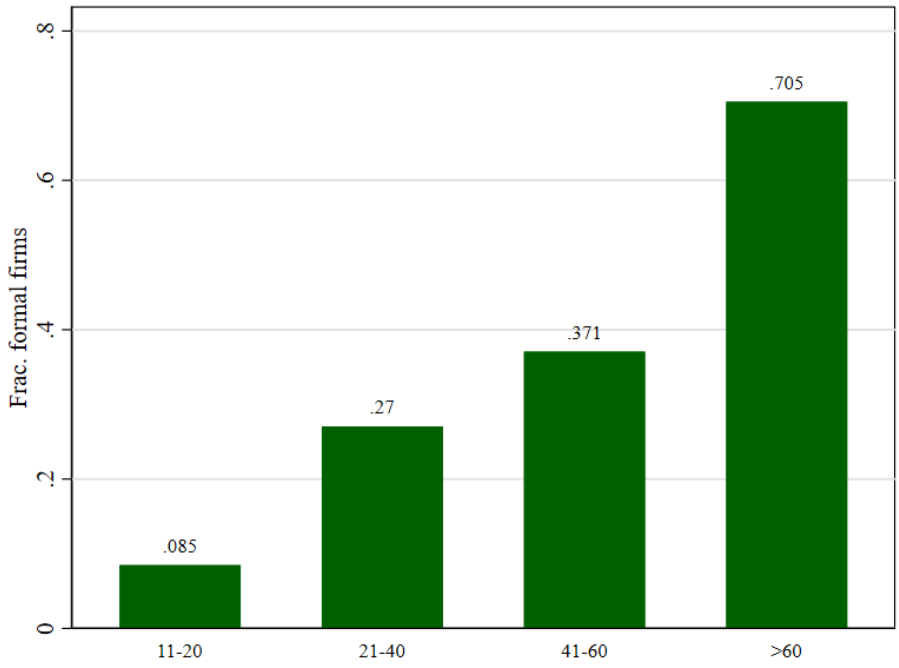
Table A1: Total firm size and composition across gender and sectors

| | Log(L) | | Frac. female emp. | |
|---|---|---|---|---|
| | 1998 | 2005 | 1998 | 2005 |
| | (1) | (2) | (3) | (4) |
| *Panel A: Without industry fixed effects* | | | | |
| Female | -0.0162 | -0.0604*** | 0.298*** | 0.587*** |
| | (0.0176) | (0.00856) | (0.0138) | (0.0145) |
| Formal | 2.448*** | 2.415*** | 0.0647*** | 0.149*** |
| | (0.0328) | (0.0308) | (0.00941) | (0.00972) |
| Female × Formal | 0.234* | 0.195*** | -0.122*** | -0.262*** |
| | (0.141) | (0.0328) | (0.0401) | (0.0277) |
| $R^2$ | 0.210 | 0.18 | 0.34 | 0.457 |
| *Panel B: With industry fixed effects* | | | | |
| Female | -0.0123 | -0.113*** | 0.233*** | 0.524*** |
| | (0.0135) | (0.00749) | (0.00956) | (0.0112) |
| Formal | 2.132*** | 2.221*** | 0.0428*** | 0.108*** |
| | (0.0340) | (0.0355) | (0.00818) | (0.00918) |
| Female × Formal | 0.329** | 0.279*** | -0.0920*** | -0.207*** |
| | (0.166) | (0.0349) | (0.0282) | (0.0271) |
| $R^2$ | 0.338 | 0.256 | 0.472 | 0.509 |
| $N$ | 12.48m | 38.75m | 12.48m | 38.75m |
| Male, Informal | 1.007 | 1.02 | 0.19 | 0.1 |
| Firm controls | Yes | Yes | Yes | Yes |
| District FE | Yes | Yes | Yes | Yes |

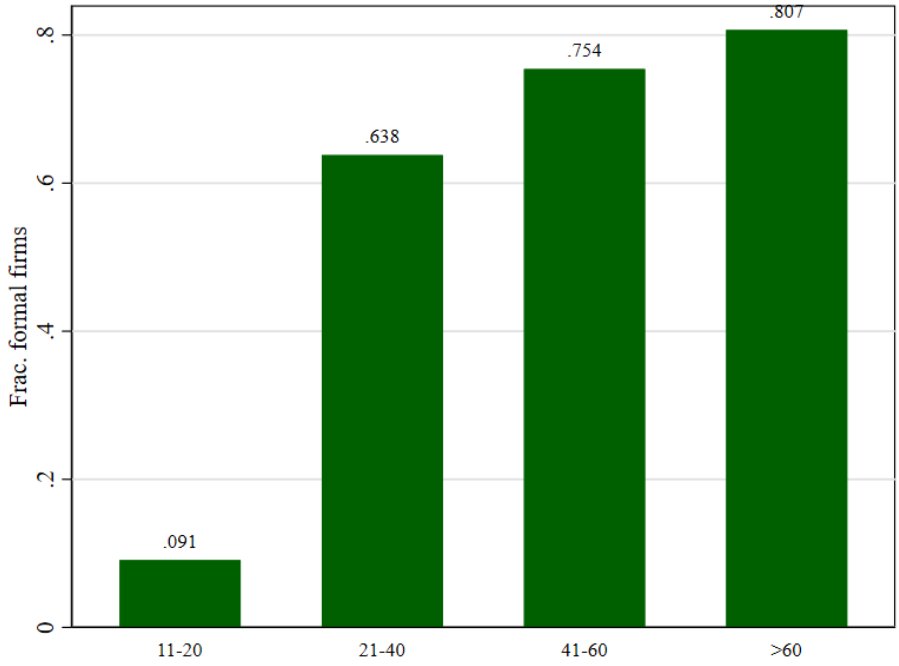
*Notes:* Female and Formal are dummy variables that take the value 1 if the firm is female-owned or if it is in the formal sector and 0 otherwise. Firm controls used are: whether the firm has access to power; dummy variables for different forms of financial access; whether the firm is primarily agriculture-based; and whether the firm is in the rural or urban area. Industry fixed effects are at the four-digit level using the NIC98 for 1998 and NIC04 for 2005. Standard errors are clustered at the district level. * p< 0.1, ** p<0.05 and *** p< 0.01 level of significance.

Figure A1: Fraction of formal firms and firm size

(a) 1998

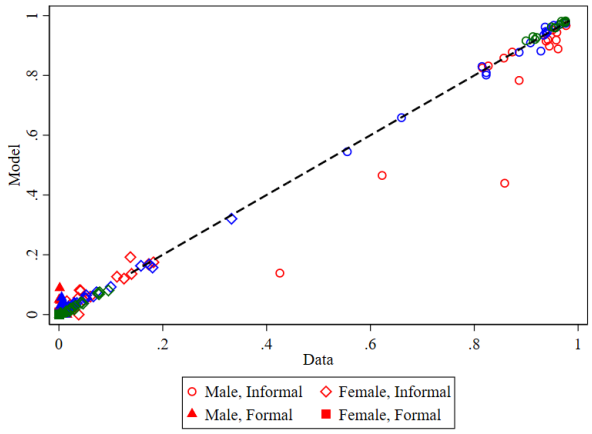



(b) 2005

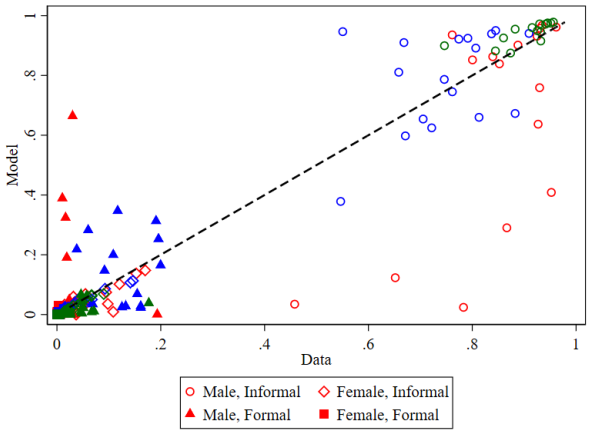



*Notes:* The above graphs plot the fraction of formal firms within each bin of firm size. Firm size is categorized into 11-20 workers, 21-40 workers, 41-60 workers, and more than 60 workers. Graph (a) plots the distribution in 1998, while plot (b) plots the distribution in 2005.

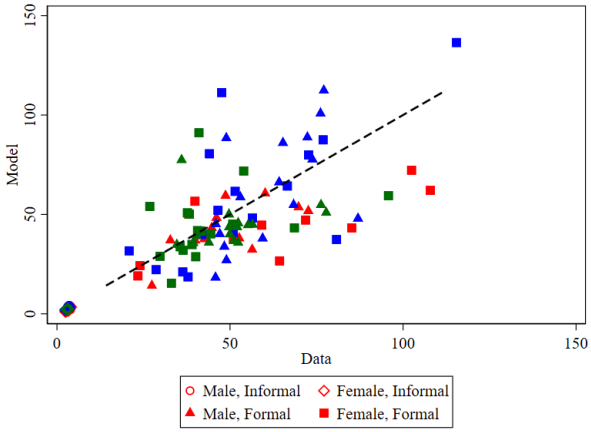
Figure A2: Model fit across firm types, industries and regions




(a) Fraction of firms




(b) Fraction of workers




(c) Average firm size

*Notes:* Figures A2(a), A2(b) and A2(c) show the fit of the model with the data for the fraction of firms, fraction of workers, and average firm size for the four types of firms, namely: male and female firms in the formal and informal sectors. The color shows the industry, where red is agriculture, blue is manufacturing and green is services. Each point is therefore the data/model pair for a firm type in an industry and region. The dotted line is the 45-degree line of equality. The horizontal axis shows the data, while the vertical axis shows the model.

# B Firm ownership in the Enterprise Surveys

We use the World Bank's Enterprise Surveys World Bank (2020), henceforth WBES, to compare the characteristics of male-owned and female-owned firms. As discussed earlier, the WBES are detailed firm-level surveys of a representative sample of the economy's private sector. In particular, they have detailed questions with respect to the output and revenue of firms, along with the composition of their employment, interactions with the government (inspections, bribes, taxes, etc.) and lastly, the subjective evaluation of the respondent on the obstacles faced by the firm. We use a sample of around 140k firms across 141 countries and 13 years (2006-2018). We report the difference across male-owned and female-owned firms. We define a firm as female-owned if a majority of its owners are women.

To begin, in table B1, we report the characteristics of these firms and examine whether they differ based on the gender of the owner. For each charactersitic $y$, the average value for male and female-owned firms is reported in columns (2) and (3). Column (4) reports the raw difference between these means, while column (5) reports the coefficient $\beta$ from the following regression:

$$y_{ict} = \alpha_c + \alpha_t + \alpha_s + \beta F_i + \gamma X_{ict} + \varepsilon_{ict} \tag{9}$$

where: $y_i$ is the characteristic of interest for a firm $i$ in country $c$ in year $t$ (such as sales, wage bill, etc.); $\alpha_c$, $\alpha_t$, $\alpha_s$ are country, year and sector fixed effects respectively; $F_i$ is a dummy variable for whether the firm is female-owned or not; $X_i$ are a set of firm-level controls (such as firm age). We cluster the standard errors at the country-level. $\beta$, our coefficient of interest is reported in column (5) or table B1. Lastly, for a better interpretation of the values in columns (5) and (6), we report them as a percent of the male-owned firm average (column (2)) in parentheses below.

We now turn to interpreting the results in table B1, where it is evident from column (4) that female-owned firms are smaller than male-owned firms along almost every dimension. First, they have around 5-8% lower sales and wages, 8% less number of establishments and 32% lesser workers. They are also around 30% less likely to take out formal loans, spend 11% more time on dealing with bureaucracies and 33% less likely to have secured a government contract in the last year. As

reported in column (6), these results are robust after controlling unobservable differences across countries, sectors and over time, and the differences remain substantial in magnitude. One striking difference between these two types of firms however is the fraction of female workers is 17.9 pp (over 70%) higher in female-owned firms, while the probability that a firm with a top manager (not the owner) as a female is almost 45 pp (over 700%) more likely to be a female-owned firm.

Table B1: Differences in male-owned and female-owned firms

| | | Male | Female | (3)-(2) | |
|---|---|---|---|---|---|
| | $N$ | Mean/SE | Mean/SE | Raw | Coefficient |
| | (1) | (2) | (3) | (4) | (5) |
| Sales (USD millions) | 101024 | 2.773 | 2.625 | -0.147*** | -0.547*** |
| | | [0.025] | [0.053] | [-5.3%] | [-13.44%] |
| Wage bill (USD millions) | 98847 | 0.268 | 0.246 | -0.022*** | -0.0498*** |
| | | [0.002] | [0.005] | [-8.21%] | [-12.15%] |
| Number of Establishments | 109490 | 1.685 | 1.551 | -0.134*** | -0.135 |
| | | [0.014] | [0.026] | [-7.95%] | [-8.1%] |
| Total workers | 118436 | 55.024 | 37.558 | -17.466*** | -21.74*** |
| | | [1.079] | [1.753] | [-31.74%] | [-22.95%] |
| Frac. Female workers | 118392 | 0.249 | 0.428 | 0.179*** | 0.124*** |
| | | [0.001] | [0.002] | [71.89%] | [55.86%] |
| Frac. With top manager female | 96373 | 0.062 | 0.509 | 0.447*** | 0.455*** |
| | | [0.001] | [0.004] | [720.97%] | [700%] |
| Frac. Formal loans | 106579 | 0.112 | 0.079 | -0.032*** | -0.00784 |
| | | [0.001] | [0.001] | [-28.57%] | [-6.17%] |
| Pct. Time spent on bureaucracy | 110131 | 0.090 | 0.080 | -0.010*** | 0.0142*** |
| | | [0.001] | [0.001] | [-11.11%] | [13.65%] |
| Secured govt. contract in last year? | 101571 | 0.142 | 0.096 | -0.046*** | -0.0397** |
| | | [0.001] | [0.002] | [-32.39%] | [-22.69%] |

*Notes:* The mean for every variable is reported for male-owned firms in column (2) and female-owned firms in column (3). The standard errors are reported in parentheses for both these columns. Column (4) reports the raw difference between columns (3) and (2), while column (5) reports the regression coefficient for the regression. The coefficient as a percentage of the male-owned mean in column (2) are reported in parentheses below. All regressions have country, year and sector fixed effects. Standard errors are clustered at the country-level. * p< 0.1, ** p<0.05 and *** p< 0.01 level of significance.

# C Mathematical Proofs

## C.1 Incumbent firm decisions

### Informal Sector

The problem of a firm with productivity $z$ in the informal sector is given by:

$$\max \; pzl^{\widetilde{\rho}} - w_m l_m - w_f l_f - pf_e$$

where $\widetilde{\rho} = \rho/\theta$. Define:

$$w_I = \left[ w_m^{\frac{-\gamma}{1-\gamma}} + w_f^{\frac{-\gamma}{1-\gamma}} \right]^{\frac{-(1-\gamma)}{\gamma}}$$

Then we can rewrite the maximization problem as a two-step problem where in the first step, the firm chooses labor $l$ to maximize profits: $\max pzl^{\rho} - wl^{\theta} - pf_e$ and then minimizes expenditure on male and female workers, given this choice of $l$. Taking the FOC and solving we get:

$$l_I^* = \left[ \widetilde{\rho} \times \frac{pz}{w} \right]^{\frac{1}{1-\widetilde{\rho}}}$$
$$\pi_I^* = \left( 1 - \widetilde{\rho} \right) \times pzl_I^{*\widetilde{\rho}} - pf_I$$

Cost minimization in the second stage implies:

$$\min \; w_m l_m + w_f l_f$$
$$\text{s.t.} \; \left[ l_m^{\gamma} + l_f^{\gamma} \right]^{\frac{1}{\gamma}} = l_I^*$$

Taking the first order conditions and rearranging, we get:

$$
\begin{aligned}
l_g^{1-\gamma} &= \frac{\gamma}{w_g} \ , \forall g \in \{m, f\} \\
\Rightarrow l^{\gamma} &= \gamma^{\frac{\gamma}{\gamma-1}} \times \underbrace{\left[ w_m^{-\frac{\gamma}{\gamma-1}} + w_f^{-\frac{\gamma}{\gamma-1}} \right]}_{=w^{-\frac{\gamma}{\gamma-1}}} \\
\Rightarrow \gamma &= \frac{w}{l^{1-\gamma}} \\
\Rightarrow l_g &= \left( \frac{w_g}{w} \right)^{-\frac{1}{1-\gamma}} \times l(z)
\end{aligned}
$$

**Formal sector**

The problem in the formal sector is the same as that in the informal sector, except that a firm earns $(1-t)p$ per unit of output sold and there is no size-based monitoring, implying that $\theta = 1$. Putting this together, we have that the optimal firm-size and profits of a firm with productivity $z$ in the formal sector will be given by:

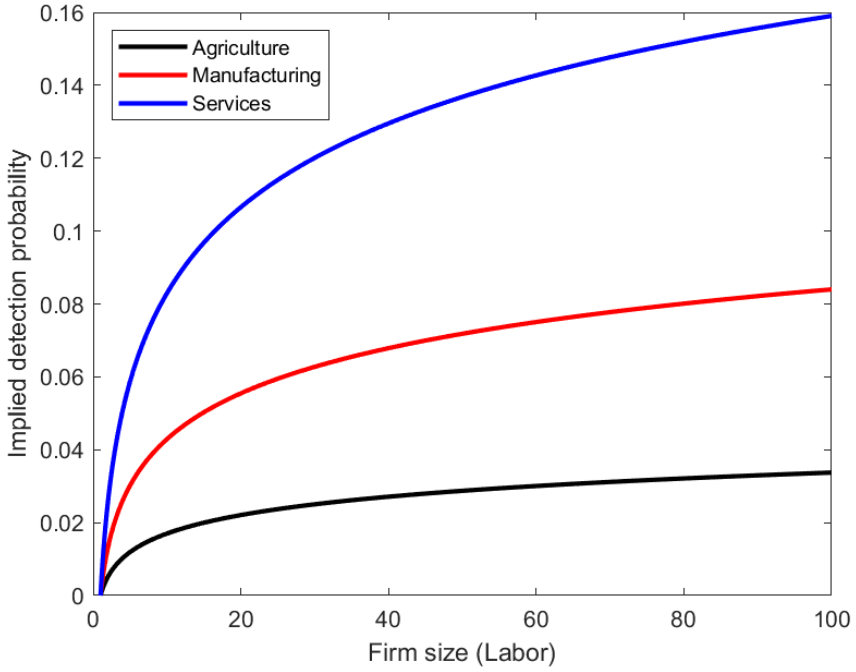
$$
\begin{aligned}
l_F^* &= \left[ \rho \times \frac{(1-t)pz}{w_F} \right]^{\frac{1}{1-\rho}} \\
\pi_F^* &= \left( 1 - \rho \right) \times pz l_F^{*\rho} - p f_F
\end{aligned}
$$

## C.2 Cost function and probability of detection

An alternative way to present the model is to allow for a size-dependent probability of detection by the government in the informal sector. This implies that larger firms operating in the informal sector face a penalty for evading taxes. Let $\tau(l)$ be the probability of detection such that $\tau(0) = 0$, $\tau'(l) > 0$ and $\lim_{l \to \infty} \tau(l) = 1$. The firm's maximization problem can be written as:

$$
\max_l (1 - \tau(l)) \times pz l^{\rho} - wl
$$

Figure C1: Probability detection function




*Notes:* The above graph plots the size-based penalty of operating in the informal sector, as implied by the model calibration in the paper.

Our baseline profit maximization condition for the informal sector is given by:

$$\max_{l} l^{-\rho(1-\frac{1}{\theta})} \times pzl^{\rho} - wl$$

This implies that: $\tau(l) = 1 - l^{-\rho(1-\frac{1}{\theta})}$. A larger $\theta$ therefore implies a higher probability of being detected conditional on the same firm-size ($l$). From the calibration exercise in the paper, we have $\rho = 0.738$ and $\theta = \{1.01, 1.03, 1.05\}$ in 2005. This would imply a size-based penalty function shown in Figure C1.

# D Identification of Model Parameters

Table D1 provides the key relationships used to identify the relevant parameters.

**Distortions for hiring female workers ($\tau_{gs}^f$):** As evident from the table, $\tau_{gs}^f$ can be identified by comparing the ratios of female-to-male workers hired across different types of firms. Note that we normalize $\tau_{mI}^f = 0$. Therefore, $R_{mF}/R_{mI} = (1+\tau_{mF}^f)^{-1/(1-\gamma)}$. Similarly, $R_{fI}/R_{mI} = (1+\tau_{fI}^f)^{-1/(1-\gamma)}$ and lastly, $R_{fF}/R_{mF} = (1+\tau_{fF}^f)^{-1/(1-\gamma)}$.

**Productivity distribution parameters:** For each gender $g$, and a given value of: (i) $\{\sigma_x^2, \sigma_\varepsilon^2\}$; (ii) the mass of potential entrants $M_g$; (iii) the number of firms in the informal and formal sectors ($N_{gI}$ and $N_{gF}$ respectively), we can identify the (gender-specific) productivity thresholds $x_{gI}^*$ and $x_{gF}^*$ of entering each sector, which are given by: $x_{gF}^* = F^{-1}(1 - N_{gF}/M_g)$ and $x_{gI}^* = F^{-1}(1 - (N_{gI}+N_{gF})/M_g)$. From table D1, the variance of log-labor in each sector can be given by:

$$\begin{aligned}\text{Var}(\ln l_I) &= \frac{\sigma_{xI}^2}{(\theta-\rho)^2} \\ \text{Var}(\ln l_F) &= \frac{\sigma_{xF}^2 + \sigma_\varepsilon^2}{(1-\rho)^2}\end{aligned}$$

where $\sigma_{xI}^2 = \text{Var}(x|x \in (x_I^*, x_F^*))$ and $\sigma_{xF}^2 = \text{Var}(x|x \geq x_F^*)$. The observable data moments on the variance of log-firm size, can be used to identify $\{\sigma_x^2, \sigma_\varepsilon^2\}$ for each gender separately.

**Informal sector penalty ($\theta$):** From table D1, note that the average firm size in each sector (for each firm owned by gender $g$) is given by:

$$\begin{aligned}\overline{\ln l_{gI}} &= \frac{1}{1-\widetilde{\rho}} \ln\left(\widetilde{\rho} \times \frac{p}{w_{gI}}\right) + \frac{\overline{\ln x_{gI}}}{1-\widetilde{\rho}} \\ \overline{\ln l_{gF}} &= \frac{1}{1-\rho} \ln\left(\rho(1-t) \times \frac{p}{w_{gF}}\right) + \frac{\overline{\ln x_{gF}}}{1-\rho}\end{aligned}$$

where: $\overline{\ln x_{gI}} = E(\ln x | x \in \{x_{gI^*}, x_{gF^*}\})$ and $\overline{\ln x_{gF}} = E(\ln x | x \geq x_F^*)$ is the average firm productivity on entry into the formal and informal sectors respectively. Therefore, note that the difference in the log-firm size for male-owned firms in the formal and informal sectors can be used to identify $\widetilde{\rho}$, and hence $\theta$, as follows:

$$(1-\rho)\overline{\ln l_{mF}} - (1-\widetilde{\rho})\overline{\ln l_{mI}} = \ln(\rho(1-t)) - \ln\widetilde{\rho} + \frac{1-\gamma}{\gamma} \times \ln\left(\frac{1+R_{mF}^{\gamma}}{1+R_{mI}^{\gamma}}\right) + \overline{\ln x_{mF}} - \overline{\ln x_{mI}}$$

**Hiring distortions ($\tau_{fs}$):** Since we normalize the hiring distortions faced by male-owned firms ($\tau_{ms}$) to be equal to zero in both the informal and formal sector, we can compare the average log-firm size for male- and female-owned firms to identify hiring distortions for the female-owned firms *relative* to male-owned firms ($\tau_{fs}$) as follows:

$$(1-\widetilde{\rho})(\overline{\ln l_{mI}} - \overline{\ln l_{fI}}) = \ln(1+\tau_{fI}) + \frac{1-\gamma}{\gamma} \times \ln\left(\frac{1+R_{mI}^{\gamma}}{1+R_{fI}^{\gamma}}\right) + \overline{\ln x_{mI}} - \overline{\ln x_{fI}}$$

$$(1-\rho)(\overline{\ln l_{mF}} - \overline{\ln l_{fF}}) = \ln(1+\tau_{fF}) + \frac{1-\gamma}{\gamma} \times \ln\left(\frac{1+R_{mF}^{\gamma}}{1+R_{fF}^{\gamma}}\right) + \overline{\ln x_{mF}} - \overline{\ln x_{fF}}$$

**Fixed costs:** Lastly, to calculate the fixed costs, we need to know the labor used by the marginal firm in each sector. Let $l_{gs}^*$ be the labor used by the firm with threshold productivity $x_{gs}^*$. Note from Table D1 that:

$$l_{gI}^* = \left(\frac{z_{gI}^*}{\overline{z_{gI}}}\right)^{\frac{1}{1-\widetilde{\rho}}} \times \overline{l_{gI}}$$

$$l_{gF}^* = \left(\frac{z_{gF}^*}{\overline{z_{gF}}}\right)^{\frac{1}{1-\rho}} \times \overline{l_{gF}}$$

We can then use the zero-profit conditions in (7) to calculate the fixed costs of entry and registration.

Table D1: Relationships used for parameter identification

| | Male owned firms | Female owned firms |
|---|---|---|
| Informal sector | $l_{mI} = \left[\widetilde{\rho}\frac{p}{w_{mI}} \times z\right]^{\frac{1}{1-\widetilde{\rho}}}$ | $l_{fI} = \left[\widetilde{\rho}\frac{p}{w_{fI}} \times z\right]^{\frac{1}{1-\widetilde{\rho}}}$ |
| | $R_{mI} \equiv \frac{l^f_{mI}}{l^m_{mI}} = \left(\frac{\tilde{w}^f}{\tilde{w}^m}\right)^{\frac{-1}{1-\gamma}}$ | $R_{fI} \equiv \frac{l^f_{fI}}{l^m_{fI}} = \left(\frac{(1+\tau^f_{fI})\tilde{w}^f}{\tilde{w}^m}\right)^{\frac{-1}{1-\gamma}}$ |
| | $w_{mI} = \tilde{w}^m\left[1 + (R_{mI})^\gamma\right]^{-\frac{1-\gamma}{\gamma}}$ | $w_{fI} = (1+\tau_{fI})\tilde{w}^m\left[1 + (R_{fI})^\gamma\right]^{-\frac{1-\gamma}{\gamma}}$ |
| Formal sector | $l_{mF} = \left[\rho\frac{(1-t)p}{w_{mF}} \times z\right]^{\frac{1}{1-\rho}}$ | $l_{fF} = \left[\rho\frac{(1-t)p}{w_{fF}} \times z\right]^{\frac{1}{1-\rho}}$ |
| | $R_{mF} \equiv \frac{l^f_{mF}}{l^m_{mF}} = \left(\frac{(1+\tau^f_{mF})\tilde{w}^f}{\tilde{w}^m}\right)^{\frac{-1}{1-\gamma}}$ | $R_{fF} \equiv \frac{l^f_{fF}}{l^m_{fF}} = \left(\frac{(1+\tau^f_{mF})(1+\tau^f_{fF})\tilde{w}^f}{\tilde{w}^m}\right)^{\frac{-1}{1-\gamma}}$ |
| | $w_{mF} = \tilde{w}^m\left[1 + (R_{mF})^\gamma\right]^{-\frac{1-\gamma}{\gamma}}$ | $w_{fF} = (1+\tau_{fF})\tilde{w}^m\left[1 + (R_{fF})^\gamma\right]^{-\frac{1-\gamma}{\gamma}}$ |