Aggregate Implications Of Barriers To Female Entrepreneurship*

Gaurav Chiplunkar
University of Virginia

Pinelopi Koujianou Goldberg
Yale University, BREAD, CEPR and NBER

December 2023

Abstract

We develop a framework for quantifying barriers to labor force participation (LFP) and entrepreneurship faced by women in developing countries, and apply it to the Indian economy. We find that women face substantial barriers to LFP. The costs for expanding businesses through the hiring of workers are also substantially higher for women entrepreneurs. However, there is one area in which female entrepreneurs have an advantage: the hiring of female workers. We show that this is not driven by the sectoral composition of female employment. Consistent with this pattern, we find even without explicitly targeting female LFP, policies that boost female entrepreneurship can significantly increase female LFP. Counterfactual simulations indicate that removing all excess barriers faced by women entrepreneurs would substantially increase the fraction of female-owned firms, female LFP, earnings, and generate substantial gains for the economy. These gains are due to higher LFP, higher real wages and profits, and 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

*This is a substantially revised version of a paper circulated in February 2021. We would like to thank Oriana Bandiera, Latika Chaudhary, Rafael Dix-Carneiro, Erik Hurst, Rachel Ngai, Rohini Pande, Michael Peters, Imran Rasul, Gabriel Ulyssea, and seminar participants at the University of Virginia, MIT, Harvard University, University of Alberta, Ohio University, the IGC-Stanford Conference on Firms, Trade and Development, and the Sustaining Economic Growth: Fall 2022 Conference Of The Working Group on the Foundations of Long-run Prosperity at Stanford, 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, 87 Trumbull St., 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 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

[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. Bento (2020) and Morazzoni and Sy (2022) focus on entrepreneursip in the U.S., and estimate large productivity and welfare gains associated with the elimination of gender-specific distortions.

[3] See Jayachandran (2021), Quinn and Woodruff (2019), and Cuberes and Teignier (2014) for reviews.

low (Fletcher, Pande and Moore, 2019; Deshpande and Kabeer, 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 stylized model of occupational choice along the lines of Roy (1951) and Banerjee and Newman (1993) that captures some important features of developing economies. The model features an economy with multiple industries and a mass of individuals (men and women), who decide whether to participate in the labor force, and conditional on participation, whether to be self-employed, earn wages as workers, or start a business as entrepreneurs. 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.[4] Firms (entrepreneurs) need to pay an entry cost to operate in either sector and an additional registration cost to formalize.[5] Firms in the informal sector avoid paying the registration cost as well as taxes, 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.[6] Given that entry and regulation costs are lower in the informal than in the formal sector, the presence of a large informal sector could, in principle, benefit women who want to enter entrepreneurship. Indeed, while women are under-represented among entrepreneurs, they are over-represented in the informal sector in developing economies (World Bank, 2012). There is only one input in production: labor. Entrepreneurs choose the sector (i.e., formal versus informal) and industry in which they operate. Conditional on these choices, they 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 and to have different productivities. Second, we allow

[4]See LaPorta and Shleifer (2014); Ulyssea (2018, 2020); Dix-Carneiro, Goldberg, Meghir and Ulyssea (2021); Rao, Verschoor, Deshpande and Dubey (2008).

[5]The importance of these fixed entry and registration costs has been emphasized across many contexts. See comprehensive reviews by Jayachandran (2021) and Quinn and Woodruff (2019).

[6]See Beck and Hoseini (2014), Nikaido, Pais and Sarma (2015), Chaudhuri, Sasidharan and Raj (2020), Raj and Sasidharan (2020) and Morazzoni and Sy (2022).

for men and women to face different costs of participating in the labor force. Third, we allow men and women entrepreneurs to face different costs to start their business, and formalize/register it with the government. 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. We then use the structure of the model, in conjunction with the rich data of the Census to estimate these frictions, and examine their implications for various aggregate outcomes (such as labor force participation, wages, productivity, income, etc.). 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).[7] While we do not measure these factors directly (but model them as "wedges"),[8] our estimated frictions are correlated with various indices of women empowerment across regions in India, constructed on the basis of comprehensive indicators (such as inputs in household decisions, patriarchal norms, asset ownership patterns, access to education and health, etc.). This increases our confidence that our estimates are meaningful in capturing various underlying barriers and frictions that women face in the labor force.

We have three key findings. First, the excess costs faced by women are substantial. Labor force participation costs are roughly twice as large for women than for men on average despite a significant decline over time. Similarly, women entrepreneurs face a 10-20% higher cost of expanding their business through hiring (both in the informal and formal sectors), compared to their male counterparts. Second, the only area where female entrepreneurs seem to have a significant advantage over their male counterparts is in the employment of female workers (particularly in the informal sector). We show that this advantage is not driven by sectoral effects (i.e., it holds even within narrowly defined industries at the 4-digit National Industry Classification level), or by family-owned firms (i.e., it holds even when family-owned businesses or non-hired workers are excluded from the analysis). This comparative advantage of female entrepreneurs in employing females is especially important in a context like India, where female labor force participation is low and women workers are scarce. Third, *conditional on female labor force participation*, constraints on the intensive margin (i.e., expanding the business) are more severe than constraints on the extensive margin (i.e., fixed costs of entry into entrepreneurship). In fact, we find no evidence that (conditional on LFP), females face higher fixed costs of entry into informal entrepreneurship; however, we do find that they face significantly higher fixed costs of formalization. These patterns

[7] For comprehensive surveys of this literature, see Altonji and Blank (1999), Bertrand (2011), Blau, Ferber and Winkler (2014).

[8] 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), Deshpande and Kabeer (2019) and Ashraf, Bau, Nunn and Voena (2020) among others.

confirm the prior that the presence of a large informal sector in developing countries could be advantageous to female entrepreneurs.

Given these results, we investigate in a series of counterfactual scenarios the potential gains to the economy of eliminating gender-related barriers. Specifically, we first examine the impact of affirmative action policies that 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 cases where women have an advantage over men (i.e., in attracting female workers), we do not eliminate this advantage. Next, we repeat this exercise by imposing complete gender equality, i.e., also eliminating the advantage that women have in hiring other women.

The counterfactual simulations lead to several policy-relevant insights. First, conditional on labor force participation, policies that target the intensive margin (hiring barriers) have substantially larger effects than policies that focus on the extensive margin (i.e., fixed costs) of entrepreneurship. Intuitively, eliminating entry and registration excess costs has little effect when barriers to operating and growing a business remain in place. Second, policies promoting female entrepreneurship can have large effects on female LFP, even when LFP is not directly targeted by policy makers. This is not only because more women become entrepreneurs, but also because female entrepreneurs tend to hire more female workers. Third, it is important to target distortions not only on the labor supply, but also on the demand side. Specifically, eliminating frictions to female labor force participation has – as expected – large effects on women's labor force participation. However, without any additional measures to boost demand for female workers, this increase implies a large decline in the real wages of women. In contrast, policies that target both labor supply and demand frictions boost female LFP while increasing profits of women entrepreneurs and only marginally decreasing real wages of women. Fourth, 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. This more efficient reallocation results in substantial gains in aggregate productivity and welfare (as measured by real income). Removing all types of barriers facing women while preserving their comparative advantage in the employment of female workers boosts labor force participation in the economy with female LFP doubling, and raises aggregate productivity by 3% and welfare (real income) by 43%. These gains are large and suggest that promoting gender equality in entrepreneurship can contribute meaningfully to economic development.

Importantly, all these effects are also present in the case where all gender-related frictions are eliminated, so that women no longer have an advantage in the employment of females. However, they are muted, as expected. This is because the positive effect of female entrepreneurs on female employment is no longer present. Nevertheless, the effects remain large and positive, suggesting that even without affirmative action (i.e., policies that give women, or let them retain, an advantage in some areas), policies targeted at achieving gender-parity can generate substantial benefits for the economy.

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 three studies that are closest in spirit to our work are the U.S.- focused papers by Hsieh, Hurst, Jones and Klenow (2019) and Bento (2020), and the cross-country analysis of Cuberes and Teignier (2016) and Ranasinghe (2021). 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 and reduced-form 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 considers a counterfactual where all gender-related frictions are removed (i.e., there is no affirmative action). Section 8 concludes.

2 Theory

2.1 Conceptual Issues

Before presenting the model, we provide a brief overview of our treatment of gender in the paper in order to make explicit which primitives in the model are assumed to be constant across genders and which are allowed to be different. A framework in which men and women are allowed to differ in every possible dimension will attribute all differences in economic outcomes to differences in primitives by construction.

We assume that men and women are the same in the following respects: First, they have the same preferences for work. Accordingly, we interpret differences in the estimated “disutility” of work as capturing distortions preventing women from participating in the labor force rather than innate stronger dislike for work. Part of the reason for doing so is that we have no way of identifying preferences versus distortions in our setting. Fundamentally, it might be impossible to separate the two as preferences are themselves endogenous: in the presence of distortions, e.g., norms discouraging female LFP, lack of child care, etc., women may find employment outside the home less appealing, even though they are not innately different from men. Second, we assume that male and female entrepreneurs active in the same industry employ the same production technology. This assumption is necessitated by data constraints as well as the desire to keep the framework tractable. While our primary data source does not allow us to directly test this assumption, we use additional data that cover a subset of firms (NSS) to provide corroborative evidence in Appendix E.2. Third, we assume that men and women do not differ in their innate entrepreneurial ability, though their ex-post realization of productivity, which is industry specific, may be different. We discuss this assumption extensively when we introduce it in the model as well as in Appendix F, where we also partially relax it.

On the other hand, we allow men and women to differ in all those dimensions that can be identified based on the information available in our data. Specifically, we allow male and female workers to be imperfect substitutes in the production function. We also allow the productivity of male and female workers to differ, with these differences being sector-specific; this treatment accounts for the brawn vs. brain hypothesis and its implications for comparative advantage. Given these differences, male and female wages are not equalized. However, the wage differences are attributed to the imperfect substitutability of male and female workers and their differential productivity by sector as opposed to wage discrimination. We abstract from wage discrimination because we cannot identify it in our setting. By doing so, we ignore an additional, potentially important source of misallocation, and bias our results towards underestimating gender-related distortions. However, given that the rest of our findings suggest large distortions, this limitation makes our conclusions even stronger. Finally, we allow for men and women to face differential fixed costs of entry in all employment choices as well as differential distortions in expanding their businesses.

2.2 Setup

The economy consists of a mass of N_g individuals of gender g (male or female) and 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.[9] The only difference between firms in the formal and informal sectors is in their compliance with

[9]Using additional survey data, we provide suggestive evidence for these assumptions in Appendix F.2. We find no gender differences in how entrepreneurs report the innovation/quality of their product as well as competition faced by other businesses offering similar products or services.

regulations, and the potential distortions they face in expanding via hiring workers (discussed below).

2.3 Entrepreneurs

Each entrepreneur or firm (we use these terms interchangeably throughout the paper) of gender g in industry j and sector s (we will subsequently drop the j and s indices for notational convenience) is indexed by her/his individual entrepreneurial productivity $z \sim H(z)$. Labor is the only input in production. This assumption implies that any potential distortions affecting other inputs (e.g., capital) will be loaded onto labor in our framework, so that it is more appropriate to interpret distortions in hiring as distortions in expanding the business more generally. Appendix E discusses the implications of this modeling in detail, as well shows (for a subset of firms) that it does not affect our main conclusions.[10] Entrepreneurs hire male and female workers to produce output that is sold in a competitive market. We allow for male and female workers to be imperfect substitutes in production with differential productivity A^g across industry-sectors. For example, we allow for the productivity of a female worker (relative to male) to be different in formal agriculture as compared to informal agriculture, formal manufacturing, etc. A worker of gender $g \in \{m, f\}$ can be hired in a competitive labor market at a wage $\widetilde{w}^g$. The setup is static so that after entry, firms stay active forever.[11]

For notational consistency, we will henceforth use $x_{gsj}^{g'}$ to denote a variable x (e.g., wages, labor, etc.) that refers to an entrepreneur of gender g, in sector s and industry j, and a worker of gender g' (that is, the subscripts in our notation will refer to the gender of entrepreneurs and the superscripts to the gender of workers). Output y of a firm with productivity z is given by:

$$y = zl^{\rho}$$
$$l = \left[\sum_{g} (A^g)^{\frac{1}{\gamma}} (l^g)^{\frac{\gamma-1}{\gamma}} \right]^{\frac{\gamma}{\gamma-1}}$$

where: $0 < \rho < 1$, A^g is the productivity of a gender g worker, and γ is the elasticity of substitution between male and female workers in production.

[10] The Census data we use to estimate the model does not provide information on other inputs, and hence we are not able to estimate a more general model that includes multiple inputs. In Appendix E.2, we consider such a model and estimate it using the 62nd Round of the National Sample Survey data (2005). The NSS provides information only on a subset of firms, namely those operating in informal manufacturing. The results obtained using the NSS and a model with multiple inputs are very similar to the ones obtained in our baseline model for the corresponding sector, i.e., informal manufacturing. For a detailed discussion see Appendix E.2.

[11] As reported by Hsieh and Klenow (2009), most firms in India are small, never grow, and never die.

The distinction between firms in the formal and informal sectors is that firms in the formal sector have to pay a per-unit sales tax t, while firms in the informal sector do not pay any taxes, but face a size-dependant penalty of being informal.[12]

Entrepreneurs in the Informal Sector: The profit maximization problem of a firm in the informal sector of industry j (dropped for notational convenience), owned by an entrepreneur of gender g with productivity z, is given by:

$$\max_{\{l^m, l^f\}} \pi_{gI}(z) = pzl_{gI}^{\rho_I} - w_{gI}^m l_{gI}^m - w_{gI}^f l_{gI}^f$$

where: $\rho_I = \lambda\rho < \rho$ captures a size-based penalty faced by firms operating in the informal sector.[13] This penalty implies that it is less desirable for larger firms to remain informal, which is plausible in the Indian context given that informal firms are more constrained in their access to formal channels of finance (Beck and Hoseini, 2014), and that large informal firms have a higher probability of being detected and penalized for failing to register their business.[14]

The terms w_{gI}^m and w_{gI}^f denote the *effective* wages facing entrepreneurs in the informal sector. Entrepreneurs differentiated by gender may face frictions in growing their businesses. 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 τ_{gI} for hiring a worker in the informal sector, and a further cost τ_{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.[15] 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.

[12]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.

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

[14]We later show in Appendix Section B.2 that this size-based penalty can be re-written as a per-unit tax of operating in the informal sector. 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 λ 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.

[15]In theory, instead of assuming potential wedges on inputs (labor), we could put them on output. However, they would be isomorphic in our context. Given that we have data on firm-size, but not output, we put these potential barriers on labor.

While such "cultural" factors and norms are considered important for employment decisions, they are difficult, if not impossible, to credibly quantify based on existing data. Accordingly, we do not attempt to measure them in this paper, but model them using the shorthand described above as distortions that increase the effective cost of labor. 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 therefore given by $\boldsymbol{w_{gI}} \equiv \{w^m_{gI}, w^f_{gI}\} = (1+\tau_{gI})\{\widetilde{w}^m, (1+\tau^f_{gI})\widetilde{w}^f\}$. The first order conditions imply that demand for male and female workers, optimal firm size, and profits (dropping j for notational convenience) are given by:

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

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

$$\pi_{gI}(z) = \frac{1-\rho_I}{\rho_I} \times w_{gI} l_{gI}(z) \tag{3}$$

$$\text{where: } w_{gI} = \left[\sum_{g'} A^{g'}(w^{g'}_{gI})^{1-\gamma}\right]^{\frac{1}{1-\gamma}}$$

Mathematical proofs are provided in Appendix B.1.

Entrepreneurs in the Formal Sector: A firm in the formal sector, owned by an entrepreneur g with productivity z, chooses labor to maximize profits given by:

$$\max_{\{l^m, l^f\}} \pi_{gF} = (1-t)pl^{\rho}_{gF} - w^m_{gF}l^m_{gF} - w^f_{gF}l^f_{gF}$$

As with the informal sector, we assume that an entrepreneur g faces hiring frictions, modeled as an additional cost τ_{gF} and τ^f_{gF} of hiring a worker and female worker respectively in the formal sector. Therefore, the *effective* wage is given by $\boldsymbol{w_{gF}} \equiv \{w^m_{gF}, w^f_{gF}\} = (1+\tau_{gF})\{\widetilde{w}^m, (1+\tau^f_{gF})\widetilde{w}^f\}$. The first order conditions imply that demand for workers of gender g', optimal firm size, and profits (dropping j for notational clarity) are given by:

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

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

$$\pi_{gF}(z) = \frac{1-\rho}{\rho} \times w_{gF} l_{gF}(z) \tag{6}$$

$$\text{where: } w_{gF} = \left[\sum_{g'} A^{g'} (w_{gF}^{g'})^{1-\gamma} \right]^{\frac{1}{1-\gamma}}$$

Mathematical proofs are provided in Appendix B.1.

2.4 Labor Supply Decisions

Individuals decide whether and how to participate in the labor force. Conditional on working, an individual can be self-employed (i.e., operate an owner-only firm), work as a worker, or be an entrepreneur.

To model the decision to participate in the labor force, we adopt a structure similar to Bick, Fuchs-Schündeln, Lagakos and Tsujiyama (2022), and assume that an individual consumes a bundle of consumption goods $C = \prod_j C_j^{\alpha_j}$ ($\sum_j \alpha_j = 1$) and has a disutility of working, so that:

$$U(x, \eta) = \max_C C - \mathbf{1}_{LFP} \times \eta \overline{u}_g$$
$$\text{s.t.} \sum_j p_j c_j \leq I(x) + b$$

where: $I(x)$ is the income earned by an individual if (s)he participates in the labor force as self-employed, a worker, or an entrepreneur. The term x denotes entrepreneurial ability and its role in the model will be explained shortly; b are benefits received by all agents in the economy from the government (financed through taxes); and $\eta \overline{u}_g$ are gender-specific utility costs of working (this term subsumes cultural and social norms discouraging women from participating in the labor force). $\eta \sim F_\eta(\eta)$ are idiosyncratic utility costs that vary across individuals, while $\overline{u}_g$ captures average differences across gender. We use the term "disutility" to capture all costs associated with work outside the home, and not just dislike of work (in that sense, the term may be misleading, but we nevertheless use it for convenience as it is standard in the literature). As noted earlier in 2.1, an important premise of this work is that men and women do not differ in their innate dislike of work, so that differences in "disutility" reflect gender barriers to LFP.

Let $P = \prod (p_j/\alpha_j)^{\alpha_j}$ be the price index of the economy, which we normalize to 1. We assume that individuals cannot save, i.e., they consume their entire income. An individual will participate in the labor force as long as the real-income from working is greater than the disutility of participating in the labor force, i.e., $\eta \bar{u}_g < \frac{I(x)}{P}$. This implies that the labor force participation rate for gender g will be given by $F_\eta(\eta^*)$, where $\eta^* = \frac{I(x)}{P\bar{u}_g}$ is – according to the LFP indifference condition – the threshold disutility of working for an individual who is indifferent between working or not. All individuals with $\eta < \eta^*$ will participate in the labor force, while those with $\eta > \eta^*$ will not.

Entrepreneurship, Self-, and Wage-Employment: Conditional on participating in the labor force, individuals choose between being entrepreneurs, self-employed, or wage workers. Individuals draw an entrepreneurial ability x from an ability distribution $x \sim G(x)$. We assume that $G(x)$ is continuous with support $(0, \infty)$, has finite moments, and is identical and independently distributed for all individuals within an industry, but can vary across industries. An entrepreneur of gender g and ability x earns an expected profit denoted by $E(\Pi_{gs}(x))$ in sector s.

Note that this specification does not allow for innate entrepreneurial ability to vary across genders (though, as will become clear shortly, we do allow the realization of entrepreneurial productivity to vary across genders in an industry-sector specific way). A potential caveat is that even if men and women may not have innate differences in their suitability to entrepreneurship, they may differ in other characteristics, e.g., education, that affect their entrepreneurial performance. We examine this possibility in Appendix F. A comparison of educational attainment of men versus women in India does not allow for clear conclusions. While men have more years of schooling and higher literacy rates on average than women, adjusting for their learning gives a different picture: quality-adjusted years of schooling tend to be higher for women (consistent with Angrist et al. (2021)). Along the same lines, surveys investigating perceptions about the entrepreneurial ability of men versus women in India do not reveal that women perceive themselves as inferior entrepreneurs (though of course, these perception could reflect selection bias).

An entrepreneur of gender g pays a fixed sunk cost of entry 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 formalization/registration of the business. 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 associated with formality.[16] Alternatively, an individual can pay a fixed cost E_{gW} and work for a (gender-specific) wage $\tilde{w}^g$. Lastly, an individual could pay a fixed cost E_{gO} to enter self-employment, in which case, s/he earns a stochastic income $\Pi_{gO} = \zeta \tilde{w}^g$, where $\zeta \sim H(\zeta)$, $H(0) = 0$, and $H(1) = 1$.[17] As with entrepreneurship, we assume that the fixed costs of entering wage- and self-employment

[16] See reviews by Jayachandran (2021) and Quinn and Woodruff (2019).

[17] The above assumption implies that the variable returns (i.e., excluding the fixed costs) of wage employ-

are gender-specific. We do not make any a-priori assumptions about the relative magnitudes of the fixed costs of the various options (so the fixed costs of entering wage-employment for instance could be lower or higher than the fixed costs of entering entrepreneurship), but let the data determine these magnitudes. Section 4.3 discusses how they are identified.

To summarize, the expected income for an individual with entrepreneurial ability x, who chooses to participate in the labor force is given by:[18]

$$I(x) = \begin{cases} b + \zeta\widetilde{w}^g - PE_{gO} & \text{(Self employment)} \\ b + \widetilde{w}^g - PE_{gW} & \text{(Wage employment)} \\ b + E(\Pi_{gI}(x)) - PE_{gI} & \text{(Informal entrepreneurship)} \\ b + E(\Pi_{gF}(x)) - PE_{gF} & \text{(Formal entrepreneurship)} \end{cases} \tag{7}$$

An individual will choose the occupation that maximises her/his expected income. They will work in wage employment as long as $\zeta \leq \zeta^*$, where $\zeta^* = 1 - \frac{E_{gW} - E_{gO}}{\widetilde{w}^g / P}$ and the fraction of individuals in wage-employment will be given by $H(\zeta^*)$. Since we observe non-zero entry in both the formal and informal sectors, there is a (gender-specific) threshold level of entrepreneurial ability in each sector x^*_{gs}, such that an individual with ability x^*_{gI} is indifferent between informal entrepreneurship and non-entrepreneurship (wage- or self-employment); and with ability x^*_{gF} is indifferent between informal and formal entrepreneurship. Lastly, from the LFP indifference condition discussed above, the threshold disutility that determines participation in the labor force is given by $\eta^*_g = E(I(x))/(\overline{u}_g P)$, and the fraction of individuals who participate in the labor force is given by $F(\eta^*_g)$.

Entrepreneurial choice across industries: We now turn to the decision of an entrepreneur to enter a particular industry j in sector s. We assume that an entrepreneur with entrepreneurial ability x and conditional on starting a firm in sector s, draws her/his ex-post industry-specific productivity $z_{sj} = x\varepsilon_{sj}$, where ε_{sj} is drawn from a gender-specific Frechet distribution, i.e., $\varepsilon_{sj} \sim Frechet(T_{sj}, \theta_g)$ with a CDF given by $F(\varepsilon) = e^{-(\varepsilon/T_{sj})^{-\theta_g}}$.

Proposition 1. *For each gender g, the share of entrepreneurs, their average firm size and profits in a sector*

ment are higher than the returns of self-employment. Since we only observe the number of wage-earners and self-employed in the Economic Census (and not their wages or profits), this assumption helps us incorporate the self-employed in the analysis while keeping the model tractable. From the 62nd Round (in 2005) of the National Sample Survey, a nationally representative survey of individuals, earnings of men and women in self-employment are 42.1% and 50.1% of those in wage employment on average.

[18]The fixed costs are measured in units of output, which implies that their expenditure is given by $PE_{(.)}$

s and industry j are given by:

$$
\begin{aligned}
&(a)\varphi_{gsj} = \frac{\left[\frac{p_{sj}T_{sj}}{w_{sj}^{\rho_s}}\right]^{\theta}}{\sum_k \left[\frac{p_{sk}T_{sk}}{w_{sk}^{\rho_s}}\right]^{\theta}} && [\ \textit{Share of Firms}\ ] \\
&(b)E[l_{gsj}(x)] = \varphi_{gsj}^{-1/\widetilde{\theta}_s}\Gamma_{\widetilde{\theta}_s}\left[\rho_s \frac{T_{sj}x}{w_{gsj}/p_{sj}}\right]^{\frac{1}{1-\rho_s}} && [\ \textit{Avg. Firm Size}\ ] \\
&(c)E[\pi_{gsj}(x)] = \frac{1-\rho_s}{\rho_s} \times w_{gsj}E[l_{gsj}(x)] && [\ \textit{Avg. Profits}\ ] \qquad (8)
\end{aligned}
$$

where: $\widetilde{\theta}_s = (1-\rho_s)\theta$, $\Gamma_a = \Gamma(1-1/a)$, $\{\rho_I, p_{Ij}\} = \{\lambda\rho, p_j\}$ *and* $\{\rho_F, p_{Fj}\} = \{\rho, (1-t_j)p_j\}$

Mathematical proof provided in Appendix B.3.

Summary: To summarize the above discussion, each individual in this economy is indexed by $\{g, x, \zeta, \eta\}$, i.e., gender g, entrepreneurial ability x, stochastic profit in self-employment ζ, and disutility of labor force participation η. An individual will enter the labor force as long as $\eta < \eta_g^*$. Conditional on working, individuals with $\zeta > \zeta^*$ will be self-employed, while those with $\zeta < \zeta^*$ will either become entrepreneurs or wage workers. Among those, individuals with $x < x_{gI}^*$ will enter wage employment, $x \in [x_{gI}^*, x_{gF}^*]$ will enter the informal sector as entrepreneurs, and those with $x > x_{gF}^*$ will enter the formal sector as entrepreneurs. Conditional on sector choice s, entrepreneurs draw an ex-post productivity signal z_{sj} that determines the industry j in which they operate.

2.5 Equilibrium

To close the model, we aggregate across all agents in the economy. Total income in the economy is given by $I = \widetilde{\boldsymbol{w}}\bar{\boldsymbol{L}} + \Pi + B$. The first term, $\widetilde{\boldsymbol{w}}\bar{\boldsymbol{L}}$, is the income received by the workers in the economy, and it is equal to $\sum_g \widetilde{w}^g L_{supply}^g$, where $L_{supply}^g = F(\eta_g^*)H(\zeta^*)G(x_{gI}^*)N^g$. The second term, Π, denotes the total profits of the firms in the economy net of their entry costs, i.e. it consists of: (i) earnings of the self-employed, given by $\Pi_O = \sum_g N^g F(\eta_g^*) \int_{\zeta_g^*}^1 (\zeta\widetilde{w}^g - PE_{gO})dH(\zeta)$; (ii) profits of the firms in the informal sector $\Pi_I = \sum_g \sum_j N^g F(\eta_g^*)H(\zeta_g^*) \times \int_{x_{gI}^*}^{x_{gF}^*} \varphi_{gIj}(E\Pi_{gIj}(x) - PE_{gI})$ and the profits of the firms in the formal sector $\Pi_F = \sum_g \sum_j N^g F(\eta_g^*)H(\zeta_g^*) \times \int_{x_{gF}^*} \varphi_{gFj}(E\Pi_{gFj}(x) - PE_{gF})$. The third term, B, denotes total benefits.

The total taxes collected in the economy are given by $TX = \sum_g \sum_j t_j p_j Y_{gjF}$, where $p_j Y_{gjF}$ is the total revenue of formal firms of gender g in industry j. Taxes are redistributed as benefits b across all individuals in the economy. Given the utility function, individuals spend a share α_j of their income on consuming goods from industry j. Labor demand for workers of gender g across all firms in the economy, denoted by L^g_{demand}, is given by $L^g_{demand} = \sum_{g'} \sum_j \sum_s L^g_{g'sj}$, where $L^g_{g'sj}$ is the total labor of gender g, demanded by entrepreneurs of gender g' in sector s and industry j given by Equations (1), (4) and (8). The equilibrium in this economy is defined by the following conditions:

(i) the labor markets clear for both genders; the goods market clears for each industry.

(ii) the zero-profit conditions for the formal and informal sectors, and the indifference conditions for LFP and self-employment hold with equality for both genders.

(iii) the total benefits received by individuals are equal to the taxes collected.

3 Data and Reduced Form Evidence

Our primary data comes from two rounds of the Economic Census of India (EC) for 1998 and 2005.[19] 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 1998 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 public-sector firms and co-operatives from our analysis since they do not have information on gender-ownership. We restrict our sample to the 18 major states of India, which cover 95-97% of male- and female-owned firms in both rounds of the EC.[20] Lastly, we define a "firm" as an establishment that hires at least

[19] 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 (2014), Ulyssea (2018), Ulyssea (2020)).

[20] 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.

one worker, while those that do not hire any workers are classified as "self-employed". Our final sample consists of 26.23 million firms in 1998 and 40.86 million firms in 2005.

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. Five stylized facts stand out. First, over half of Indian establishments are owner-operated, i.e., self-employed individuals, and the overwhelming majority of them are male. Second, the majority of (not owner-operated) firms in India are male-owned as well, and more than 99% of them (both male and female) operate in the informal sector. Firms in the formal sector are however significantly bigger. Third, female-owned firms (excluding self-employed) account for only 3% of the total establishments in the country. Fourth, as reported in Columns (3) and (4), female-owned firms are smaller (larger) than male-owned firms in the informal (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)) has decreased for all categories between 1998 and 2005. At the same time, the number of self-employed individuals as well as firms (both male and female) have significantly increased in this period. The combination of these two patterns suggests a decline in entry costs into both labor force and entrepreneurship. The decline in firm-size is particularly pronounced for formal firms (both male- and female-owned), suggesting a decline in the costs of formalization, especially for women.[21] The fraction of female employees (Columns (5) and (6)) across firms in both sectors has remained relatively stable in male-owned firms, but increased in female-owned firms.

To explore whether these patterns are driven by firm sorting either across space (districts in India), or across industries, we estimate regressions of the form:

$$y_{fjd} = \alpha_d + \alpha_j + \beta_1 \text{Female}_f + \beta_2 \text{Formal}_f + \beta_3 \text{Female}_f \times \text{Formal}_f + \delta X_{fjd} + \varepsilon_{fjd} \tag{9}$$

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.), and a dummy for whether the firm operates in a rural or urban area. We cluster standard errors at the district level.

[21]This is also 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 (MSMEs).

Table 2 reports the results. Panel A reports the regressions with district fixed effects (α_d), but without industry fixed effects (α_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 4.35 log-points (or 4.4%) smaller in size than male-owned informal firms, but 14.05 log-points (or 15%) 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 23.5 p.p. in the informal sector, and 16.74 p.p. 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 not significantly 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.[22]

4 Model Estimation

The purpose of quantifying the model is twofold. First, we estimate the hiring wedges and excess fixed costs of entry and registration. Second, we evaluate the impact of counterfactual policies that eliminate the entry, registration and hiring barriers faced by female entrepreneurs. Table 3 lists the model parameters. Given data limitations, we use a combination of calibration and estimation to set their values.

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[23]. As noted earlier, we use the 1998 and 2005 rounds of the Economic Census and allow for different parameters for each round.

[22]These results are also robust to excluding "family-owned" firms, which are defined as those where more than half the employees are not hired on wage contracts. The results are reported in Table A1.

[23]In principle, our data allows for a more disaggregate analysis at the 4-digit NIC level, and we have in fact experimented with specifications based on this more disaggregate industry definition. The key challenge however is that there are very few female firms (especially formal female firms) in several of these industries, the disaggregate analysis is not particularly meaningful as we cannot recover the entry costs, hiring frictions, etc., in these NIC industries that have very few or no female entrepreneurs. We discuss this in Appendix C.1. However, the reduced form results of Table 2 suggest that the estimates are not severely affected by the inclusion of industry fixed effects. For these reasons, and also to facilitate presentation of the estimates, we have opted to aggregate the 4-digit level NIC data to three broader "industries": Agriculture; Manufacturing; and Services in our baseline model. However, in Appendix C.1,

We classify our parameters into two sets:

(a) Fundamental parameters $\{\Gamma, \Psi\} = \left\{ \{\rho, \gamma, \alpha_j, t_{jr}\}, \{\lambda_j, A_{sjr}, T_{sjr}, \sigma_x^2, \theta_g\} \right\}_{\forall g,j,r}$

(b) "Barriers" faced by individuals and entrepreneurs, including the fixed costs of entry into the various employment options $\Upsilon = \{\overline{u}, E_O, E_w, E_I, E_R\}_{\forall g,r}$ and hiring wedges $\Theta = \{\tau_{fI}, \tau_{fF}, \tau_{fI}^f, \tau_{fF}^f\}_{\forall j,r}$.

The parameters in Γ are determined based on statutory values or taken from the literature. The parameters in Ψ and all barriers faced by entrepreneurs (Υ,Θ) are estimated. Note that conditional on the expected profits in each occupation (self-employment, workers, entrepreneurship, etc.), the occupational choice depends on the differences in the fixed costs as opposed to their levels, i.e., they are invariant to an additive scale. Moreover, from the expression for η^*, we can only identify either $\overline{u}_g$ or E_O. Therefore, we therefore normalize $E_O = 0$ for both genders, so that all other fixed costs are interpreted *relative to* the fixed costs of entering self-employment.

Similar to Bick, Fuchs-Schündeln, Lagakos and Tsujiyama (2022), we assume the individual disutility of work follows a uniform distribution, i.e., $\eta \sim U(0,1)$. This implies that the average disutility by gender, $\overline{u}_g$, is distributed according to $\eta\overline{u}_g \sim U(0,\overline{u}_g)$. We also assume $\zeta \sim U(0,1)$ and entrepreneurial ability for an entrepreneur g follows a log-normal distribution with mean 0 and variance σ_x^2, i.e., $x \sim \log N(0, \sigma_x^2)$. Further, we assume the realized industry-sector specific productivity $z_{sj} = x\varepsilon_{sj}$, where $\varepsilon \sim Frechet(T_{sj}, \theta_g)$.[24]

We normalize the productivity of male workers A^m to be 1, and the hiring barriers faced by male entrepreneurs to be zero, i.e., $\tau_{mI} = \tau_{mF} = 0$ and $\tau_{mI}^f = \tau_{mF}^f = 0$. These normalizations are harmless, but imply that the productivity of female workers as well as the hiring barriers faced by female entrepreneurs (i.e., τ_{fs} and τ_{fs}^f) are to be interpreted *relative to* their male counterparts. Finally, the relative worker productivities (A^g) are identified from the ratio of female to male workers in male-owned firms across industries (within a sector). However, the levels for A^g cannot be separately identified from gender-specific wages. We can only identify the relative productivity across industries and hence have to assume that the productivity of male and female workers is equal to 1 in one industry (services in our case).[25]

we estimate the model for nine 1-digit industries (rather than the three aggregate industries described above), as the 1-digit level is the most disaggregate level at which we can meaningfully estimate fixed costs of entry into entrepreneurship. The results are similar to those obtained when the model is estimated at the more aggregate level. Importantly, our finding that female entrepreneurs hire more female workers is robust to the estimation at the more disaggregate level.

[24] As discussed earlier, our baseline specification does not allow for the distribution of innate entrepreneurial ability to vary by gender. However, note that the ex-post realizations of productivity can be gender-specific across industries and sectors. Appendix F provides a detailed discussion of these assumptions. Moreover, in Appendix F.3, we allow the variance σ_x to vary by gender, and show that it does not affect our results in a meaningful way. If anything, the results become quantitatively stronger.

[25] The industry we base the normalization on does not affect our results. Services is a natural choice given

4.2 Exogenous Model Inputs from the Literature

The parameters in Γ are determined using statutory values or values taken from the literature as follows: We fix the share of consumer expenditure on an industry, i.e., $\{\alpha_j\}_{\forall j}$, 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 parameter γ measures the elasticity of substitution between male and female workers in production. A rich literature estimates this elasticity of substitution and the values typically vary between 1.7 to 2.3 across studies and contexts.[26] We set $\gamma = 2.1$, which the average of the values estimated in this literature. Lastly, the 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. The tax rates are between 5-8% across industries and are consistent with the sales tax on most products in India during that period.

4.3 Estimation Strategy

This section outlines the estimation procedure and provides some heuristic arguments of how the remaining parameters are identified (conditional on the parameters in Γ)[27]. In a nutshell, we jointly use moments from male- and female-owned firms to estimate the parameters in Ψ and Υ, and then use the differences between moments of male-owned and female-owned firms to identify the parameters in Θ. We use our model to simulate moments that we can observe in the data. We employ a Simulated Minimum Distance (SMD) estimator, which minimizes the distance between the simulated and actual moments in the data. Table 3 provides a list of all the parameters along with the moments that are targeted to identify them.

We first discuss the moments in the data we target to estimate the parameters in Ψ. We normalize the productivity of female workers (relative to male) in services to equal 1 in both the formal and informal sectors, i.e., we set $A_{s,Services,r} = 1$. From Equations (1) and (4), the ratio of female to male workers in a given sector, industry (and region) is given by $A_s(w^f_{gs}/w^m_{gs})^{-\gamma}$. We target the

that the literature documents the lowest gender productivity gaps in services, compared to agriculture and manufacturing (Pitt et al., 2012; Rendall, 2013; Olivetti and Petrongolo, 2014).

[26] See Hamermesh (1996); Udry (1996); Weinberg (2000); Acemoglu, Autor and Lyle (2004); De Giorgi, Paccagnella and Pellizzari (2013); Johnson and Keane (2013); Olivetti and Petrongolo (2014); Ghosh (2018).

[27] We discuss identification more systematically in Section 5.4 where we employ an approach similar to Kaboski and Townsend (2011) and Bick, Fuchs-Schündeln, Lagakos and Tsujiyama (2022) to establish identification.

ratio of female to male workers in male-owned firms in agriculture and manufacturing (relative to services) ($2 \times 2 \times R$ moments) to estimate $\{A_{Ijr}, A_{Fjr}\}_{\forall j,r}$. Similarly, we normalize $T_{Services,r} = 1$ in both the formal and informal sectors, and identify $\{T_{sjr}\}_{\forall j,r}$ for agriculture and manufacturing using the share of male-owned firms in each sector in these industries (Equation 8), relative to services ($2 \times R$ moments). The penalty of informality $\{\lambda_j\}_{\forall j}$ is identified using the average ratio (across all regions) of firm-size of male-owned firms in the informal to formal sector (Equations 2 and 5) for each industry separately (3 moments). Lastly, we use the variance of firm-size for male-owned and female-owned firms in the formal and informal sector (4 moments) to estimate $\{\sigma_x^2, \theta_{gs}\}_{\forall g}$.

Regarding the parameters in Υ, we identify the fixed costs of labor force participation, wage work, and informal and formal entrepreneurship i.e., $\{\overline{u}_{gr}, E_{W,gr}, E_{I,gr}, E_{R,gr}\}_{\forall g,r}$ using the labor force participation rate for men and women ($2 \times R$ moments), number of men and women in wage work ($2 \times R$ moments), and the number of male-owned and female-owned firms (as a fraction of the gender-specific labor force) in the informal and formal sectors ($2 \times 2 \times R$ moments). As noted earlier, we can only identify these fixed costs of wage work and entrepreneurship relative to self-employment, which also cannot be separately identified from a (gender-specific) LFP cost. Accordingly, we normalize $E_{O,gr}$, i.e., the fixed cost of self-employment conditional on participating in the labor force, to be equal to 0.

Turning to hiring frictions faced by female entrepreneurs (Θ), since we normalize τ_{mI} and τ_{mF} to be equal to zero, we use the ratio of the average firm size of male-owned and female-owned firms in the formal and informal sectors ($2 \times J \times R$ moments) to identify $\{\tau_{fI,jr}, \tau_{fF,jr}\}_{\forall j,r}$. Similarly, we use the ratio of the ratio of female to male workers in male-owned and female-owned firms in the formal and informal sectors ($2 \times J \times R$ moments) to identify $\{\tau^f_{fI,jr}, \tau^f_{fF,jr}\}_{\forall j,r}$.

Given a guess of the parameter vector $X = \{\Psi, \Upsilon, \Theta\}$, we simulate the above moments from the structure of the model to obtain the vector $M(X)$. The data counterpart is denoted by M_{data}. We then choose the parameter vector $\hat{X} = \arg\min g(X)'g(X)$, where $g(X) = (M(X) - M_{data})/M_{data}$.

5 Parameter Estimates

We start by discussing the parameter estimates for entrepreneurial ability, technology, worker productivity, and penalty of informality (i.e., Ψ) in Section 5.1, fixed costs of entrepreneurship and LFP (i.e., Υ) in Section 5.2, and finally the barriers faced by women entrepreneurs (i.e., Θ) in Section 5.3. In Section 5.4 and Section 5.5 we discuss identification and model fit. In Appendix D we correlate our estimates with existing measures of women empowerment and gender-specific policies and show that they are consistent with common wisdom.

5.1 Comparative Advantage, Technology, Informality, and Entrepreneurial Ability

Table 4 reports the estimates for the parameters of the productivity of female relative to male workers in production in the informal and formal sectors (A), technology (T), penalty of operating in the informal sector (λ), and entrepreneurial ability distribution $\{\sigma_x^2, \theta\}$.

The estimates for A_I show that women working in the informal sector do not have a significant comparative (dis)advantage relative to men in either agriculture or manufacturing (relative to services). In the formal sector (A_F), women have a disadvantage of working in agriculture (especially in 2005) and manufacturing relative to services.[28] Relative to services, the technology parameter (T) is around half in agriculture (in both sectors) and about 0.60-0.70 in manufacturing. The size-based penalty of operating in the informal sector (λ) is 0.97 in agriculture and manufacturing, and 0.92 in services in 1998. In 2005, it is 0.89, 0.92, and 0.93 in agriculture, manufacturing, and services respectively. In Appendix B.2, we discuss how these estimates relate to size-based penalties (such as the probability of detection for example).

Despite allowing for the ability distribution to vary across years, the parameter estimates are remarkably similar between 1998 and 2005 ($\sigma_x \approx 0.30$). In our baseline estimation, we restrict σ_x to be the same across genders, so that the ex-ante entrepreneurial ability distribution is the same for men and women. In Appendix F we provide an extensive discussion of this assumption by relating it to data on education as well as to responses on surveys of women's aptitude for entrepreneurship. In the same appendix (Section F.3), we also examine the ramifications of relaxing this assumption by allowing the shape parameter σ_x to vary by gender, and obtain very similar results to the baseline specification. The parameter $\widetilde{\theta}_{gs} \equiv \theta_{gs}(1-\rho)$ is 2.19 (2.22) in the informal sector, and 2.02 (2.04) in the formal sector for men (women) in 1998. It is 2.44 (2.74) in the informal sector, and 2.14 (2.15) in the formal sector for men (women) in 2005. These values are consistent with estimates from Hsieh, Hurst, Jones and Klenow (2019), who using a similar modeling structure for the US, and estimate a value of 2.57. Note that our parameter estimates imply that *ex-post*, women are slightly more productive than men (though their *ex-ante* ability is assumed to be the same). These results are in line with those of Ashraf, Bandiera, Minni and Quintas-Martinez (2022), who also find that women are more productive than men using an entirely different methodology and data from a large multinational firm covering 101 countries.

[28] This is consistent with a literature on the importance of brawn versus brain (Pitt et al., 2012) as well as the impact of the rise of service industries on female labor force participation (Rendall, 2013; Olivetti and Petrongolo, 2014, 2016; Ngai and Petrongolo, 2017).

5.2 Fixed Costs of Entrepreneurship and LFP

The fixed costs include those for entry and formalization in the informal and formal sector, entry into wage work, and the disutility of participating in the labor force (which we shall call LFP costs). We estimate these costs separately for male and female entrepreneurs, region (r), and year (t). Figure 2, reports the average (across regions) values for men and women using blue triangles and red circles respectively. The dash (solid) lines show a one standard error band around the averages in 1998 (2005). Several interesting patterns emerge.

First, women faced around 2.4 times the cost of entering the labor force compared to men in 1998. While LFP costs have declined for both men and women over time, women still face around twice the cost of participating in the labor force in 2005. Second, conditional on entering the labor force, the fixed costs of working in wage work or starting informal entrepreneurship (relative to self-employment) are quite low–both for men and women, and over time. In contrast, there is a stark gender-difference in the costs of formalizing informal businesses. Compared to men, women faced around twice the cost to formalize their business in 1998. Despite a reduction in these costs over time, for both men and women, they remained around 25% higher for women than men in 2005.

The low fixed costs of wage work and informal entrepreneurship (relative to self-employment), for both men and women, may seem surprising at first, given that wage work is considered highly desirable in many low-income countries, and women have been shown to be reluctant entrepreneurs (Jensen, 2022; Schoar, 2010). The low estimates likely reflect the heterogeneity of wage employment and informal entrepreneurship. Many wage jobs are low-paying and provide no benefits. Similarly, some informal enterprises barely differ from self-employment (in the sense that they may employ two, instead of just one, people, but are otherwise of similar size and productivity as owner-operated businesses). Such options may not entail the high fixed costs of entry one typically associates with "good" wage jobs or successful enterprises.

Appendix C.2 examines one particular source of heterogeneity related to the employment of "non-hired" workers[29]. Non-hired labor is pervasive in the informal sector, in both male- and female-owned firms. Non-hired workers are treated as "wage workers" in our framework. But given that they do not go through a formal hiring process, they presumably face lower fixed costs of entering wage employment. To understand the role of non-hired labor in the fixed cost estimation, we classify non-hired workers as self-employed, and then re-estimate the model to obtain new fixed cost estimates. This scenario, though extreme, is useful as a benchmark because classifying non-hired workers as self-employed implies that their income is lower in expectation than that of hired wage employees. As Figure C3(b) demonstrates, treating non-hired workers as

[29]Non-hired workers are typically household members working in smaller firms and/or apprentices.

self-employed substantially increases the fixed costs of wage employment relative to the baseline, for both genders, but especially for women (by more than six times). These results are reassuring, in the sense that they rationalize our baseline estimates. The effects on the other types of fixed costs are very small. Given that our focus is on entrepreneurship, and not on wage work, we proceed with our baseline classification.

In Appendix D, we use region-specific measures of women empowerment from various sources in the literature to examine whether our implied measures of gender-related barriers correlate with the documented level of women empowerment in these regions. We find that states with more conservative gender norms do have higher relative LFP costs for women as compared to men. There is no statistical association between gender empowerment and formalization costs, though the estimated coefficients have the expected signs.

5.3 Distortions in Hiring Workers

For each industry j, region r, and year t, we quantify two types of barriers that may distort the expansion/hiring of female-owned firms. First, τ_{fsj} is the additional cost of employing a worker for a female entrepreneur in sector s and industry j, relative to her male counterpart. We remind the reader that we have normalized $\tau_{msj} = 0$. Accordingly, the marginal cost faced by female entrepreneurs (relative to male entrepreneurs) is expressed in relative terms as $1 + \tau_{fsj}$. Similarly, $1 + \tau^{f}_{fsj}$ is the additional marginal cost incurred by women entrepreneurs relative to male entrepreneurs, in employing female workers relative to male workers, in sector s and industry j (again, we remind the reader that we have normalized $\tau^{f}_{msj} = 0$).

As shown in Table 5, the cost of employing a worker is on average around 20-25% (10-15%) higher for women entrepreneurs in the informal (formal) sector compared to men. Figures 3(a) and 3(b) plot the distribution for $1 + \tau_{fI}$ and $1 + \tau_{fF}$ across regions and industries in 2005. A value greater than 1 implies female-owned firms face a higher marginal cost than male-owned firms. $1 + \tau_{fI}$ ranges from approximately 1.10 to 1.30, while $1 + \tau_{fF}$ ranges from 1.00 to 1.40 across most industries and states. These results indicate that women-owned businesses (both formal and informal) face substantial barriers in employing workers, both across industries and states. In Appendix D, we show that these barriers are lower in states with more progressive gender norms.

Turning to the gender composition of workers, i.e., $1 + \tau^{f}_{fs}$, the estimates indicate that this is the only area in which female entrepreneurs have an advantage, and more so in the informal sector. From Table 5, female entrepreneurs in the informal sector incur 5-6% lower costs to hire a female (relative to male) worker, relative to male entrepreneurs. This advantage is still present, but muted in the formal sector, where despite the average being equal to 1, the median is 0.93 in

1998 and 0.95 in 2005. Figures 3(c) and 3(d) display the heterogeneity across industries and states. The advantage for female entrepreneurs in employing women (relative to men) in the informal sector is quite substantial, over 10% in some industry-regions. It is also present in most cases in the formal sector.

Note that Figures 3(c) and 3(d), as well as the reduced form results of Table 2, show that this pattern is not driven by selection of female workers and entrepreneurs into a few industries, as it is present in every industry, even at a highly disaggregate level as in Table 2[30]. Furthermore, in Appendix C.1, we estimate the model at a more disaggregate level: the NIC 1-digit industry level instead of the level of three aggregate industries (agriculture, manufacturing, and services).[31] We find that the estimates of the hiring frictions are very similar to those in our baseline estimation, and that the distributions of these frictions (especially τ^{f}_{fsj}) overlap greatly.

The comparative advantage of female entrepreneurs in employing female workers may itself 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 households if they seek work outside the home, such resistance may be less pronounced in cases where they work for other women. Rigorously examining the sources of the pattern we document using micro-economic data and surveys is an interesting question for future research.

To summarize, the estimates suggest that while the excess barriers female entrepreneurs face have been reduced over time, there nevertheless remains a substantial gender gap across industries and regions. On the extensive margin, the excess barriers women face affect primarily their labor force participation decisions and entry into formal entrepreneurship. In contrast, the fixed costs of entering wage work or starting an informal enterprise, *conditional on LFP*, are not excessive. On the intensive margin, female entrepreneurs face additional costs of expanding their businesses, so even though entry into the informal sector may not be particularly costly, growth still is. The only exception is in the employment of female workers, where female entrepreneurs appear to have an advantage.

5.4 Identification

Section 4.3 provided heuristic arguments of how various data moments help identify the key parameters of the model. We now adopt a more systematic approach for establishing identification

[30] As reported in Table A1, these patterns are robust even when we exclude "family-owned" firms.

[31] As discussed earlier, in principle, our data allows for a more disaggregate analysis, and we have in fact attempted estimation based on more disaggregate industry definitions. The key challenge however is that there are very few female firms in several disaggregate industry-region pairs. As a result, with highly disaggregate data, we cannot recover the entry costs in those NIC industries that have very few or no female entrepreneurs. We discuss this issue in detail in Appendix C.1.

in the spirit of Kaboski and Townsend (2011) and Bick, Fuchs-Schündeln, Lagakos and Tsujiyama (2022). Specifically, for each of the seven sets of key model parameters, namely: hiring distortions faced by women entrepreneurs ($\{\tau_{fIj}, \tau_{fFj}\}$ and $\{\tau^f_{fIj}, \tau^f_{fFj}\}$), relative productivity of female workers (A_{Ij} and A_{Fj}) and penalty of operating in the informal sector (λ_j), we compute the derivative of a moment with respect to each parameter.[32] To do so, we re-solve the model each time by increasing one parameter by 1 p.p. above its estimated value (keeping all others the same) and compute the resulting percentage changes in the simulated moments. We report the results in Table A2. Each number in a row r and column c is the percentage change in the moment in row r (averaged across regions, industries and gender) when the parameter in column c is increased by 1 p.p. (keeping all other parameters the same). We bold-face and underline the largest derivative in each column to highlight which moment responds the most when the parameter in that column is changed. Panel A (B) in Table A2 reports the results using the 1998 (2005) Round of the Economic Census.

As the table shows, the results are consistent with the discussion in Section 4.3. From Columns (1) and (2), we see that the ratio of female to male workers in a male-owned firm in the informal and formal sectors is sensitive to changes in the relative female to male worker productivity (A_I and A_F). On the other hand, from Columns (3) and (4), the ratio of female to male workers in female-owned (relative to male-owned) firms in the informal (formal) sector is substantially affected by the change in τ^f_I (τ^f_F). From Columns (5) and (6), the ratio of female to male firm-size in the informal and formal sectors is most responsive to the hiring barriers that female entrepreneurs face (τ_I and τ_F). Lastly, in Column (7), the ratio of firm-size of male-owned firms in the formal and informal sectors is most sensitive to the penalty of operating the informal sector (λ).

5.5 Model Fit

Tables A3-A5 in the Appendix show the fit of the model for the 2005 data.[33] In Panel A of Table A3, we start by discussing the allocation of men and women across the economy. Since these moments are generated at the region-level, we average them across regions and report the standard deviations in parentheses. In particular, we show the model fit across five moments, the fraction of men and women in: (a) the labor force; (b) self employment; (c) wage employment; (d) informal entrepreneurship and (e) formal entrepreneurship. In Panel B, we also examine the ratio of female to male workers in informal and formal male/female-owned firms. These sets of moments were directly targeted by the model, and we fit them very well.

[32]Note that the fixed costs of entry and formalization are not identified based on data moments, but are computed based on the the zero-profit conditions. Accordingly, we do not discuss them in this subsection.

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

In Table A4, we examine moments related to the distribution of firm size across the four types of firms in our data. In Panel A, we examine the fit of our model using moments related to the ratio of firm size of female-owned to male-owned firms in the informal/formal sector. We also examine the ratio of firm-size between the informal and formal firms for the same gender-owned firms. Our model fits these moments very well. Note that we do not target moments related to the ratio of formal to informal firm size for female-owned firms, yet our model fits those well. In Panel B of Table A4, we show the fit of the model for average firm size. Note that our estimation strategy only targets the *ratios* of firm size across gender/sector, but not the levels. Despite, this, the model does a decent job at fitting the levels as well.[34] Lastly, Panel C shows that the model also fits the standard deviation of firm size in the formal and informal sectors.

Lastly, in Table A5, we examine moments related to the share of male-owned and female-owned firms across sectors and regions. Panel A reports the average (and standard deviations) across states for the three industries in the informal sector, while Panel B reports the same for the formal sector. As we can see, the model is able to replicate the sorting of male-owned (Columns 1-2) and female-owned firms (Columns 3-4) across industries and both sectors. Note that we only target the relative shares for male-owned firms in both sectors and yet, the model is able to match their levels for both men and women (who we do not target at all). Figure A2 then shows the distribution across all regions and industries. Each dot on the figure is the share of male-owned (blue triangles) and female-owned firms (red circles) in each industry and region in the data (horizontal axis) and model counterpart (vertical axis). The solid line is the 45 degree line. As we can see the model does a good job at matching these moments well in both sectors.

6 Impact of Affirmative Action Policies

Apart from quantifying the various types of barriers faced by female entrepreneurs, the advantage of our 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 five policies that sequentially eliminate the excess barriers faced by females in the economy on both the extensive margin (i.e., participation in the labor force, wage work, self-employment, and informal/formal entrepreneurship) and intensive margin (i.e., expansion through hiring workers). This exercise allows us to identify the barriers that are most consequential for aggregate productivity and welfare. We consider the following scenarios that eliminate:

(i) **Excess fixed costs:** We eliminate the excess fixed costs faced by women in both wage work and

[34] In Figure A1, we show the model fit in firm size across all industries and regions for both male and female firms in the formal and informal sectors.

entrepreneurship (informal and formal), i.e., we set $E_{fW} = \min\{E_{mW}, E_{fW}\}$, $E_{fI} = \min\{E_{mI}, E_{fI}\}$, and $E_{fR} = \min\{E_{mR}, E_{fR}\}$.
(ii) **Excess hiring barriers:** We set $\{E_{gW}, E_{gI}, E_{gR}\}$ to their baseline values, but eliminate excess hiring barriers. That is, we set $\tau_{fs} = \min\{\tau_{fs}, 0\}$ and $\tau^f_{fs} = \min\{\tau^f_{fs}, 0\}$, for $s = \{I, F\}$.
(iii) **Excess fixed costs and hiring barriers:** We eliminate all excess entrepreneurial costs as well as all hiring barriers in (i) and (ii).
(iv) **Excess LFP costs:** In scenarios (i)-(iii), we do not change the excess LFP costs faced by women, which from Table 3 are substantial. In scenario (iv), we set all fixed entrepreneurial costs and hiring barriers to their baseline values, and remove only the excess costs faced by women for participating in the labor force i.e., set $\overline{u}_f = \min\{\overline{u}_f, \overline{u}_m\}$.
(v) **All excess barriers:** In a final counterfactual, we remove all excess barriers faced by women on labor force participation, fixed costs of informal and formal entrepreneurship, and intensive margin hiring barriers.

We examine the effects of these policies on labor force participation rates for men and women, the allocation of men and women across wage employment and entrepreneurship, and the earnings of men and women workers (measured by real wages) as well as entrepreneurs (measured by average real profits). The results are displayed graphically in Figure 4. Then, for each region, we aggregate across workers to measure the impact of each policy on productivity, which we measure as the average productivity of firms in a region across sectors and industries, and real income, which given our preference structure, is a natural candidate for measuring welfare. These results are shown in Figure 5. Figures 4 and 5 display results for all scenarios, but we focus our discussion on the last three.

Removing Excess Fixed Costs of Wage Work, Entrepreneurship, and Hiring Frictions

Consider a counterfactual that removes all excess costs faced by female entrepreneurs, while leaving the costs to labor force participation unchanged. From Figure 4(a), female labor force participation increases by 4 p.p. compared to the baseline, the fraction of women who are entrepreneurs increases to 5% (from 2% at baseline), and those in self-employment decrease by 0.2 p.p. (5.7%). Real wages for women increase by 6.5% (Figure 4(c)), and real profits of women entrepreneurs increase by around 25% (Figure 4(d)). Real wages and profits increase for men too, but in relative terms, both female workers and female entrepreneurs gain relative to male workers and male entrepreneurs.

Turning to productivity, average productivity of female-owned firms *decreases* by 10.7% (Figure 5(a)). This can be rationalized in Figure 5(b), which shows, for the baseline as well as for all counterfactual scenarios, the productivity of the marginal entrepreneur, i.e., the entrepreneur who is just indifferent between starting a firm or not. 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 30 percent more productive than her male counterpart. The removal of excess fixed costs allows more women to enter entrepreneurship, presenting male entrepreneurs with more competition. Accordingly, the productivity of the marginal female (male) entrepreneur decreases (increases), which results in a decrease (increase) in the productivity of the average female (male) entrepreneur. Nevertheless, the marginal female entrepreneur is still 8 p.p. more productive than her male counterpart, and this compositional shifts translates into median aggregate productivity gains of 2.4% across Indian states (with a 25th-75th percentile increase of 1.78-2.95%) and median real income gains of 10% (with a 25th-75th percentile gains of 6.57-11.10%), as reported in Figures 5(c) and 5(d) respectively.

Removing Excess LFP Costs

From Section 5.2, we know that LFP costs are substantial and twice as large for women on average. We therefore consider a counterfactual policy that sets the entrepreneurial costs and hiring barriers to their baseline values, but removes excess LFP costs for women, i.e., we set $\overline{u}_f = \min\{\overline{u}_m, \overline{u}_f\}$.

The fraction of women in the labor force increases by 21 p.p. (68%) in this scenario (Figure 4(a)). The fraction of female entrepreneurs increases by 2.7 times to 5.4% (from 2% at baseline), the fraction of those in self-employment increases to around 4% (from 3.5% at baseline), and the fraction of those in wage work increases by around 1.7 times. There are no substantial changes in the allocation of men in the economy (Figure 4(b)), but there is still a modest increase of 4 p.p. in the fraction of men who participate in the labor force, and an equally large increase in the fraction of men who are wage earners. This increase is due to the 7.3% increase in the earnings of male workers (Figure 4(c)). On the other hand, real wages for female workers *decrease* by around 10%. This decrease is due to the fact that we do not change the entrepreneurial barriers faced by women and therefore, while this counterfactual increases female labor supply, it does not adequately stimulate labor demand through the creation of new female-owned firms. Similarly, the average profits of male-owned firms increase by 10%, while the profits of female-owned firms do not change (Figure 4(d)). Turning to productivity (Figures 5(a) and 5(b)), while the average and marginal ability of the male entrepreneurs do not change much relative to the baseline, the average productivity of female entrepreneurs decreases by 4.2%, and the threshold productivity of the marginal woman entrepreneur also decreases (by 7 p.p.). These changes translate into median aggregate productivity gains of 1.14% across Indian states (with a 25th-75th percentile increase of 0.64-1.74%) and median real income gains of 30% (with a 25th-75th percentile gains of 17.7-37.7%). The effects on productivity are lower than in the previous counterfactual that removed all excess entrepreneurial barriers faced by women. However, the effects on income are larger, despite the fall in real wages, because now there are more women who choose to participate in the labor force and earn income.

Removing All Excess Barriers

The last counterfactual we consider is the removal of all excess barriers faced by women. This includes not only the labor force participation barriers, but also the excess barriers to wage work, entrepreneurship and hiring. As a result, labor force participation of women doubles (to 60%), and the fraction of women who are entrepreneurs increases more than five-fold, to 11% (Figure 4(a)). Interestingly, for men (Figure 4(b)), there is 7 p.p. increase in wage workers, with a corresponding 2.4 p.p. decrease in entrepreneurs.

These patterns can be rationalized by examining the changes in real wages (Figure 4(c)) and profits (Figure 4(d)). Male workers experience a 13% increase in real wages and a 8% increase in real profits, which explains the reallocation of men into wage employment discussed previously. For women on the other hand, real wages decrease slightly by 3.5%, while real profits increase by 25%. These effects contrast sharply with the previous scenario that reduced only labor supply barriers and which resulted in a large (almost 10%) decrease of real wages and no increase in real profits for women.

Lastly, the average ability of male (female) entrepreneurs increases (decreases) relative to the baseline by 6% (13.4%), which is rationalized by the increase (decrease) in the ability of the marginal male (female) entrepreneur (Figures 5(a) and 5(b) respectively). This scenario equalizes the ability of the marginal male and female entrepreneur (Figure 5(b)). However, the less productive male entrepreneurs (who exit) are now replaced by more productive female entrepreneurs (who enter). This reallocation channel improves the aggregate productivity in the economy by 3% across Indian states (with a 25th-75th percentile increase of 2.5-3.6%) and results in median real income gains of 43.5% (with a 25th-75th percentile gains of 35.7-55.3%).

Discussion of Results

The counterfactual scenarios considered above lead to several policy-relevant insights.

First, the barriers faced by women are substantial, both with respect to entrepreneurship and with respect to their participation in the labor force. Their removal has quantitatively meaningful impacts on aggregate productivity and welfare.

Second, policies targeting the intensive margin of growing a business through the hiring of workers have far greater impact than those targeting the fixed costs of entry and formalization. Intuitively, interventions that lower the costs of entry will have minimal impact on women's labor allocation decisions if distortions preventing them from succeeding post-entry remain in place.

Third, removing the barriers to operating businesses not only helps female entrepreneurs, it also benefits female workers relative to male workers in the form of higher real wages.

Fourth, policies that target women entrepreneurship, also improve female labor force participation, which is particularly important in the Indian setting where female labor force participation is low. This is both because more women become entrepreneurs and because, with more female entrepreneurs, more women are willing to enter the labor force as wage earners given that female entrepreneurs hire more female workers.

Fifth, policies that mitigate the excess costs to labor force participation alone significantly boost the labor supply of women. However, they do not boost the creation of female-owned businesses as much, thus depressing real wages and profits of women in equilibrium. Nevertheless, despite these lower wages, since more women are now wage earners, and the marginal women entrepreneurs who start firms replace less-productive male entrepreneurs, there are aggregate productivity and welfare gains in the economy.

Sixth, our results highlight the importance of addressing both labor supply and labor demand distortions. The elimination of barriers to female labor force participation increases (as expected) female labor force participation and boosts productivity and average real income. But the larger supply of women results in substantially lower real wages for them, while average profits in female-owned firms increase marginally. In contrast, boosting labor demand (in addition to labor supply) for women mitigates real wage declines stemming from the increase in labor supply, and results in additional profits for female-owned firms, as well as larger aggregate productivity and real income gains.

Lastly, all our 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. In conclusion, affirmative action policies that can effectively target both the constraints to labor force participation as well as barriers to entrepreneurship are highly effective in boosting productivity and welfare, both for women and the economy as a whole.

7 Impact of Removing all Gender-Based Frictions

The counterfactual exercises in Section 6 studied the implications of removing excess barriers faced by women, either in entering entrepreneurship, or in expanding their businesses. In cases where women had an advantage, i.e., hiring female workers, we left the advantage intact. In this section, we examine the implications of counterfactual exercises that equalize all gender-based frictions. Specifically, we consider two scenarios that equalize:

(i) **Entry and Hiring Barriers:** We equalize the fixed costs and hiring barriers between men and women. First, we set the fixed costs for an activity A (wage work, informal and formal entrepreneurship) to be equal between men and women $E_{fA} = E_{mA}$. Second, we set $\{\tau_{fs}, \tau_{fs}^{f}\}$ to be equal to 0.

(ii) **Removing All Barriers:** In addition to equalizing all fixed costs and hiring barriers, we now also equalize LFP costs between men and women i.e., we set $\overline{u}_f = \overline{u}_m$.

The above scenarios correspond to those discussed in Section 6. We report the results in Figure 6.

From Figure 6(a), we see that equalizing hiring and entry barriers increases the number of women entrepreneurs by 50% in the economy (from 2% to 3.2%). It also increases female LFP by 2 p.p. The fraction of women in wage work declines by 6 p.p. with a corresponding increase in self-employment (7.5 p.p.). Equalizing all barriers (including LFP costs) increases the fraction of women entrepreneurs by 6 p.p. (3.5 times), and those in self-employment by 9.5 p.p. (three-fold increase). It almost doubles FLFP, and increases the fraction of women in wage-work by 10 p.p. These results are similar in direction to the ones discussed in Section 6 (Figure 4(a)), but smaller in magnitude. Notably, the impacts are different for the share of self-employed women.

From Figure 6(b), we see that profits of women-owned firms decrease by 3.5%, and female real wages increase by 14% when we equalize entry and hiring barriers. However, when we equalize all barriers, there is no change in real wages for women, with a 7.7% decline in real profits. Lastly, the impact on average ability of women entrepreneurs and the ability of the marginal entrepreneur are similar to the previous case. Specifically, the average ability of women entrepreneurs (Figure 6(b)) decreases by around 9% and 13% (as compared to baseline), driven by a decline in the ability of the marginal woman entrepreneur (Figure 6(c)). However, the women who enter entrepreneurship replace male entrepreneurs of even lower productivity (Figure 6(c)), thus increasing median aggregate productivity by around 2% across all Indian states across both counterfactuals. Median real income increases by 1.10% when we remove all entry and hiring barriers, and by 30.5% when we remove all barriers (including those for LFP).

Discussion: The counterfactual exercises we discuss in Section 7 provide a few insights in addition to the ones discussed previously. Eliminating *all* gender-based frictions generates ca. 83% of the increase in FLFP, ca. two-thirds the change in aggregate productivity, and 70% of the gains in real income generated when compared to the scenario in which female entrepreneurs retain the advantage they have in hiring female workers (Section 6). This suggests that this advantage is significant; when eliminated, the positive effects discussed in Section 6 are smaller.

On the other hand, the effects remain large and positive, suggesting that even without affirmative action (i.e., policies that give women, or let them retain, an advantage in some areas), policies targeted at achieving gender-parity can generate substantial benefits for the economy. Interestingly however, in our case, affirmative action provides additional gains, not only to the women whom it explicitly targets, but also to the economy in the aggregate.

8 Conclusion

Our analysis demonstrates that eliminating excess barriers to entrepreneurship facing women is beneficial not only to women, but to the entire economy. But it does 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 identified based on the data patterns in the Census data related to entrepreneurship. Further research needs to relate the estimated wedges to actual policies to assess which interventions are most effective. The main challenge is that several of these barriers are not due to legal constraints, but to norms and attitudes, which are more difficult to measure. This challenge notwithstanding, our work has two main policy-relevant messages: First, absent a comprehensive approach towards eliminating *all* gender distortions in the labor market, policies focused exclusively on increasing female LFP may have unintended adverse effects on female wages and profits of female entrepreneurs; complementing such policies with measures supporting female entrepreneurship ensures that the additional supply of women on the labor market is met with additional demand, and results in larger benefits for women. Second, interventions aimed at supporting female entrepreneurship will be more effective if they target the intensive margin (i.e., support existing female-owned enterprises) than the extensive margin (i.e., encourage new entry of female entrepreneurs).

Testing and implementing policy interventions at scale requires not only studying their implications for the labor force participation and entrepreneurial decisions of the women they directly target, but also assessing their impact on the labor supply decisions of all men and women, along with the resulting changes in wages and prices in equilibrium. In this regard, our analysis can prove helpful. Combining case studies of specific interventions to empower women with our framework can be a fruitful approach towards identifying the most promising policies in equilibrium.

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.

Angrist, Noam, Simeon Djankov, Pinelopi K Goldberg, and Harry A Patrinos, "Measuring human capital using global learning data," *Nature*, 2021, *592* (7854), 403–408.

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

__, Oriana Bandiera, Virginia Minni, and Victor Quintas-Martinez, "Gender Roles and the Misallocation of Labor Across Countries," *Working Paper*, 2022.

Banerjee, Abhijit V. and Andrew F. Newman, "Occupational Choice and The Process of Development," *Journal of Political Economy*, 1993, *101* (2), 274–298.

Bansal, Samarth, "Indian States and Women: Where are They Empowered, Where are They Not," *Hindustan Times. https://www. hindustantimes. com/interactives/women-empowerment-index*, 2017.

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

Bento, Pedro, "Female Entrepreneurship in the US 1982-2012: Implications for Welfare and Aggregate Output," *Working Paper*, 2020.

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

Bick, Alexander, Nicola Fuchs-Schündeln, David Lagakos, and Hitoshi Tsujiyama, "Structural change in labor supply and cross-country differences in hours worked," *Journal of Monetary Economics*, 2022, *130*, 68–85.

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.

Chaudhuri, Kausik, Subash Sasidharan, and Rajesh Seethamma Natarajan Raj, "Gender, small firm ownership, and credit access: some insights from India," *Small Business Economics*, 2020, *54* (4), 1165–1181.

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.

Desai, Sonalde, Reeve Vanneman, and NCAER, "India Human Developement Survey (IHDS)," *Inter-University Consortium for Political and Social Research*, 2005.

Deshpande, Ashwini and Naila Kabeer, "(In) Visibility, Care and Cultural Barriers: The Size and Shape of Women's Work in India," 2019.

Dix-Carneiro, Rafael, Pinelopi K Goldberg, Costas Meghir, and Gabriel Ulyssea, "Trade and Informality in the Presence of Labor Market Frictions and Regulations," Technical Report, National Bureau of Economic Research 2021.

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.

Ghosh, Pallab K, "A Study of the Long-run Substitutability Between Men and Women," 2018.

Giorgi, Giacomo De, Marco Paccagnella, and Michele Pellizzari, "Gender Complementarities in the Labor Market," *Bank of Italy Occasional Paper*, 2013, (183).

Gruber, Siegfried and Mikołaj Szołtysek, "The Patriarchy Index: A Comparative Study of Power Relations Across Historical Europe," *The History of the Family*, 2016, *21* (2), 133–174.

Hamermesh, Daniel S, *Labor Demand*, Princeton University Press, 1996.

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.

Jayachandran, Seema, "Microentrepreneurship in developing countries," *Handbook of Labor, Human Resources and Population Economics*, 2021, pp. 1–31.

Jensen, Anders, "Employment structure and the rise of the modern tax system," *American Economic Review*, 2022, *112* (1), 213–234.

Johnson, Matthew and Michael P Keane, "A Dynamic Equilibrium Model of the US Wage Structure, 1968–1996," *Journal of Labor Economics*, 2013, *31* (1), 1–49.

Kaboski, Joseph P and Robert M Townsend, "A structural evaluation of a large-scale quasi-experimental microfinance initiative," *Econometrica*, 2011, *79* (5), 1357–1406.

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

Morazzoni, Marta and Andrea Sy, "Female entrepreneurship, financial frictions and capital misallocation in the US," *Journal of Monetary Economics*, 2022, *129*, 93–118.

Ngai, L Rachel and Barbara Petrongolo, "Gender Gaps and the Rise of the Service Economy," *American Economic Journal: Macroeconomics*, 2017, *9* (4), 1–44.

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.

Olivetti, Claudia and Barbara Petrongolo, "Gender Gaps Across Countries and Skills: Demand, Supply and the Industry Structure," *Review of Economic Dynamics*, 2014, *17* (4), 842–859.

— **and** — , "The Evolution of Gender Gaps in Industrialized Countries," *Annual Review of Economics*, 2016, *8*, 405–434.

Pitt, Mark M, Mark R Rosenzweig, and Mohammad Nazmul Hassan, "Human Capital Investment and the Gender Division of Labor in a Brawn-based Economy," *American Economic Review*, 2012, *102* (7), 3531–60.

Plan International, "Plan for Every Child Gender Vulnerability Index Report-1," 2017. https://smartnet.niua.org/sites/default/files/resources/gender-vulnerability-index.pdf.

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

Raj, Rajesh SN and Subash Sasidharan, *Small Firm Ownership and Credit Constraints in India*, Routledge India, 2020.

Ranasinghe, Ashantha, "Misallocation across Establishment Gender," *University of Alberta Department of Economics Working Paper*, 2021, (2021-02).

Rao, Nitya, Arjan Verschoor, Ashwini Deshpande, and Amaresh Dubey, "Gender Caste and Growth Assessment–India," *Report submitted to DFID. http://www.uea.ac.uk/dev/publications/RPP8*, 2008.

Rendall, Michelle, "Structural Change in Developing Countries: Has it Decreased Gender Inequality?," *World Development*, 2013, *45*, 1–16.

Restuccia, Diego and Richard Rogerson, "The Causes and Costs of Misallocation," *Journal of Economic Perspectives*, 2017, *31* (3), 151–74.

Reynolds, Paul D, Michael Hay, and S Michael Camp, "Global entrepreneurship monitor," *Kansas City, Missouri: Kauffman Center for Entrepreneurial Leadership*, 1999.

Roy, Andrew Donald, "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 1951, *3* (2), 135–146.

Schoar, Antoinette, "The divide between subsistence and transformational entrepreneurship," *Innovation policy and the economy*, 2010, *10* (1), 57–81.

Singh, Abhishek, Praveen Chokhandre, Ajeet Kumar Singh, Kathryn M Barker, Kaushalendra Kumar, Lotus McDougal, KS James, and Anita Raj, "Development of the India Patriarchy Index: Validation and Testing of Temporal and Spatial Patterning," *Social Indicators Research*, 2021, pp. 1–27.

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.

Weinberg, Bruce A, "Computer Use and the Demand for Female Workers," *ILR Review*, 2000, *53* (2), 290–308.

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

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

Tables

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, Self-Employed	12.68	21.14				
	(48.35%)	(51.26%)				
Male, Informal	11.58	15.83	3.29	3.02	0.10	0.10
	(44.13%)	(38.37%)	(2.83)	(2.12)	(0.21)	(0.22)
Male, Formal	0.08	0.14	77.31	67.54	0.21	0.25
	(0.31%)	(0.34%)	(440.9)	(166.58)	(0.25)	(0.3)
Female, Self-Employed	1.07	2.50				
	(4.07%)	(6.06%)				
Female, Informal	0.82	1.24	3.01	2.81	0.70	0.76
	(3.13%)	(3.04%)	(2.61)	(1.83)	(1.86)	(0.37)
Female, Formal	0.00	0.01	97.59	76.53	0.37	0.48
	(0.01%)	(0.02%)	(1197.03)	(130.34)	(0.33)	(0.40)
Total	26.23	40.86				

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: 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	-0.0167	-0.0346***	0.304***	0.298***
	(0.0175)	(0.00485)	(0.0126)	(0.0111)
Formal	2.348***	2.536***	0.0904***	0.0970***
	(0.0364)	(0.0332)	(0.00951)	(0.00990)
Female × Formal	0.135*	0.196***	-0.180***	-0.111***
	(0.0689)	(0.0452)	(0.0231)	(0.0176)
R^2	0.212	0.280	0.328	0.301
Panel B: With Industry Fixed Effects				
Female	-0.00962	-0.0435***	0.232***	0.235***
	(0.0135)	(0.00642)	(0.00953)	(0.00786)
Formal	2.079***	2.385***	0.0520***	0.0692***
	(0.0347)	(0.0361)	(0.00831)	(0.00885)
Female × Formal	0.170**	0.184***	-0.120***	-0.0676***
	(0.0672)	(0.0480)	(0.0191)	(0.0164)
R^2	0.338	0.344	0.472	0.404
N	12.48m	17.22m	12.48m	17.22m
Male, Informal	1.007	0.970	0.190	0.205
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. All regressions control for district fixed effects, along with whether the firm has access to power, dummy variables for different forms of financial access, 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.

Table 3: List of Parameters

Parameter	Details	Targeted Moments
α	Share of each industry in consumer demand	Share of firm sales in industry j as a frac. of the economy
ρ	Returns to scale in production	Avg. labor share in sales
γ	EoS b/w male and female workers	Set to 2.1 from the literature
t_{jr}	Tax in formal sector	Avg. sales tax in ASI
λ_j	Size-based penalty of operating in the informal sector	Ratio of avg. firm size of informal and formal male-owned firms
T_{sjr}	Aggregate production technology	Share of male-owned formal firms across industries in each sector
A_{sjr}	Female (relative to male) worker productivity	Ratio of female-male workers in male-owned firms across industries
$\{\sigma_x^2, \theta_g\}$	Variance of the productivity distribution	Variance of male & female firm size in the formal sector
$\overline{u}_g$	Disutility of LFP	Gender-specific LFP rate
$\{E_w, E_I, E_R\}$	Fixed costs of entrepreneurship and formalization	No. of entrepreneurs in the formal & informal sector as a frac. of the labor force
τ_{gs}	Hiring barriers	Ratio of avg. firm size of female-owned to male-owned firms
τ_{gs}^f	Hiring barriers	Ratio of avg. female-male workers in female-owned to male-owned firms

Table 4: Parameter Values

	1998			2005		
	Agri.	Manf.	Services	Agri.	Manf.	Services
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Parameter values that vary by industry						
T_I	0.51	0.72	1.00	0.51	0.62	1.00
T_F	0.43	1.06	1.00	0.44	0.74	1.00
A_I	1.03	1.03	1.00	1.04	1.03	1.00
A_F	1.01	0.94	1.00	0.94	0.88	1.00
λ	0.97	0.97	0.92	0.89	0.92	0.93
Panel B: Ability distribution parameters						
$\widetilde{\theta}_{mI}$	2.19			2.44		
$\widetilde{\theta}_{fI}$	2.22			2.74		
$\widetilde{\theta}_{mF}$	2.02			2.14		
$\widetilde{\theta}_{fF}$	2.04			2.15		
σ_m	0.30			0.31		

Notes: Each of the first three rows in Panel A reports the average values for the parameter across regions. The parameter λ varies only by industry. Parameters in Panel B do not vary by industry or regions, and hence only the values for each year are reported in columns (1) and (4) for 1998 and 2005 respectively.

Table 5: Estimates for Hiring Distortions

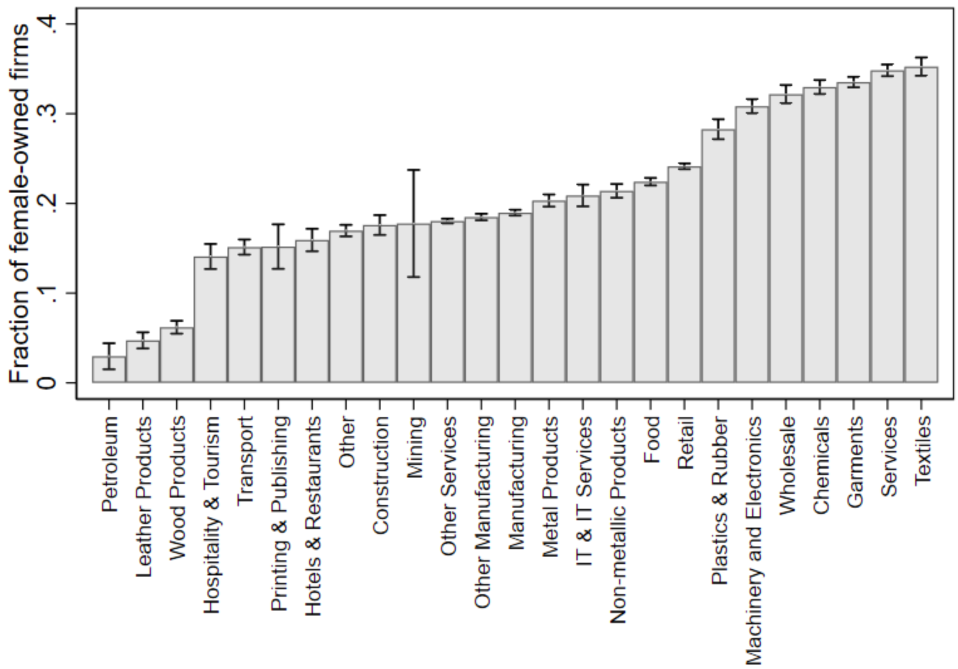
	1998	2005	(2)-(1)
	(1)	(2)	(3)
$1+\tau_{fI}$	1.24	1.18	-0.06
	[0.12]	[0.08]	
$1+\tau_{fF}$	1.07	1.14	0.07
	[0.38]	[0.19]	
$1+\tau_{fI}^{f}$	0.96	0.95	-0.01
	[0.04]	[0.03]	
$1+\tau_{fF}^{f}$	0.99	1.00	0.01
	[0.20]	[0.25]	

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 value for 1998 and 2005 respectively. Column (3) reports the difference between columns (2) and (1).

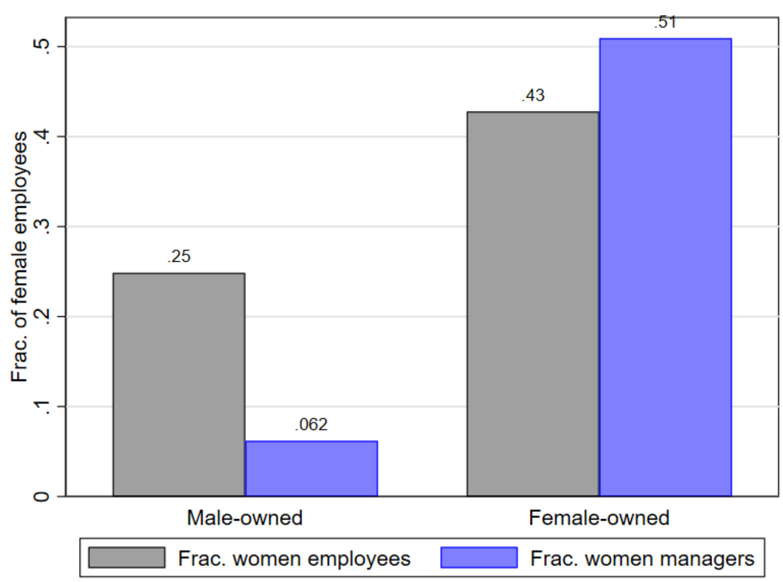
Figures

Figure 1: Share of Women Entrepreneurs, Employees and Managers

(a) Fraction of female entrepreneurs across industries

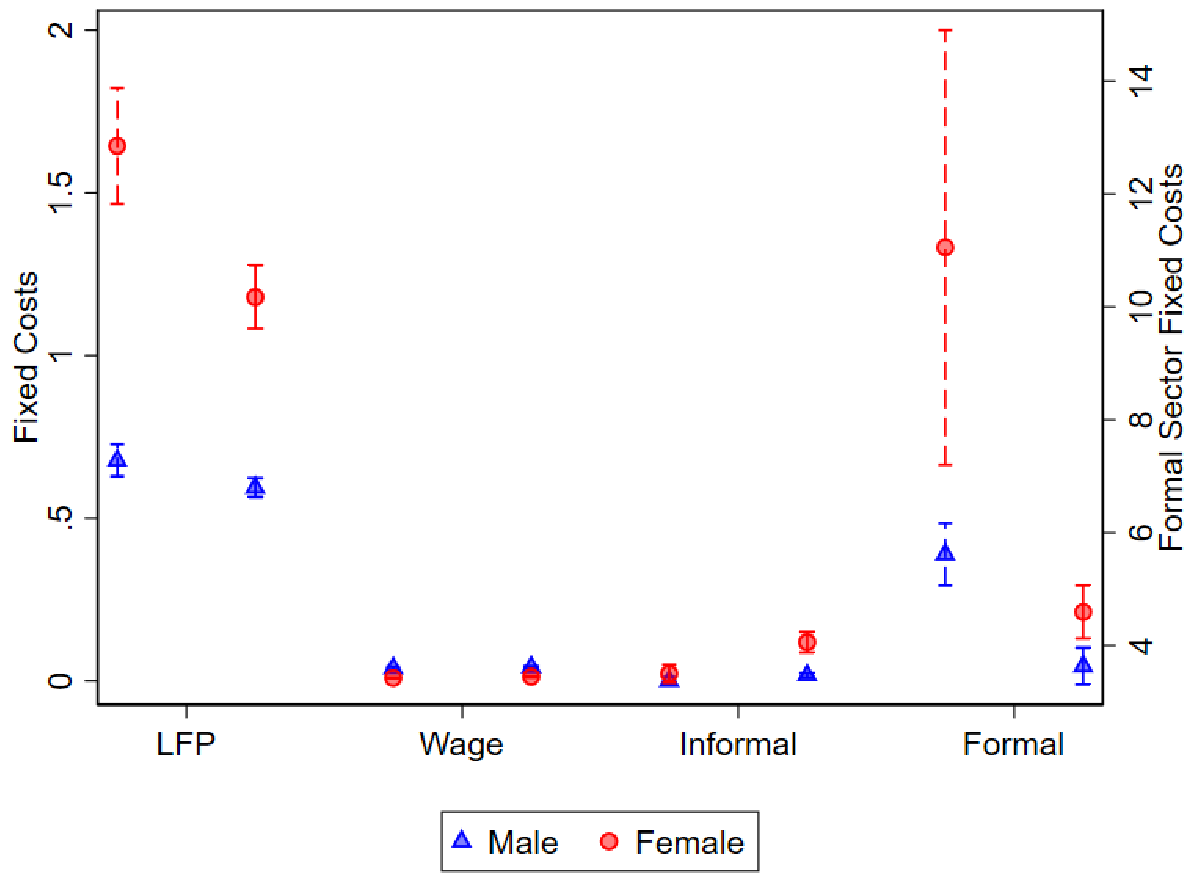


(b) Fraction of women employees and managers



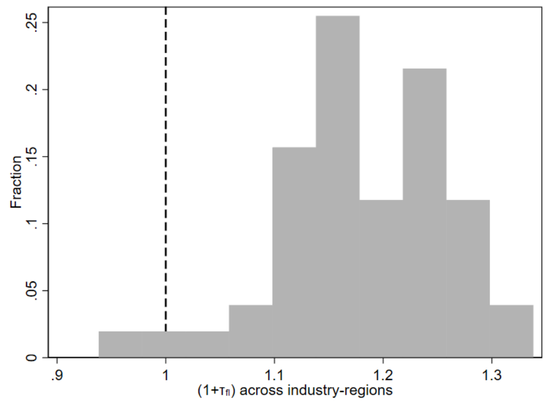
Notes: Both figures use the World Bank Enterprise Surveys. Figure 1(a) plots the average fraction of female-owned firms across 25 sectors. Figure 1(b) plots the fraction of women employees and the probability that the top manager in a firm is a female.

Figure 2: Fixed Costs of LFP, Wage-Employment, and Entrepreneurship

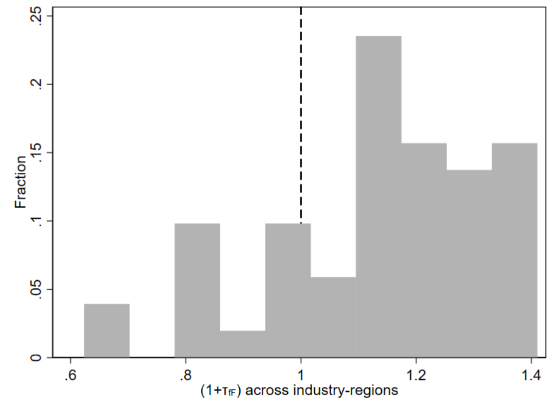


Notes: The above figure shows the estimates for the fixed costs to participate in the labor force (LFP), wage employment, informal and formal entrepreneurship. For each parameter, the estimates for 1998 (2005) are reported by the dash (solid) lines, with a 1 standard error range around the mean. Blue triangles and red triangles are the estimates for men and women respectively.

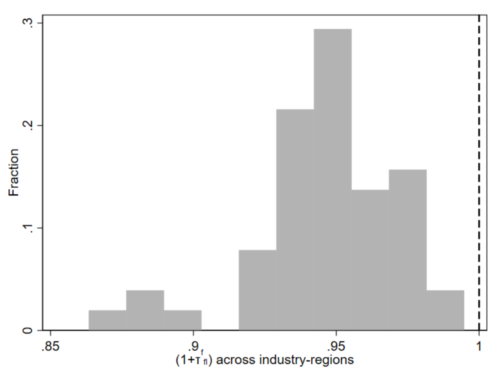
Figure 3: Hiring Barriers in the Formal and Informal Sectors



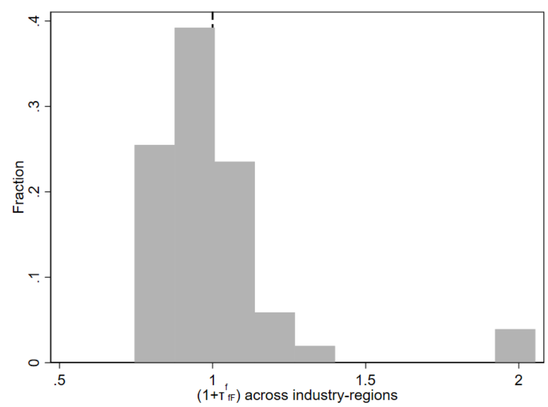
(a) $1 + \tau_{fI}$



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



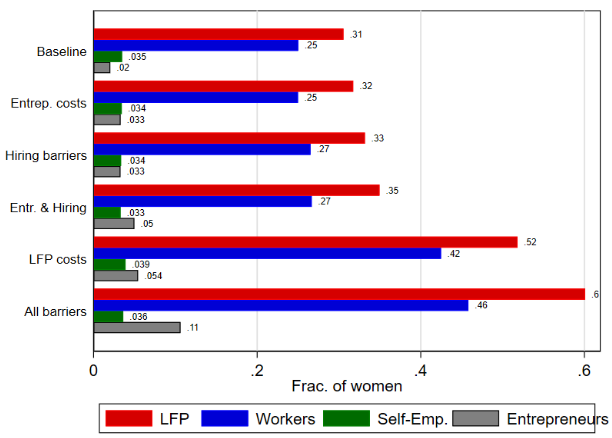
(c) $1 + \tau^f_{fI}$



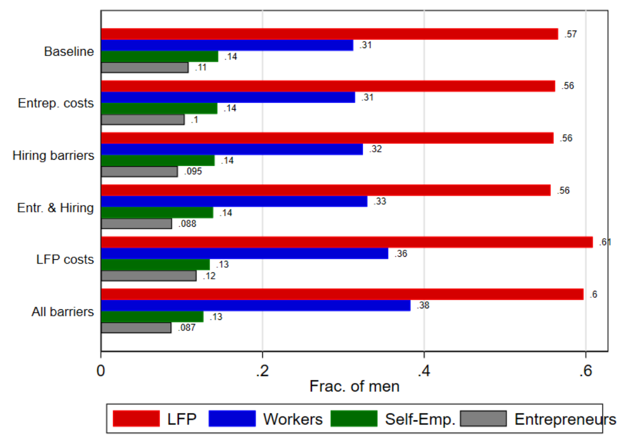
(d) $1 + \tau^f_{fF}$

<u>Notes:</u> Figures (a)-(b) plot the distribution of hiring barriers faced by women entrepreneurs (relative to men) across regions and industries in the informal and formal sectors in 2005 i.e., $1 + \tau_{fs}$. Figures (c)-(d) plot the distribution of barriers faced by women entrepreneurs (relative to male entrepreneurs) in hirng female workers (relative to male workers) i.e., $1 + \tau^f_{fs}$.

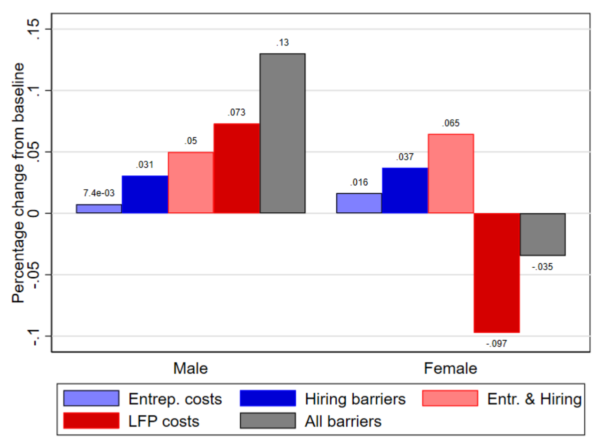
Figure 4: Impact of Affirmative Action Policies on FLFP, Entrepreneurship and Earnings



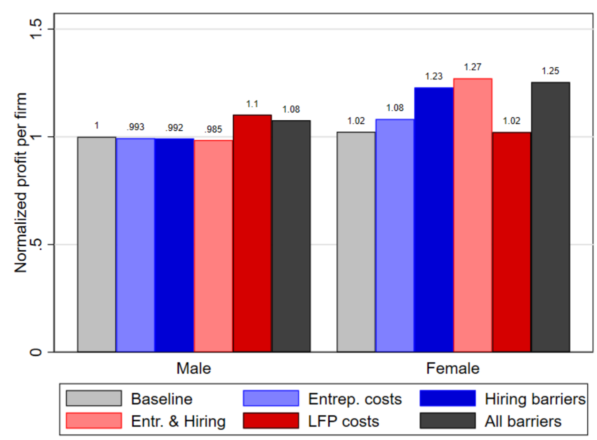
(a) Distribution of Women in the Economy



(b) Distribution of Men in the Economy



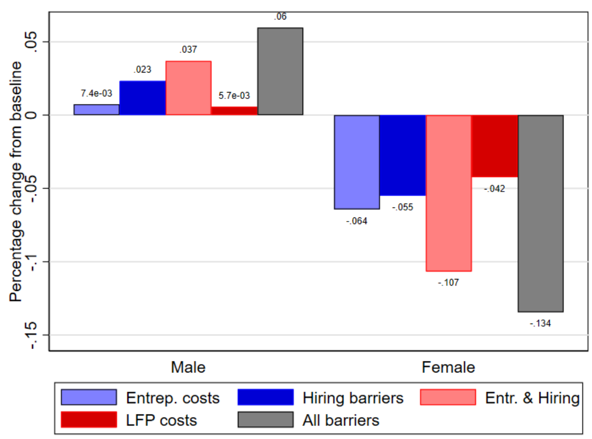
(c) Real Wages



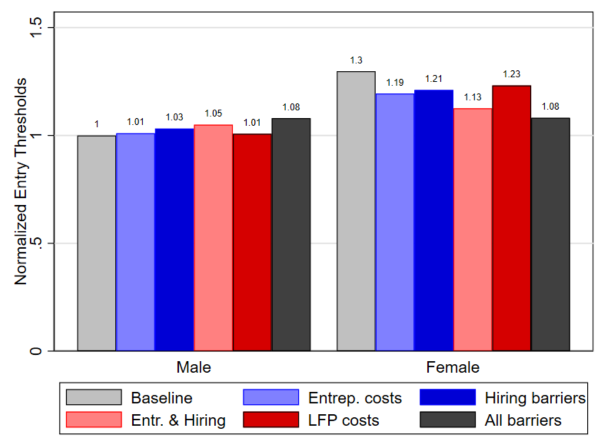
(d) Real Profits per Firm

Notes: The above figures report the impact of five affirmative action policies described in Section 6 of the paper. For each policy, Figures (a) and (b) show the distribution of women and men in the labor force, wage- and self-employment, and entrepreneurship. Figures (c) and (d) show the effect of each policy on gender-specific real wages and real average profits for workers and entrepreneurs respectively.

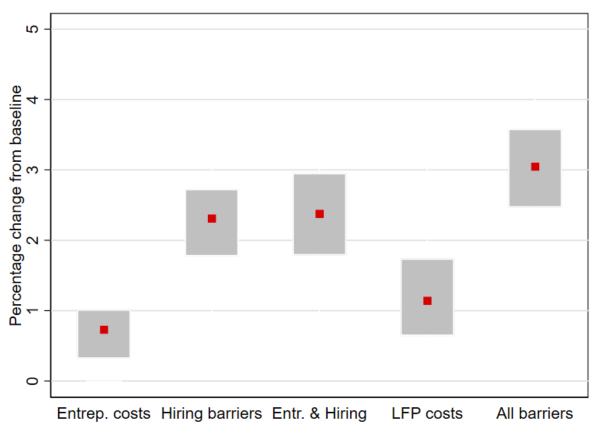
Figure 5: Impact of Affirmative Action Policies on Productivity and Welfare



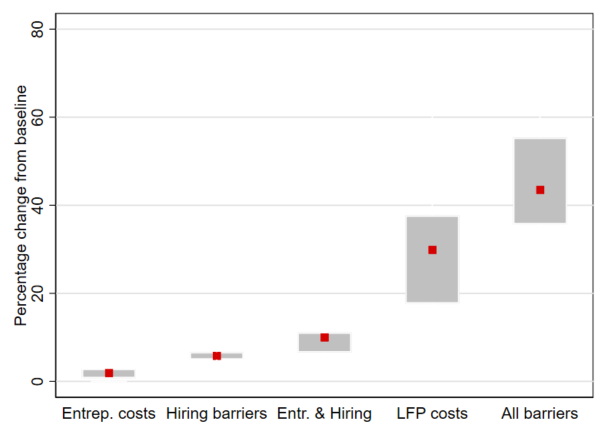
(a) Average Ability of Entrepreneurs Relative to Baseline



(b) Ability of the Marginal Entrepreneur



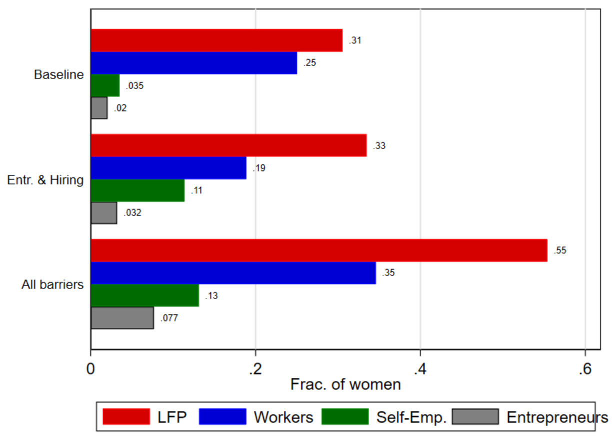
(c) Change in Aggregate Productivity



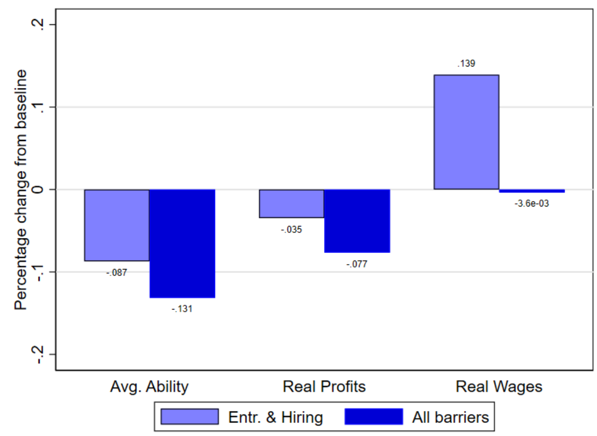
(d) Change in Real Income

Notes: The above figures report the impact of five affirmative action policies described in Section 6 of the paper. For each policy, Figure (a) reports the change in the average ability of an entrepreneur relative to the baseline scenario. Figure (b) reports the ability of the marginal entrepreneur. Figures (c)-(d) report the changes in aggregate productivity and real income across the economy as compared to the baseline.

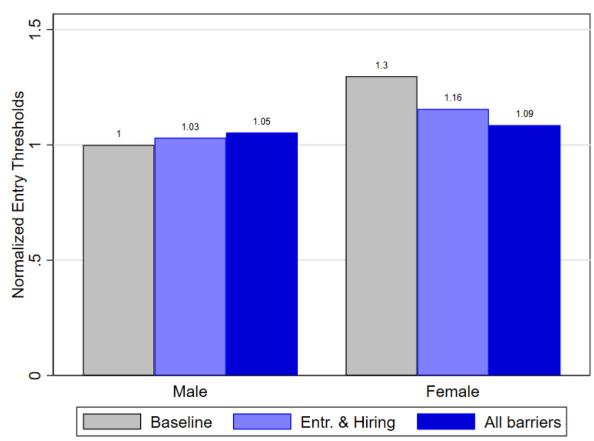
Figure 6: Impact of Policies That Remove All Gender-Based Frictions



(a) Distribution of Females



(b) Changes in Real Profits and Wages for Women



(c) Ability of Marginal Entrepreneur

(d) Change in Agg. Productivity and Real Income

Notes: The above figures report the impact of policies described in Section 7 of the paper. For each policy, Figures (a) and (b) report the ability of the average and marginal entrepreneur respectively. Figures (c)-(d) show the changes in aggregate productivity and real income across the economy as compared to the baseline.

ONLINE APPENDIX: NOT FOR PUBLICATION

A Additional Tables and Figures

Table A1: Composition across Gender and Sectors, Excluding Family-owned Firms

	Log(L)		Frac. female emp.	
	1998	2005	1998	2005
	(1)	(2)	(3)	(4)
Panel A: Without industry fixed effects				
Female	0.0484	-0.0473***	0.326***	0.331***
	(0.0449)	(0.00773)	(0.0225)	(0.0115)
Formal	2.200***	2.475***	0.120***	0.125***
	(0.0348)	(0.0334)	(0.00915)	(0.00988)
Female × Formal	0.0149	0.229***	-0.184***	-0.151***
	(0.0853)	(0.0444)	(0.0289)	(0.0162)
R^2	0.226	0.305	0.231	0.210
Panel B: With industry fixed effects				
Female	-0.00646	-0.0770***	0.264***	0.266***
	(0.0279)	(0.00811)	(0.0169)	(0.00811)
Formal	1.889***	2.306***	0.0763***	0.0941***
	(0.0303)	(0.0365)	(0.00757)	(0.00855)
Female × Formal	0.0815	0.250***	-0.145***	-0.116***
	(0.0632)	(0.0480)	(0.0231)	(0.0139)
R^2	0.378	0.378	0.368	0.294
N	5.23m	9.88m	5.23m	9.88m
Male, Informal	1.192	1.059	0.0855	0.126
Firm controls	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes

Notes: The sample is restricted to firms that are not "family-owned". Family-owned firms are defined as those firms where more than half the employees are not hired on wage contracts. 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. All regressions control for district fixed effects, along with whether the firm has access to power, dummy variables for different forms of financial access, 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.

Table A2: Derivatives of Moments to Parameters

Moment	A_I	A_F	τ_I^f	τ_F^f	τ_I	τ_F	λ
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Sample from the 1998 Round of the Economic Census							
$R_{mI,j}/R_{mI,Serv.}$	**0.67**	0.00	0.00	0.00	0.00	0.00	0.00
$R_{mFj}/R_{mFServ.}$	0.00	**0.66**	0.00	0.00	0.00	0.00	0.00
$R_{fI,j}/R_{fI,Serv.}$	0.00	0.00	**-2.18**	0.00	0.00	0.00	0.00
$R_{fF,j}/R_{fF,Serv.}$	0.00	0.00	0.00	**-2.25**	0.00	0.00	0.00
$\bar{l}_{fI,j}/\bar{l}_{mI,j}$	0.06	0.04	-0.85	0.07	**-1.34**	0.31	0.00
$\bar{l}_{fF,j}/\bar{l}_{mF,j}$	-0.13	0.14	-0.24	-0.31	-0.40	**-1.27**	-2.46
$\bar{l}_{mF,j}/\bar{l}_{mF,Serv.}$	-0.20	0.08	-0.01	0.00	0.03	0.09	**-3.30**
Panel B: Sample from the 2005 Round of the Economic Census							
$R_{mI,j}/R_{mI,Serv.}$	**0.67**	0.00	0.00	0.00	0.00	0.00	0.00
$R_{mFj}/R_{mFServ.}$	0.00	**0.65**	0.00	0.00	0.00	0.00	0.00
$R_{fI,j}/R_{fI,Serv.}$	0.00	0.00	**-2.19**	0.00	0.00	0.00	0.00
$R_{fF,j}/R_{fF,Serv.}$	0.00	0.00	0.00	**-2.26**	0.00	0.00	0.00
$\bar{l}_{fI,j}/\bar{l}_{mI,j}$	0.03	0.04	-0.87	0.10	**-1.39**	**0.19**	-1.02
$\bar{l}_{fF,j}/\bar{l}_{mF,j}$	-0.02	0.01	-0.58	0.01	-0.99	-0.01	-0.30
$\bar{l}_{mF,j}/\bar{l}_{mF,Serv.}$	-0.19	0.10	-0.01	0.01	0.03	0.02	**-1.83**

Notes: This table reports the derivatives of each moment with respect to each parameter. Each row is a moment calculated from the model simulation. Each number in the table indexed by row R and column C, is the percent change in the moment in row R, when a parameter in column C is increased by 1 p.p. The largest value in each column is bold faced. Panel A (B) reports the results from the 1998 (2005) Round of the Economic Census. R_{gsj} and $\bar{l}_{gsj}$ are the ratio of female-male workers and the average size of a firm owned by an entrepreneur of gender g in sector s and industry j.

Table A3: Model Fit I

	Male		Female	
	Data	Model	Data	Model
	(1)	(2)	(3)	(4)
Panel A: Occupational choice of individuals				
1-LFP	0.43	0.43	0.70	0.69
	(0.04)	(0.04)	(0.08)	(0.08)
Frac. Wage Emp.	0.31	0.31	0.25	0.25
	(0.04)	(0.04)	(0.07)	(0.07)
Frac. Self Emp.	0.15	0.14	0.03	0.03
	(0.02)	(0.02)	(0.03)	(0.03)
Frac. Inf. Entrp.	0.11	0.11	0.02	0.02
	(0.01)	(0.01)	(0.01)	(0.01)
Frac. Formal Entrp.	0.001	0.001	0.000	0.000
	(0.0005)	(0.0005)	(0.0001)	(0.0001)
Panel B: Ratio of female-male workers in a firm				
Informal	0.95	0.95	1.07	1.06
	(0.06)	(0.06)	(0.08)	(0.08)
Formal	0.77	0.77	0.87	0.87
	(0.15)	(0.15)	(0.36)	(0.36)

Notes: Each row reports the average value across regions with the standard deviation in parentheses. Columns (1)-(2) report the moments for men, while (3)-(4) report those for women. Columns (1) and (3) report the moments in the Data, while (2) and (4) report their simulated counterparts from the Model. Panel A reports the allocation of men/women in the economy with the fraction of individuals who are (a) not in the labor force; (ii) in wage employment; (iii) informal entrepreneurship and (iv) formal entrepreneurship. Panel B reports the ratio of female to male workers in an informal and formal male-owned (Coumns 1-2) and female-owned firm (Columns 3-4).

Table A4: Model Fit II

	Male		Female	
	Data	Model	Data	Model
	(1)	(2)	(3)	(4)
Panel A: Ratio of average firm size				
$\bar{l}_{gI}/\bar{l}_{mI}$	1.00	1.00	1.06	1.04
	(0)	(0)	(0.18)	(0.17)
$\bar{l}_{gF}/\bar{l}_{mF}$	1.00	1.00	1.18	1.05
	(0)	(0)	(0.62)	(0.29)
$\bar{l}_{gF}/\bar{l}_{gI}$	22.69	28.70	26.15	28.66
	(9.39)	(7.55)	(20.64)	(8.99)
Panel B: Average firm size				
Informal	4.21	6.83	4.37	7.11
	(0.70)	(0.88)	(0.40)	(1.39)
Formal	95.05	193.54	113.38	199.02
	(41.85)	(45.90)	(87.40)	(59.45)
Panel C: Std. Deviation of firm size				
Informal	3.60	3.63	3.58	3.35
	(1.35)	(1.23)	(1.16)	(1.55)
Formal	184.70	191.89	160.68	200.95
	(108.70)	(92.96)	(172.76)	(102.24)

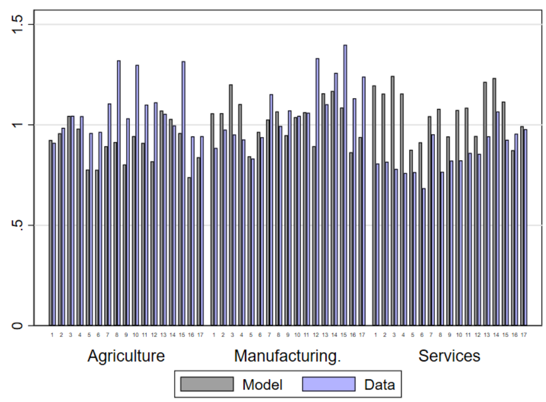
Notes: Each row reports the average value across regions with the standard deviation in parentheses. Columns (1)-(2) report the moments for men, while (3)-(4) report those for women. Columns (1) and (3) report the moments in the Data, while (2) and (4) report their simulated counterparts from the Model. Panel A reports the ratio of the average firm size for: (i) firms of gender g relative to male-owned firms in the informal sector; (ii) firms of gender g relative to male-owned firms in the formal sector and (iii) firms of gender g in the formal relative to the informal sector. Panel B reports the average firm-size in the informal and formal sector and Panel C reports the standard deviation for those firms.

Table A5: Model Fit III

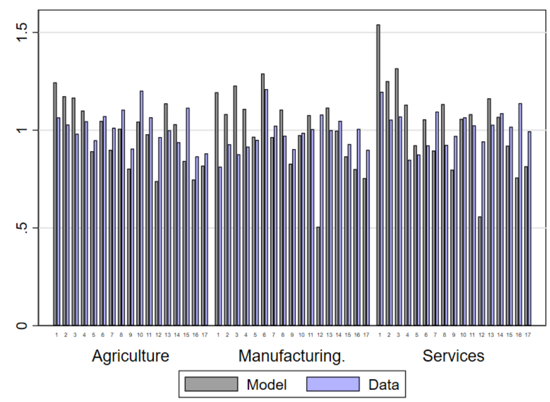
	Male		Female	
	Data	Model	Data	Model
	(1)	(2)	(3)	(4)
Panel A: Share of Firms in the Informal Sector				
Agriculture	0.15	0.16	0.25	0.15
	(0.08)	(0.09)	(0.22)	(0.17)
Manf.	0.26	0.27	0.33	0.22
	(0.06)	(0.07)	(0.16)	(0.11)
Services	0.59	0.56	0.43	0.62
	(0.07)	(0.08)	(0.17)	(0.14)
Panel B: Share of Firms in the Formal Sector				
Agriculture	0.06	0.06	0.15	0.24
	(0.05)	(0.05)	(0.14)	(0.2)
Manf.	0.58	0.59	0.35	0.37
	(0.12)	(0.11)	(0.12)	(0.17)
Services	0.37	0.35	0.50	0.38
	(0.11)	(0.1)	(0.12)	(0.16)

Notes: Each row reports the average value across regions with the standard deviation in parentheses. Columns (1)-(2) report the moments for men, while (3)-(4) report those for women. Columns (1) and (3) report the moments in the Data, while (2) and (4) report their simulated counterparts from the Model. Panel A reports the allocation of men/women in the economy with the fraction of individuals who are (a) not in the labor force; (ii) in wage employment; (iii) informal entrepreneurship and (iv) formal entrepreneurship. Panel B reports the ratio of female to male workers in an informal and formal male-owned (Coumns 1-2) and female-owned firm (Columns 3-4).

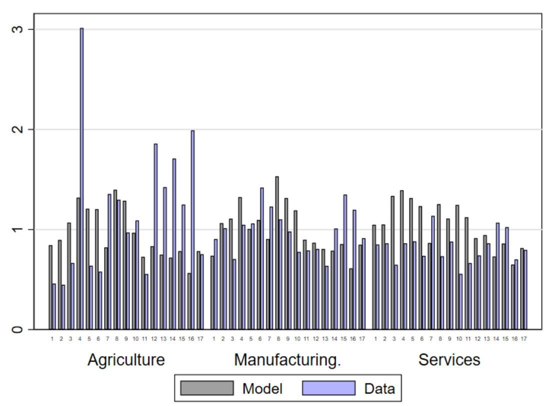
Figure A1: Model Fit: Average Firm Size



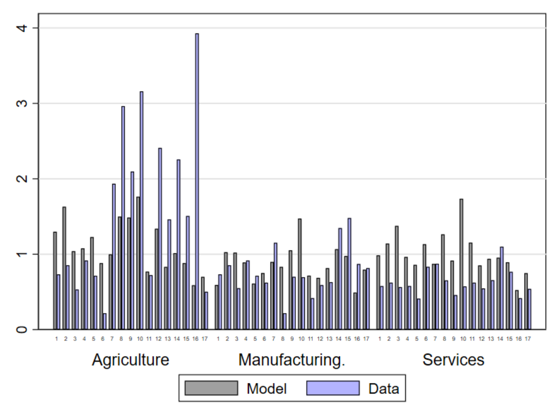
(a) Avg. Firm Size: Male, Informal



(b) Avg. Firm Size: Female, Informal



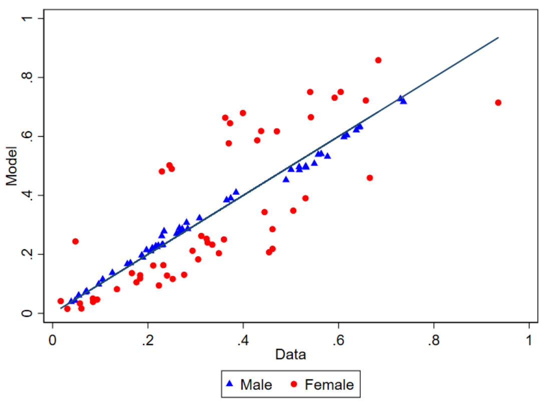
(c) Avg. Firm Size: Male, Formal



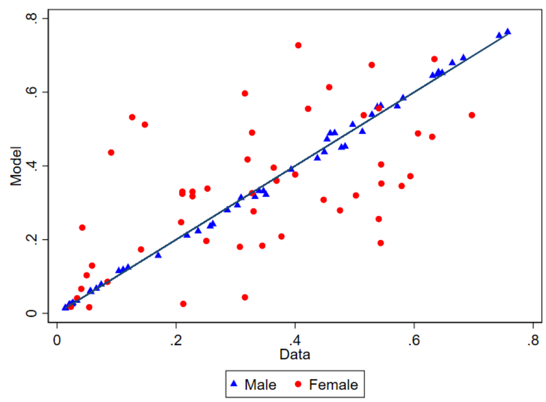
(d) Avg. Firm Size: Female, Formal

Notes: Figures (a)-(b) plot the distribution of hiring barriers faced by women entrepreneurs (relative to men) across regions and industries in the informal and formal sectors in 2005 i.e., $1 - \tau_{fs}$. Figures (c)-(d) plot the distribution of barriers faced by women entrepreneurs (relative to male entrepreneurs) in hirng female workers (relative to male workers) i.e., i.e., $1 - \tau_{fs}^{f}$.

Figure A2: Model Fit: Share of Firms in Each Industry



(a) Informal Sector



(b) Formal Sector

Notes: Figures (a)-(b) plot the distribution of hiring barriers faced by women entrepreneurs (relative to men) across regions and industries in the informal and formal sectors in 2005 i.e., $1 - \tau_{fs}$. Figures (c)-(d) plot the distribution of barriers faced by women entrepreneurs (relative to male entrepreneurs) in hirng female workers (relative to male workers) i.e., i.e., $1 - \tau_{fs}^{f}$.

B Mathematical Proofs

B.1 Incumbent Firm Decisions

The problem of a firm with productivity z in a sector s (dropping gender and industry for notational ease) is given by:

$$\max \; p_s z l^{\rho_s} - \left[w^m l^m + w^f l^f \right]$$

where $\{\rho_I, \rho_F\} = \{\lambda\rho, \rho\}$ and $\{p_I, p_F\} = \{p, (1-t)p\}$. Define:

$$w = \left[\sum_g A^g (w^g)^{(1-\gamma)} \right]^{\frac{1}{1-\gamma}}$$

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

$$l_I^*(z) = \left[\rho_s \times \frac{z}{w/p_s} \right]^{\frac{1}{1-\rho_s}}$$
$$\pi_I^*(z) = \frac{1-\rho_s}{\rho_s} \times w l_I^*(z)$$

Cost minimization in the second stage implies:

$$\begin{aligned} &\min \; w^m l^m + w^f l^f \\ &\text{s.t.} \; \left[\sum_g A^g (l^g)^{\frac{\gamma-1}{\gamma}} \right]^{\frac{\gamma}{\gamma-1}} = l_I^* \end{aligned}$$

Taking the first order conditions and rearranging, we get:

$$w^g l^g(z) = A^g \left(\frac{w^g}{w} \right)^{1-\gamma} \times w l^*(z)$$

B.2 Penalty of Operating in the Informal Sector

An alternative way to present the model is to allow for a size-dependent penalty of operating in the informal sector. Let $\tau(l)$ be the penalty function such that $\tau(l) > 0$, $\tau'(l) < 0$ and $\tau(\infty) \to 0$. One can think of $t_I(l)$ as a per-unit size-dependent tax of operating in the informal sector, such that $\tau(l) = 1 - t_I(l)$. Accordingly, the maximization problem of the firm can be written as:

$$\max_{l} \tau(l) p z l^{\rho} - wl$$

Taking the first order condition and rearranging:

$$\left[\rho \tau(l) + l \tau'(l)\right] p z l^{\rho} = wl \tag{10}$$

Compared to the baseline model, we have:

$$\left[\widetilde{\rho} \times l^{\tilde{\rho}-\rho}\right] p z l^{\rho} = wl \tag{11}$$

Equations (10) and (11) are therefore connected through the $\tau(l)$ function, so that:

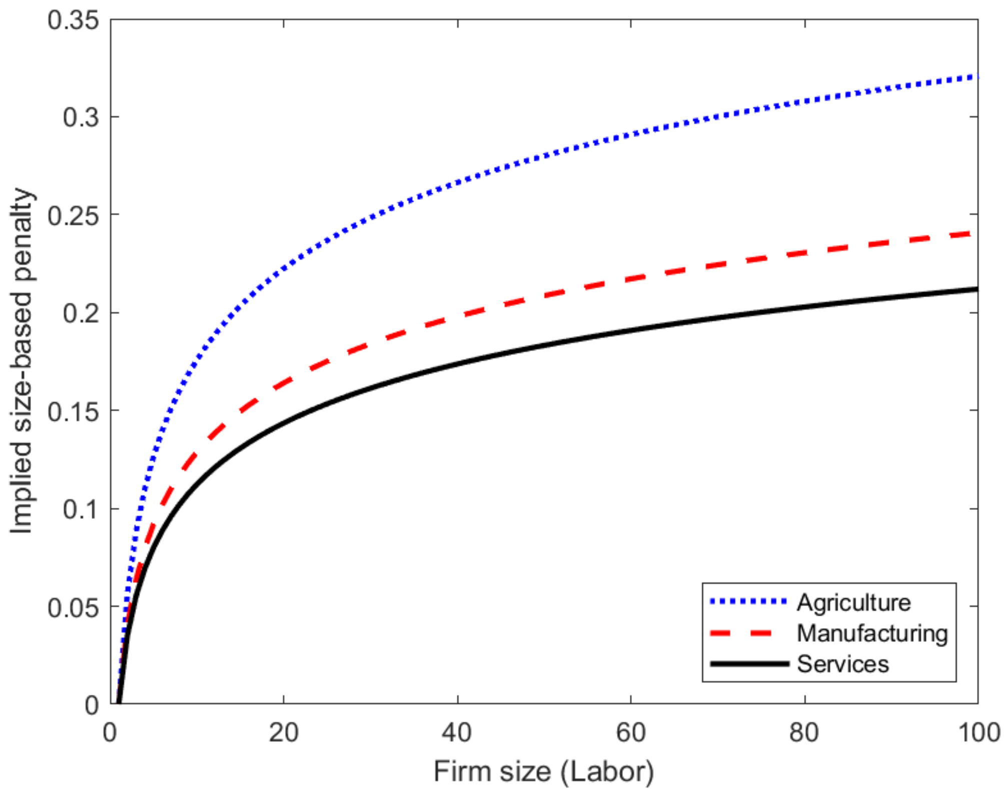
$$\rho \tau(l) + l \tau'(l) = \widetilde{\rho} \times l^{\tilde{\rho}-\rho} \tag{12}$$

This is a differential equation of the form $ay + xdy/dx = bx^c$, where $y = f(x)$. This has a general solution of the form $y = \frac{bx^c}{a+c} + \frac{k}{x^a}$ where k is an integration constant. Therefore the general solution to $\tau(l)$ is given by:

$$\tau(l) = \left[l^{\tilde{\rho}} + k\right] l^{-\rho} \tag{13}$$

To restrict $0 < \tau(l) < 1$, we assume $k = 0$ and plot $t_I(l) = 1 - \tau(l)$ in Figure B1.

Figure B1: Size-based Penalty Function



Notes: The above graph plots the size-based penalty function of operating in the informal sector as a function of firm size.

B.3 Allocation of entrepreneurs across industries

From Equations (2), (3), (5) and (6), the general form of the profit function and wage bill for a firm in sector s (dropping g for notational convenience) is given by:

$$b_{sj} \equiv \frac{w_{sj} l_{sj}}{T_{sj}} = \eta_{L,sj} \times \varepsilon^{\frac{1}{1-\rho_s}}$$

$$\pi_{sj} = \eta_{\pi,sj} \times \varepsilon^{\frac{1}{1-\rho_s}}$$

where:

$$\eta_{L,sj} = \frac{w_{sj}}{T_j} \left[\rho_s \frac{T_j}{w_{sj}/p_{sj}} \times x \right]^{\frac{1}{1-\rho_s}}$$

$$\eta_{\pi,sj} = \frac{1-\rho_s}{\rho_s} \times \eta_{L,sj}$$

Let $\widetilde{\theta}_s = \theta(1-\rho_s)$. Dropping s for notational ease, the distribution of π_j within a sector s will follow a Frechet distribution given by $\pi_j \sim Frechet(\widetilde{\theta}, \eta_{\pi,j})$ with a CDF given by:

$$F(\pi) = exp\left[-\left(\frac{\pi}{\eta_\pi} \right)^{-\widetilde{\theta}} \right]$$

Note that the share of firms in an industry k will be the probability that the profits in industry k are higher than in all other industries. This implies that:

$$\begin{aligned}
\varphi_k &= Pr(\pi_k = \max\{\pi_j\}_{\forall j}) \\
&= \int \prod_{j \neq k} F(\pi_k) \times dF(\pi_k) d\pi_k \\
&= \int \prod_{j \neq k} e^{-(\pi_k/\eta_{\pi,j})^{-\widetilde{\theta}}} \times e^{-(\pi_k/\eta_{\pi,k})^{-\widetilde{\theta}}} \times \widetilde{\theta} \eta_{\pi,k}^{\widetilde{\theta}} \times \pi_k^{-\widetilde{\theta}-1} d\pi_k \\
&= \int e^{-(\sum_j \eta_{\pi,j}^{\widetilde{\theta}}) \pi_k^{-\widetilde{\theta}}} \times \widetilde{\theta} \eta_{\pi,k}^{\widetilde{\theta}} \times \pi_k^{-\widetilde{\theta}-1} dx \\
&= \frac{\eta_{\pi,k}^{\widetilde{\theta}}}{\sum \eta_{\pi,j}^{\widetilde{\theta}}} \times \int \underbrace{e^{-\sum \eta_{\pi,j}^{\widetilde{\theta}} \pi_k^{-\widetilde{\theta}}} \times \widetilde{\theta} (\sum \eta_{\pi,k}^{\widetilde{\theta}}) \pi_k^{-\widetilde{\theta}-1}}_{\text{Frechet distribution}} dx \\
&= \frac{\eta_{\pi,k}^{\widetilde{\theta}}}{\sum_j \eta_{\pi,j}^{\widetilde{\theta}}}
\end{aligned}$$

Substituting the values in the expression above, we have:

$$\eta_{\pi,j} = \frac{1-\rho_s}{\rho_s} \times \frac{w_{sj}}{T_j}\left[\rho_s \frac{p_{sj}}{w_{sj}/T_j} \times x\right]^{\frac{1}{1-\rho_s}}$$
$$= \left\{\frac{1-\rho_s}{\rho_s} \times (\rho_s x)^{\frac{1}{1-\rho_s}}\right\} \times \left[\frac{p_{sj}}{(w_{sj}/T_j)^{\rho_s}}\right]^{\frac{1}{1-\rho_s}}$$
$$\Rightarrow \frac{\eta_{\pi,j}^{\widetilde{\theta}}}{\sum_j \eta_{\pi,k}^{\widetilde{\theta}}} = \frac{\left[\frac{p_{sj}}{(w_{sj}/T_j)^{\rho_s}}\right]^{\theta}}{\sum_k \left[\frac{p_{sk}}{(w_{sk}/T_k)^{\rho_s}}\right]^{\theta}}$$

Note that since $\pi_k \sim Frechet(\widetilde{\theta}, \eta_{\pi,k})$, the distribution of maximum profits $\pi_j = \max\{\pi_k\}_j$ will also follow a Frechet distribution where $\pi_j \sim Frechet(\widetilde{\theta}, (\sum \eta_{\pi,k}^{\widetilde{\theta}})^{1/\widetilde{\theta}})$, so that:

$$E(\pi_j | \pi_j = \max\{\pi_k\}_{\forall k}) = (\sum \eta_{\pi,k}^{\widetilde{\theta}})^{1/\widetilde{\theta}} \Gamma_{\widetilde{\theta}}$$
$$= \Gamma_{\widetilde{\theta}} \times \varphi_j^{-1/\widetilde{\theta}} \eta_{\pi,j}$$

where $\Gamma_{\widetilde{\theta}} = \Gamma(1 - 1/\widetilde{\theta})$. Lastly, turning to the wage bill (b_j), note that similar to profits, $b_k \sim Frechet(\widetilde{\theta}, \eta_{L,k})$. Note that since $\pi_k = (\frac{1-\rho}{\rho})b_k$, $\pi_j = \max\{\pi_k\}_{\forall k}$ implies that $b_j = \max\{b_k\}_{\forall k}$. This implies that similar to the profits above,

$$E(b_j | \pi_j = \max\{\pi_k\}_{\forall k}) = \Gamma_{\widetilde{\theta}} \times \varphi_j^{-1/\widetilde{\theta}} \eta_{L,j}$$

Substituting in the values for η_π and η_L, we get:

$$(a)\varphi_{sj} = \frac{\left[\frac{p_{sj}}{(w_{sj}/T_j)^{\rho_s}}\right]^{\theta}}{\sum_k \left[\frac{p_{sk}}{(w_{sk}/T_k)^{\rho_s}}\right]^{\theta}}$$
$$(b)E[l_{sj}(x)] = \varphi_{sj}^{-1/\widetilde{\theta}_s} \Gamma_{\widetilde{\theta}_s} \left[\rho_s \frac{T_j p_j}{w_{sj}}\right]^{\frac{1}{1-\rho_s}} \times x^{\frac{1}{1-\rho_s}}$$
$$(c)E[\pi_{sj}(x)] = \frac{1-\rho_s}{\rho_s} \frac{w_{gsj}}{T_j} \times \left\{ \underbrace{\varphi_{gsj}^{-1/\widetilde{\theta}_s} \Gamma_{\widetilde{\theta}_s} \left[\rho_s \frac{T_j p_{sj}}{w_{gsj}}\right]^{\frac{1}{1-\rho_s}} \times x^{\frac{1}{1-\rho_s}}}_{=E[l_{sj}(x)]} \right\}$$

C Robustness of Model Estimation and Results

C.1 Model Estimation at a More Disaggregate Industry Level

In the baseline empirical exercise (Section 4), we aggregate industries to agriculture, manufacturing and services. As we discuss in Section 3 (and Table 2), it is unlikely that sorting into more disaggregate industries drives our results given that the main patterns of the data are also present at the NIC-4 classification level, including the fact that women hire more women. Nevertheless, to examine the robustness of our conclusions, we also conduct the analysis at a more disaggregate level – to the extent permitted by data constraints.

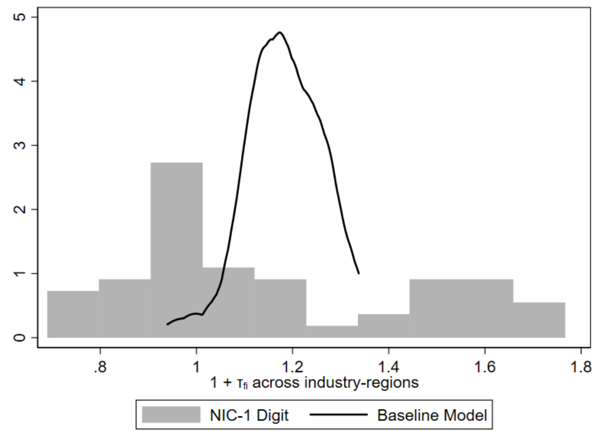
Specifically, we re-estimate our model using data at the NIC 1-digit level instead of the three aggregate industries. To facilitate comparison of the new estimates with the baseline case, we aggregate each industry-region specific estimate across regions and NIC 1-digit industries (weighted by the total individuals in that industry-region) to the three industries we consider in our baseline analysis (agriculture, manufacturing and services), and report the results in Table C1.

Table C1: Parameter Estimates based on NIC-1 Digit Classification of Industries

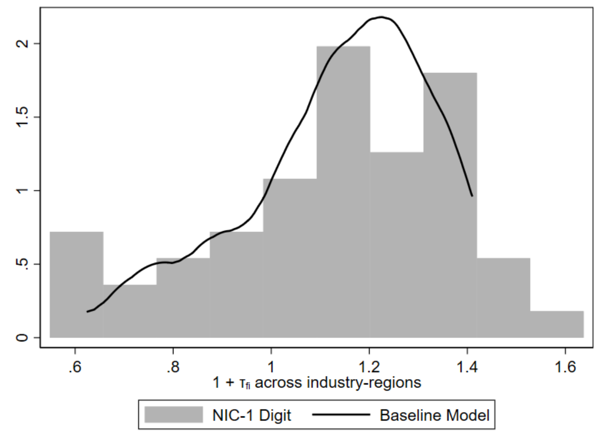
	Baseline Model	NIC 1-Digit	(1) - (2)
	(1)	(2)	(3)
τ_{fI}	1.18	1.15	0.04
	[0.08]	[0.31]	
τ_{fF}	1.14	1.12	0.01
	[0.19]	[0.25]	
τ_{fI}^{f}	0.95	0.95	-0.00
	[0.03]	[0.02]	
τ_{fF}^{f}	1.00	1.11	-0.11
	[0.25]	[0.30]	

Column (1) reports the values from the baseline model (Table 5), while Column (2) reports the values obtained by aggregating the NIC 1-digit estimates. Column (3) reports the difference between the two columns. The numbers in Column (2) are very similar to those in Column (1). Figure C1 compares the entire distribution of hiring barriers estimated based on NIC 1-digit level data (gray bars) to the distribution from the baseline model (from Figure 3). The distributions overlap greatly, except for τ_{fI}. Importantly, the distributions of the τ^f_{fs}'s, which reflect the comparative advantage of females in hiring females, and which play an important role in our counterfactual exercises, are very similar in the two cases.

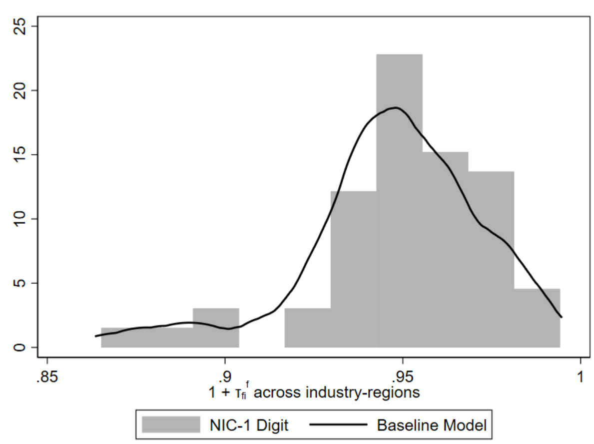
Figure C1: Comparing Parameter Estimates for Baseline Model and NIC 1-digit Level



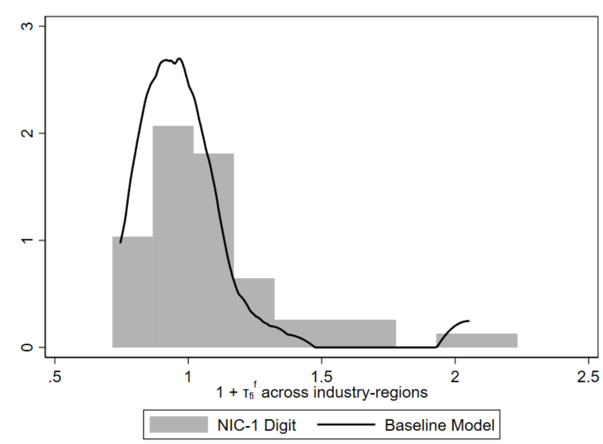
(a) $1 + \tau_{fI}$



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



(c) $1 + \tau^f_{fI}$



(d) $1 + \tau^f_{fF}$

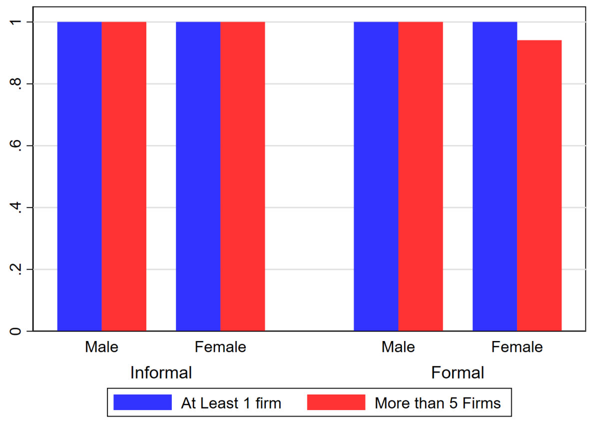
Notes: The above figures report the parameter estimates for the hiring barriers faced by women entrepreneurs. The histogram shows the estimates at the NIC 1-digit, while the solid line shows the density for the aggregate industries as reported in the paper (Figure 3).

In theory, one could re-estimate the model at even more disaggregate levels (NIC 2- or 3-digit levels). However, the poor representation of female-owned firms in several industries limits this exercise. We illustrate the problem in Figure C2 below.

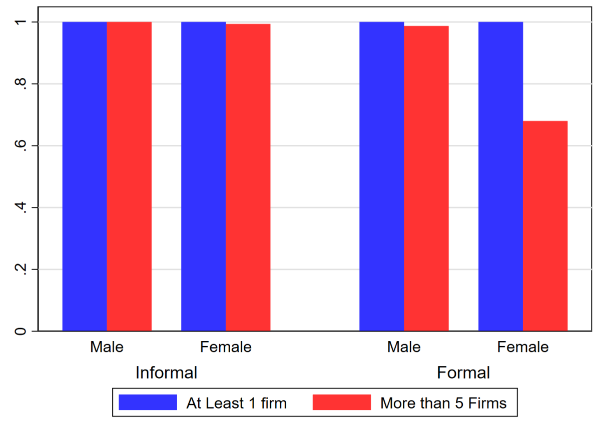
Specifically, we conduct the following exercise: consider a particular level of industry classification (NIC 1-digit, 2-digit, etc.). We calculate the fraction of industry-region pairs in 2005 that have - at that level of industry classification - at least one (five) firms of gender g in a sector s (for example, male-owned firms in the informal sector). In Figure C2(a), in which we define an industry at the aggregate level (agriculture, manufacturing and services), all industry-region pairs have at least 5 male-owned firms in both the informal and formal sectors, and 100% (95%) of industry-regions have at least 5 female-owned firms in the informal (formal) sector. At the NIC 1-digit level (Figure C2(b)), only two-thirds of industry-regions have at least 5 female-owned firms in the formal sector. At the NIC 2-digit level (Figure C2(c)), the coverage of firms drops even more. Only 85% of industry-regions have at least 5 female-owned firms in the informal sector. In the formal sector, only 79% and 31% of industry-regions have at least five male-owned and female-owned firms respectively. Finally, at the NIC 3-digit level (Figure C2(d)), only 87% (63%) of industry-regions have at least 1 (5) female-owned firm in the informal sector. In the formal sector, only 80.5% (54.5%) of industry-regions have at least 1 (5) male-owned firms and 34% (10%) of industry-regions have at least 1 (5) female-owned firms.

Having no or very few firms - especially owned by women - in several industry-region pairs does not allow us to estimate the fixed costs of entry into these industry-region pairs, prohibiting us for conducting the analysis at a more disaggregate level. However, the fact that the estimated barriers are virtually unchanged when we estimate the model at the 1-digit level (instead of the more aggregate level in the baseline case), is reassuring.

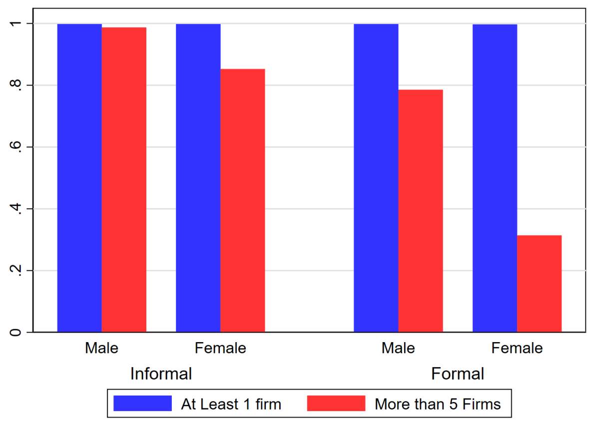
Figure C2: Fraction of Male-Owned and Female-Owned Firms at NIC 1, 2 and 3-digit Industries



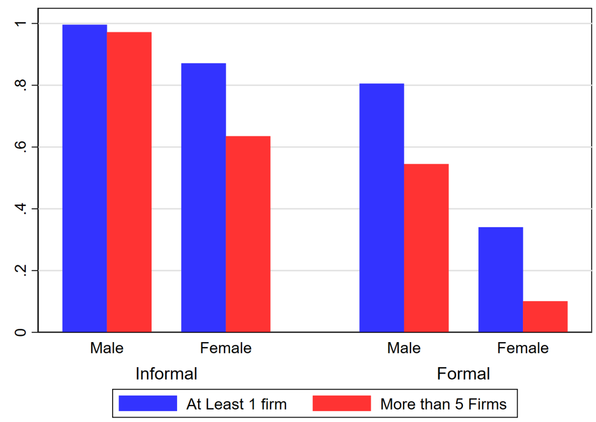
(a) Aggregate Industries



(b) NIC 1-digit Industries



(c) NIC 2-digit Industries



(d) NIC 3-digit Industries

Notes: The above figures report the fraction of industry-region pairs that have at least one firm (green bars) or five firms (orange bars) of gender g (Male, Female) and sector s (Informal, Formal). Figure (a) defines an industry at the aggregate level (agriculture, manufacturing and services). Figures (b)-(d) define an industry at the NIC 1-digit, 2-digit, 3-digit respectively.

C.2 The Role of Non-Hired Individuals

Figure 2 shows that the fixed costs of entering wage work or starting informal entrepreneurship are very low (relative to self-employment), for both men and women. This may seem surprising at first, given that wage work is considered highly desirable in many low-income countries, and women have been shown to be reluctant entrepreneurs (Jensen, 2022; Schoar, 2010).

As noted earlier, these estimates may reflect heterogeneity in wage employment and informal entrepreneurship. Many wage jobs are low-paying and provide no benefits. Similarly, some informal enterprises barely differ from self-employment (in the sense that they may employ two, instead of just one, people, but are otherwise similar in size and productivity to owner-operated businesses). Such options may not seem particularly desirable relative to self-employment. Hence, they may not entail the high fixed costs of entry one typically associates with "good" wage jobs or successful enterprises.

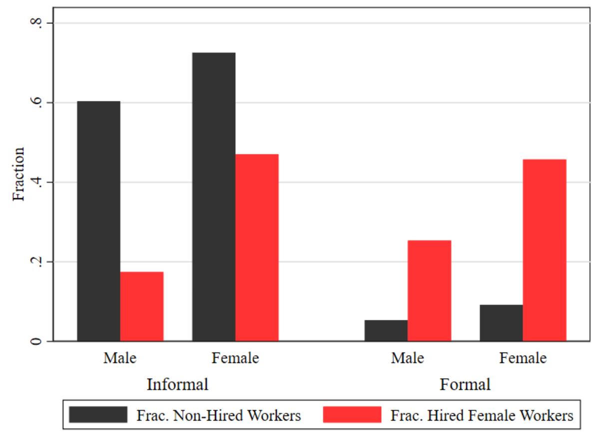
In this section, we explore one particular source of heterogeneity: the employment of "non-hired" workers. The Economic Census separately reports the number of "hired" and "non-hired" workers (by gender) within a firm. Non-hired workers are typically household members working in smaller firms and/or apprentices. Such workers are classified as "wage workers" in our baseline framework (since we do not distinguish between hired and non-hired workers). Given that they do not go through a formal hiring process, they presumably face lower fixed costs of entering wage employment.

Figure C3(a) reports the non-hired workers and the hired female workers as fractions of total workers across firms of gender g in sector s. Two observations stand out.

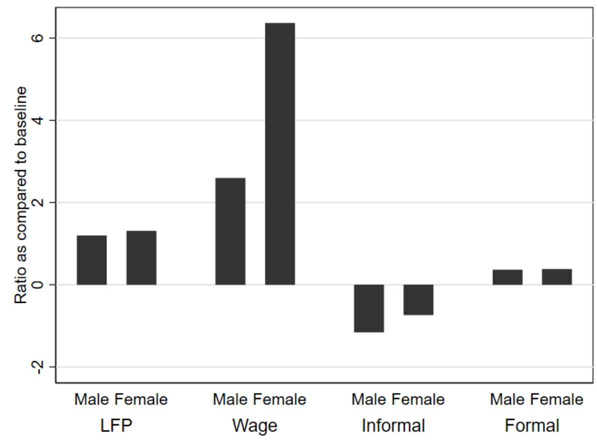
First, non-hired labor is pervasive in the informal sector for both male- and female-owned firms (60-70% on average), but less so in the formal sector (around 5%). The high incidence of non-hired labor could rationalize the low fixed costs of wage employment we estimate, shown in Figure 2. Second, the fraction of hired female workers is higher (around 40%) in female-owned firms than in male-owned firms (around 20%), indicating that the comparative advantage of female entrepreneurs in employing females is not driven by the use of "non-hired" labor, but is present among hired workers as well.

To understand the role of non-hired labor in the fixed cost estimation, we classify non-hired workers as self-employed, and then re-estimate the model to obtain new fixed cost estimates. This scenario, though extreme, is useful as a benchmark because classifying

Figure C3: Fixed Cost Estimates after Reclassifying Non-Hired Workers



(a) Frac. of Hired Female Workers



(b) Ratio of Fixed Costs

Notes: Both figures use data from the 2005 Economic Census. Figure (a) reports the non-hired workers and hired female workers as fractions of total workers in firms owned by gender g in sector s. Figure (b) reports the average ratio of the fixed costs for LFP, entry into wage work, entry into the informal, and entry into the formal sector for male-owned and female-owned firms, when non-hired workers are classified as self-employed, to the fixed costs as estimated in our baseline model.

non-hired workers as self-employed implies that they earn an income λw^g, which is lower than the market wage w^g in expectation. Figure C3(b) reports the ratio of the new gender-specific fixed costs in LFP, wage work, informal and formal entrepreneurship to the gender-specific fixed costs in our baseline framework.

The results are intuitive and confirm the hypothesis that the low estimates of the fixed costs of wage employment are driven by non-hired labor. When non-hired labor is treated as being self-employed, the big change is in the fixed costs of wage work which increase substantially for both men (2.6x) and women (6.3x) relative to the baseline. Correspondingly (and perhaps unsurprisingly), the fixed cost of informal entrepreneurship decreases slightly for both men and women (by around 1x), indicating the emergence of "reluctant" informal entrepreneurs now that the fixed costs of wage employment are higher.

Given that the focus of the paper is on entrepreneurship, and not on wage work, our baseline specification, in which all workers (hired and non-hired) are considered firm employees, remains our preferred specification. In future work, it would be interesting to explore the heterogeneity in wage employment more fully, but this is outside the scope of the present paper.

D Correlation of Parameter Estimates with Measures of Gender Norms

Figure 2 and Table 5 indicate that women face higher costs of participating in the labor force (LFP costs), formalizing their business, and hiring workers. On the other hand, they face an advantage in hiring female workers (in both the formal and informal sectors). This section explores the plausibility of the estimates. Specifically, we use region-specific measures of women empowerment from various sources in the literature to examine whether our implied measures of gender-related barriers correlate with the documented level of women empowerment in these regions.

D.1 Measuring Gender Empowerment

We use three widely used measures of gender inequality and empowerment in India: (a) Women Empowerment Index (Bansal, 2017); (b) Gender Vulnerability Index (Plan International, 2017); and (c) Patriarchy Index (Singh et al., 2021).

The Women Empowerment Index (WEI), proposed by Bansal (2017) at the Hindustan Times (a widely circulated national daily) uses data from the National Family Health Survey (NFHS), a large, nationally representative survey conducted by the Health and Family Welfare Ministry. In particular, it is based on data for eight indicators, such as the participation of women in household decisions, ownership of land, cell phones and bank account, instances of spousal violence, etc., to construct a state-specific Women Empowerment Index.

The Gender Vulnerability Index (GVI), proposed by Plan International (2017), expands the scope of the WEI by using a set of 170 indicators constructed from large nationally representative data like the Population Census of India, National Family Health Survey (NFHS), Health Management Information System, District Information for School Education (DISE), Rapid Survey on Children, Annual Economic Survey, Annual Survey on Education Report and National Achievement Survey to construct a state-specific, comprehensive measure of gender parity along various dimensions, such as Social Protection (26 indicators), Education (68 indicators), Health (57 indicators), Poverty (19 indicators). These are then aggregated to construct a state-level index of Gender Vulnerability.

Lastly, the Patriarchy Index (PI), proposed by Singh et al. (2021), adapts the Patriarchy Index developed by Gruber and Szołtysek (2016) for Europe, to the Indian context. Using the NFHS data as well, the PI uses measures that span five domains: (1) domination of men over women; (2) domination of the older generation over the younger generation; (3) patrilocality; (4) son preference; and (5) socio-economic domination that recognizes the social and economic imbalances between men and women in households in terms of both earning and control over money and education.

D.2 Gender Norms, Fixed Costs and Hiring Barriers

We begin by examining the association between LFP costs and measures of gender norms by estimating the following regression:

$$Y_{st} = \alpha_t + \beta I_s + \gamma X_{st} + \varepsilon_{st} \tag{14}$$

where Y_{st} is the percentage difference between female and male LFP costs, i.e., excess costs faced by women. We pool the 1998 and 2005 estimates, and examine their correlation with state-specific measures of women empowerment $I_s = \{GVI, WEI, PI\}$. All indices are normalized to have mean 0 and standard deviation 1. We control for state-year-specific observables such as GDP and the fraction of SC/ST population (backward castes), as well as year fixed effects that capture all observable and unobservable trends in India over this time period. Given the small sample size, we bootstrap our standard errors.

Our coefficient of interest is β. As reported in Columns (1)-(3) of Panel A in Table D1, a one standard deviation increase (decrease) in WEI/GVI (PI) is correlated with approximately a 0.5 p.p. or 35% (0.26 p.p. or 20%) decrease in the ratio of female to male LFP costs. There is no statistical association between gender empowerment and formalization costs (Panel B), though the coefficients in Panel B have the expected signs.

Next, we examine how hiring distortions (τ_{fs} and τ^f_{fs}) relate to measures of women empowerment. We re-estimate Equation (14), where Y_{jst} is now the hiring distortion in industry j, state s and year t. In addition to the variables described previously, we include industry fixed effects, α_j, to control for time-invariant differences across industries and control for the female labor force participation rate in order to net out the costs to LFP participation that were the focus of Table D1.As reported in Panel A of Table D2, we find a negative association between empowerment indices and hiring distortions in the informal and formal sectors, indicating that - conditional on entry - the barriers to business expansion for women entrepreneurs are higher in the more gender-conservative areas. Regarding the comparative advantage of female entrepreneurs in the hiring of female workers (Panel B), we find no statistically significant association. A possible interpretation is that - as noted earlier - this comparative "advantage" could itself be the result of gender-related distortions; if women are discouraged from finding work outside the home due to conservative norms, it is possible that they will only take jobs in female-owned firms, giving rise to the documented pattern in the data.

The associations documented above suggest that while the model treats barriers to entry and operation facing women as a black box, our estimates of such barriers do correlate with measures of women empowerment across Indian states.

Table D1: Correlations of Cost Estimates and Measures of Women Empowerment

	WEI	GVI	PI
	(1)	(2)	(3)
Panel A: Relative LFP Costs			
Index	-0.500***	-0.461***	0.255*
	(0.001)	(0.001)	(0.061)
R^2	0.348	0.317	0.227
Panel B: Relative Formal Sector Entry Costs			
Index	-0.185	-0.00329	0.0125
	(0.489)	(0.988)	(0.940)
R^2	0.101	0.090	0.090
N	34	34	34

Notes: The dependent variable in Panel A (B) are the relative LFP (Formal Sector Entry) costs, which is the percentage difference between female and male costs. WEI = Women Empowerment Index; GVI = Gender Vulnerability Index; PI = Patriarchal Index. All indices have been normalized to have mean 0 and standard deviation 1. All regressions control for the GDP of the state, fraction of population comprising of SC/ST castes, and year fixed effects. p-values from bootstrapped standard errors are reported in parentheses.

Table D2: Correlations of Hiring Barriers and Measures of Women Empowerment

	Informal			Formal		
	WEI	GVI	PI	WEI	GVI	PI
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Hiring barriers $(1+\tau_{fsj})$						
Index	-0.0258**	-0.0353***	0.00618	-0.0345	-0.0861***	-0.0137
	(0.026)	(0.006)	(0.531)	(0.281)	(0.019)	(0.549)
R^2	0.182	0.204	0.153	0.488	0.521	0.482
Panel B: Hiring barriers for female relative to male workers $(1+\tau^f_{fsj})$						
Index	0.0000599	-0.00375	-0.000280	0.0367	0.0124	0.00880
	(0.986)	(0.268)	(0.898)	(0.255)	(0.729)	(0.573)
R^2	0.246	0.252	0.246	0.156	0.143	0.143
N	102	102	102	102	102	102

Notes: The dependent variable in Panel A is $1+\tau_{fsj}$ and $1+\tau^f_{fsj}$ in Panel B. Columns (1)-(3) refer to the informal sector, while Columns (4)-(6) refer to the formal sector. WEI = Women Empowerment Index; GVI = Gender Vulnerability Index; PI = Patriarchal Index. All indices have been normalized to have mean 0 and standard deviation 1. All regressions control for the GDP of the state, and fraction of population comprising of SC/ST castes, female labor force participation rates, and fixed effects for year and industry. p-values from bootstrapped standard errors are reported in parentheses.

E A General Model of Production with Many Inputs

Our model in the paper considers labor as the only input in production. This modeling is driven by data constraints as we do not observe any other inputs in the Economic Census that we use for estimation. In this section, we extend our baseline model to allow for multiple inputs and examine its implications for our estimates. We use the extended model for two purposes. First, we derive expressions of the gender distortions in the extended model and compare them to those obtained in the single-input model. Second, we use the NSS data that provides information on multiple inputs for a subset of firms to estimate the distortions in informal manufacturing based on the extended model, and compare the estimates to those obtained using our baseline model.

Consider the following production function:

$$Y = z\left(L^{\alpha_L}\prod_{i\neq L}K_i^{\alpha_i}\right)^{\rho} \tag{15}$$

$$\text{where: } \alpha_L + \sum_{i\neq L}\alpha_i = 1$$

where K_i are a set of i inputs in production with an expenditure share α_i. For now, we abstract from the distinction between the formal and informal sector. Let w_{ig} be the price for input i paid by an entrepreneur of gender g such that $w_{im} = w_i$ and $w_{if} = (1+\tau_i)w_i$, i.e., women entrepreneurs face a potential distortion τ_i on input i. The profit maximization problem of an entrepreneur becomes (we drop the gender script for ease of notation):

$$\pi = \max_{\{L,K_i\}} pz\left(L^{\alpha_L}\prod_{i\neq L}K_i^{\alpha_i}\right)^{\rho} - w_L L - \sum_i w_i K_i$$

E.1 Identification of Gender Barriers and Comparison with the Single-Input Model

To solve the profit-maximization problem of the entrepreneur, we can break the optimization problem into two steps. In the first step, the profit maximization problem can be written as:

$$
\begin{aligned}
\pi &= \max_{\{M\}} pzM^{\rho} - w_M M \\
\text{where: } M &= L^{\alpha_L} \prod_{i \neq L} K_i^{\alpha_i} \\
\text{and: } w_M &= \left(\frac{w_L}{\alpha_L}\right)^{\alpha_L} \times \prod_i \left(\frac{w_i}{\alpha_i}\right)^{\alpha_i}
\end{aligned}
\tag{16}
$$

The first-order condition implies:

$$
M^*(z) = \left[\rho \frac{z}{w_M/p}\right]^{\frac{1}{1-\rho}} \tag{17}
$$

In the second step, we solve the cost-minimization problem conditional on the choice of $M^*(z)$, which implies:

$$
L^*(z) = \frac{\alpha_L}{w_L} \times w_M \left[\rho \frac{z}{w_M/p}\right]^{\frac{1}{1-\rho}} \tag{18}
$$

$$
K_i^*(z) = \frac{\alpha_i}{w_i} \times w_M \left[\rho \frac{z}{w_M/p}\right]^{\frac{1}{1-\rho}} \tag{19}
$$

Equations (17)-(19) provide important insights as to how this extension relates to our baseline model. From Equation (17), note that since $w_{if} = (1+\tau_i)w_i$, for an entrepreneur with ability z,

$$
\begin{aligned}
M_f(z) &= \Bigg[\underbrace{(1+\tau_L)^{\alpha_L} \times \prod_i (1+\tau_i)^{\alpha_i}}_{=1+\tau_M}\Bigg]^{\frac{-1}{1-\rho}} \times M_m(z) \\
\Rightarrow \frac{M_f(z)}{M_m(z)} &= (1+\tau_M)^{\frac{-1}{1-\rho}}
\end{aligned}
\tag{20}
$$

i.e., if one had information on the other inputs K_i, as we do with labor, then one could identify a composite index of distortions faced by women entrepreneurs as compared to men.

Moreover, from Equations (18) and (19), note that:

$$\frac{L_f(z)}{L_m(z)} = \frac{1+\tau_M}{1+\tau_L} \times (1+\tau_M)^{\frac{-1}{1-\rho}}$$

If we had information on the other inputs, so that we could identify τ_M, then we could separately identify the true distortion in labor hiring $1+\tau_L$, from the distortions affecting other inputs $1+\tau_M$. Instead, what we identify based on the current approach that considers labor as the only input is $(1+\widetilde{\tau_L})$, where:

$$\begin{aligned}
(1+\widetilde{\tau_L})^{\frac{-1}{1-\rho}} &= \frac{1+\tau_M}{1+\tau_L} \times (1+\tau_M)^{\frac{-1}{1-\rho}} \\
1+\widetilde{\tau_L} &= \left[\frac{1+\tau_L}{1+\tau_M}\right]^{1-\rho} \times (1+\tau_M) \\
&= \left[1+\tau_L\right]^{1-\rho}\left[1+\tau_M\right]^{\rho} \qquad (21)
\end{aligned}$$

i.e., we identify a weighted average of the true τ_L and barriers to all inputs (τ_M). This is why we interpret the distortions in hiring as distortions in expanding the business. Note however that this modeling does not affect the finding that female entrepreneurs have a comparative advantage in the hiring of female workers, since this comparative advantage is identified from the ratio of female to male workers in each firm, conditional on firm size.

E.2 Estimating A Model with Multiple Inputs Using the NSS Establishment Surveys

As noted earlier, the Economic Census provides information only on one input, labor. We use the Economic Census because it is the only data set that covers the entire firm distribution. However, if we confine the analysis to a subset of firms, then we can draw on other data sets that contain information on additional inputs. Such a data set is the Survey of Unorganized Manufacturing Firms from the National Sample Survey (Round 62) in 2005. Like the Economic Census, the NSS asks firms to report the gender of the owner as well as the number of employees and their gender. In addition, it asks firms detailed questions on their sales, wage bill, expenditure on raw materials, capital, and loans. We use the NSS to estimate a model with multiple inputs and compare it to our baseline model. However, the NSS surveys only small, informal firms, and only in the manufacturing sector. This implies that we cannot use it to estimate the barriers faced by women in agriculture or services or the formal manufacturing sector. Therefore, we use the NSS only to examine the robustness of our findings.

Gender Differences in Production Technology

A potential concern in our analysis is that distortions in input markets may affect the production technology women use relative to men. As a result, the "barriers" we estimate could reflect underlying differences in the production functions of male-owned versus female-owned firms. For instance, if female entrepreneurs do not have access to capital, they may choose to operate more labor-intensive technologies.

The NSS data allows us to examine this hypothesis. Note that according to the model presented above, the share of expenditure on an input i is equal to $\rho\alpha_i$. This share incorporates the relevant parameters of the production technology. Based on the NSS, we can calculate the expenditure shares for the three key inputs (labor, capital, materials) as follows. We define the firm expenditure on capital to be the total value of assets that are owned or hired by the firm. These include plant and machinery, transport, and expenditure on software and hardware. For expenditure on labor and materials, we use the total wage bill and the expenditure on raw materials respectively. We then calculate the expenditure share of each input (labor, capital and materials) in total sales.

As reported in Table E1, the three expenditure shares are similar across male-owned and female-owned firms. Not only is the raw difference (Column 3) negligible in magnitude,

Table E1: Share of Inputs in Total Sales

	Male	Female	Difference: (2) - (1)	
			Raw	F.E.
	(1)	(2)	(3)	(4)
Labor	0.12	0.13	0.01	-0.011
			[0.66]	[0.36]
Capital	0.15	0.12	-0.025	-0.018
			[0.33]	[0.38]
Raw Material	0.52	0.49	-0.029	0.0091
			[0.38]	[0.75]

Notes: The table shows the share of labor, capital and raw materials in sales averaged across male-owned and female-owned firms in Columns (1) and (2) respectively. Column (3) reports the raw difference between the means in the previous two columns. The discrepancies are due to rounding errors. Column (4) reports the difference based on regressions that control for an entrepreneur's education level, whether the owner works full-time in the firm, whether the firm is registered with any authority, and district and NIC 5-digit industry fixed effects. p-values calculated from robust standard errors are reported in parentheses below.

this difference is similar even after including district and NIC 4-digit industry fixed effects (Column 4), indicating that they are not driven by sorting across space or industries either. We conclude that at least in the NSS data, there is no evidence of men and women using different production technologies.

Estimating Barriers Using Measures of MRPL, MRPK, MRPR

One of the limitations of the Economic Census is that it does not report the expenditure on any input (including labor). However, given that we observe the input expenditures in the NSS, we can follow the methodology of Hsieh and Klenow (2009) to calculate measures of marginal product revenues of labor (MRPL), capital (MRPK) and raw materials (MRPR) and examine their magnitudes across male-owned and female-owned firms. The model presented above implies that:

$$
\begin{aligned}
MRPL_g &\equiv \frac{\rho\alpha_L pY_g}{L_g} = (1+\tau_L^g)w_L && \text{(Labor)} \\
MRPK_g &\equiv \frac{\rho\alpha_K pY_g}{K_g} = (1+\tau_K^g)w_K && \text{(Capital)} \\
MRPR_g &\equiv \frac{\rho\alpha_R pY_g}{R_g} = (1+\tau_R^g)w_R && \text{(Raw Materials)}
\end{aligned}
\tag{22}
$$

Given that there is no evidence (at least in the NSS data) of any differences in production technology between male- and female-owned firms, any deviations of the MRPs of female-owned firms from those of male-owned firms must reflect distortions (Hsieh and Klenow, 2009). We calculate the MRP of each of the three inputs in our data as follows.

In contrast to labor, the NSS does not provide information on the "quantity" of capital or materials. We follow an approach similar to Hsieh and Klenow (2009) to assign "prices" to capital and materials. The NSS asks firms about their total outstanding loans, along with the interest payable on these loans during the reference period. We calculate the interest rate as the ratio of these two values[35] and use it to deflate the total capital expenditure to calculate K. For raw materials, each firm reports the value and quantity for up to 5 specific products used as raw materials. We use this information to calculate the price for each product, and weight it by its share in total expenditure on raw materials to calculate an (expenditure-weighted) price of raw materials for each firm. We then deflate the expenditure on raw materials by this price index to calculate M. Given these measures, we then compute measures of MRPL, MRPK and MRPR for each firm (Equation 22) and estimate the following regression:

$$
\ln MRPx_i = \alpha_x + \beta_{x,s} FemaleOwner_i + \varepsilon_i \tag{23}
$$

where $x = \{K, L, R\}$ and from Equations (22) and (23), τ_x will be equal to $e^{\widehat{\beta_x}} - 1$. This is reported in Table E2. Columns (1)-(3) report the value for τ_x. Columns (4) and (5) use Equations (20) and (21) to calculate τ_M and $\widetilde{\tau}_L$ respectively. Note that this estimate in Column (5) is close to the value that we estimate in Section 5.3 of the paper, which is 0.21 (mean) and 0.23 (median) for informal manufacturing in 2005.

There are two main takeaways from these results. First, the distortion estimates we obtain from the NSS data when we make the assumption of a single-input ($\widetilde{\tau}_L$) are very

[35] We replace missing values with the gender-, registration-status-, and state-specific average.

Table E2: MRPL, MRPK and MRPR

	τ_L	τ_K	τ_R	τ_M	$\widetilde{\tau_L}$
	(1)	(2)	(3)	(4)	(5)
NSS	0.36	1.10	0.20	0.29	**0.30**

Notes: For an input x, $\tau_x = e^{\widehat{\beta_x}} - 1$ using estimates from Equation (23). Columns 1-3 report the estimates for τ_L, τ_K and τ_R respectively. Column 4 uses Equation (20) to calculate τ_M. Column 5 uses Equation (21) and reports the "implied" $\widetilde{\tau_L}$.

similar to those obtained from the Census data for the corresponding sector (informal manufacturing). More importantly, the estimate of the composite gender distortions in the multiple-input model, τ_M, is similar to the one obtained using the approach we described in the baseline model, $\widetilde{\tau_L}$. This gives us confidence that our estimates of "hiring" barriers reflect the combined distortions women face in *expanding* their business.

F Gender Differences in Entrepreneurial Ability

Our baseline model assumes that the entrepreneurial ability distribution is the same for men and women. This section examines the validity of this assumption and its implications for our main conclusions.

Even if men and women have the same innate ability, it is possible that gender-based discrimination leads to differences in other characteristics, most importantly education, which could make women less suitable to entrepreneurship than men. Therefore, in the next subsection, we examine gender differences in educational outcomes in India during our sample period. Education is only one among several characteristics that could affect entrepreneurial performance. Therefore, we next investigate whether surveys of the population and experts show women to have traits that are considered undesirable for entrepreneurship (of course, the survey responses could themselves reflect gender-bias, but this makes responses that do not suggest any innate differences in entrepreneurial suitability even more credible). Finally, we estimate a version of the model in which we allow the variances of the ability distributions of men and women to differ, and show that the results are virtually unchanged.

F.1 Measuring Ability based on Micro Data (IHDS)

We use data from the 2005 round of the India Human Development Survey (Desai et al., 2005) to compare the educational attainment of men and women. The IHDS is a nationally representative, multi-topic survey of 41,554 households in 1,504 villages and 970 urban neighborhoods across India.

The IHDS collects data on the educational attainment of all household members. A key advantage of this data set is that children aged 8-11 had to also complete short reading and arithmetic tests, which were implemented in a way similar to the ASER modules. For example, the reading test (implemented in the local language) had four levels corresponding to being able to recognize letters, words, paragraphs, and read stories respectively. The arithmetic test tested whether a child could recognize numbers, perform elementary operations like addition and subtraction, and more complex ones like multiplication and division.

We use this data to estimate the following regression for an individual i between the ages 18-65, living in a household h of village v:

$$Y_{i(hv)} = \alpha + \beta Female_i + \gamma X_i + \varepsilon_i \tag{24}$$

where Y_i are two outcome variables: (i) a dummy variable that takes the value 1 if the individual is literate and 0 otherwise; (ii) years of education. Female$_i$ takes the value 1 if the respondent is a female and 0 otherwise. We also control a quadratic polynomial for age, and add either village or household fixed effects to take into account unobservable differences across villages or households that could impact the educational attainment of individuals. We cluster standard errors at the village-level.

Table F1: Education Levels

	Literate		Ed. Years		LAYS #1		LAYS #2	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	-0.263***	-0.267***	-2.686***	-2.764***	0.355***	0.355***	0.347***	0.346***
	(0.005)	(0.005)	(0.044)	(0.042)	(0.006)	(0.006)	(0.006)	(0.006)
Male Mean	0.78	0.78	7.05	7.05	-0.00	-0.00	-0.00	-0.00
R2	0.34	0.66	0.43	0.76	0.39	0.63	0.40	0.63
N	113627	112798	113627	112798	113627	112798	113627	112798
Village FE	Yes	No	Yes	No	Yes	No	Yes	No
HH FE	No	Yes	No	Yes	No	Yes	No	Yes

The results are reported in Columns (1)-(4) of Table F1. Approximately 78% of men and 51% of women between the ages of 18-65 are literate. While men have around 7 years of education, women only have around 4.35 years of education. These results suggest that women lag behind men in terms of schooling.

However, as documented by Angrist, Djankov, Goldberg and Patrinos (2021), enrollment and learning are different measures of educational attainment, and they do not always go hand in had. As discussed earlier, a key advantage of the IHDS is that it measures *learning*, and not just schooling, for children between 8-11 years of age. We use this information to create a measure of "learning-adjusted" years of schooling (LAYS) in the following four steps:

1. For each child c in a household h and village v, we calculate her/his total "learning" score as the sum of the (standardized) reading and math scores.

2. Using the sample of children for whom we observe the learning scores, we estimate Equation (24), where Y_c now indicates the learning score of the child. We add village fixed effects, control for a quadratic polynomial of age, type of school (public, private, convent, madrassa, etc.), and a set of household characteristics such as size, asset index, highest educational level of parents, whether at least one person works in the household or not, (log) household income and poverty status.

3. Based on the estimated coefficients, we then *predict* the learning levels for the sample of adults (between the ages 18-64) and calculate LAYS for an individual i as the product of the years and his/her (predicted) learning level. For ease of interpretation, we standardize this measure to have mean 0 and standard deviation of 1 for men and define it as LAYS #1.

4. We repeat steps 2 and 3, but now add household fixed effects to the regression specification in Step 2 (instead of household characteristics), and calculate a second measure of LAYS #2.

We then estimate Equation (24) with the LAYS measures as our dependent variables and report the results in Columns (5)-(8) of Table F1. As is clear from the table, even though women have lower levels of literacy and schooling years, they have 0.3 standard deviations higher learning. These results are consistent with the cross-country patterns documented by Angrist, Djankov, Goldberg and Patrinos (2021).

To summarize the above discussion, the analysis in this section indicates that data on education do not provide support for the premise that women may be less suited to entrepreneurship due to lack of education. Women may have fewer years of schooling, but they exhibit higher learning. This pattern may also justify an assumption that we explore later in this section, namely that the variance of the ability distribution is higher for women than for men. Some women have very few years of schooling or are illiterate, and they may make poor entrepreneurs. But there are also other highly competent women, who have made the most of their schooling.

F.2 Entrepreneurial Ability from GEM Surveys

This subsection takes another approach for assessing entrepreneurial ability based on data from the Adult Population Surveys (APS) implemented by the Global Entrepreneurship Monitor GEM (Reynolds et al., 1999). The APS is particularly valuable since it explores the role of the individual in the entrepreneurial process. The questions focus not only on business characteristics, but also on people's motivation for starting a business, the actions taken to start and run a business, as well as entrepreneurship-related personality traits. The APS is administered to a minimum of 2000 adults in each economy, ensuring that it is nationally representative. We use all rounds of the APS in India between 2001-2007 and restrict the sample to adults between the age 18-65. We estimate the following regression specification, where i denotes a respondent:

$$Y_i = \alpha_t + \beta Female_i + \gamma X_i + \varepsilon_i \tag{25}$$

Y_i are a set of individual beliefs/opinions/outcomes that we will discuss below. Female$_i$ is a binary variable that takes the value 1 if the respondent is a female and 0 otherwise. X_i are individual controls such as age, income category and educational level. We add year fixed effects in all specifications.

Barriers to Entrepreneurship and Differences in Attitudes/Traits

We first explore gender differences in the ownership and firm size. The results are reported in Table F2. In Column (1), the outcome variable is a binary variable that take the value 1 if an individual reports owning a firm. Women (as compared to men) are 12.9 p.p. (44.4%) less likely to own a firm. Columns (2) and (3) report gender differences in the current and expected (in five years) firm size. Female-owned firms hire 1.4 fewer workers (56.4%) on average, and even expect to hire 1.8 fewer workers (27.5%) in the future as well. These patterns confirm those we documented earlier using the Census data.

Next, we examine gender differences in other variables capturing risk appetite, expectations, and other attitudes as measured in the APS. For each outcome variable, Table F5 provides the detailed questions that were asked.

Table F3 examines gender differences in attitudes towards risks associated with entrepreneurship. We do not find any gender differences with respect to: (i) fear of failure that would prevent women from starting a business (Column 1); (ii) competition faced by other businesses who offer similar products and services (Column 2); (ii) optimistic or pessimistic assessment of the novelty of the product/service provided (Column 3) or the novelty of technology (Column 4); (iii) their perception of whether starting new businesses is considered a desirable career choice (Column 5), is respected (Column 6) or reported positively in the news media (Columns 7).

Table F2: Firm Characteristics

	(1) Own	(2) Current L	(3) Expected L
Female	-0.129***	-1.370**	-1.773**
	(0.013)	(0.601)	(0.776)
Male Mean	0.29	4.66	6.45
R2	0.08	0.05	0.06
N	8306	793	793

Notes: See Table F5 for a definition of all the outcome variables. Female takes the value 1 if the respondent is a female and 0 otherwise. Male mean is the average value of the outcome variable for male respondents. All regressions control for respondents' age, education, and income category along with year fixed effects. Robust standard errors are reported in parentheses. * is $p < 0.1$, ** is $p < 0.05$, and *** is $p < 0.001$.

Lastly, Table F4 examines gender differences in reasons individuals give for starting a business. Columns (1) to (4) show no differences between men and women.

To summarize the APS analysis, there is no evidence of innate gender differences in risk appetite or entrepreneurship-related attitudes that would explain the low share of female entrepreneurs and the small size of their businesses.

F.3 Re-Estimating the Model with Gender-Specific Ability Distributions

In a final exercise, we re-estimate the model to allow for a gender-specific ability distribution i.e., $x \sim \log N(0, \sigma_x^g)$. The differences in educational attainment between men and women documented earlier suggest a larger variance for the ability distribution of women (given that some women are illiterate or have very few years of schooling, while at the other end, some women exhibit higher learning than men conditional on the same years of schooling). We assume that the means of the two distributions are the same as we cannot identify differences in means. But we remind the reader that the evidence we have presented so far does not provide any support for the hypothesis that on average, women differ from men in ways that affect their suitability for entrepreneurship and their performance.

We estimate σ_x^f to be 0.37, which is greater than 0.31 – the estimate in our baseline scenario (Table 4). Figure F1 shows however that relaxing this assumption does not

Table F3: Attitudes and Risk

	Risks		Innovation		Perception		
	(1) Failure	(2) Competition	(3) New Prod.	(4) New Tech.	(5) Desirable	(6) Prestige	(7) Media
Female	0.010	-0.007	0.041	0.026	-0.036	-0.041	0.012
	(0.016)	(0.012)	(0.046)	(0.041)	(0.030)	(0.025)	(0.029)
Male Mean	0.31	0.96	0.36	0.65	0.71	0.83	0.73
R2	0.03	0.02	0.03	0.06	0.03	0.02	0.03
N	6819	2045	718	718	1382	1382	1382

Notes: See Table F5 for a definition of all the outcome variables. Female takes the value 1 if the respondent is a female and 0 otherwise. Male mean is the average value of the outcome variable for male respondents. All regressions control for respondents' age, education, and income category along with year fixed effects. Robust standard errors are reported in parentheses. * is $p < 0.1$, ** is $p < 0.05$, and *** is $p < 0.001$.

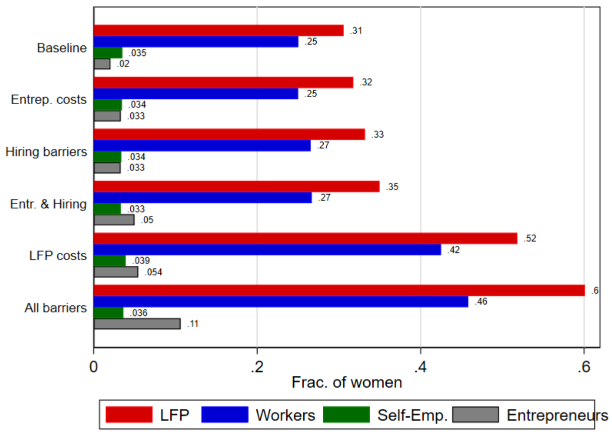
Table F4: Reason for starting business

	(1) Business Opp.	(2) Independence	(3) Higher Income	(4) Maintain Income
Female	0.015	0.050	-0.039	-0.007
	(0.031)	(0.050)	(0.053)	(0.032)
Male Mean	0.50	0.36	0.52	0.12
R2	0.04	0.09	0.08	0.02
N	2397	514	514	514

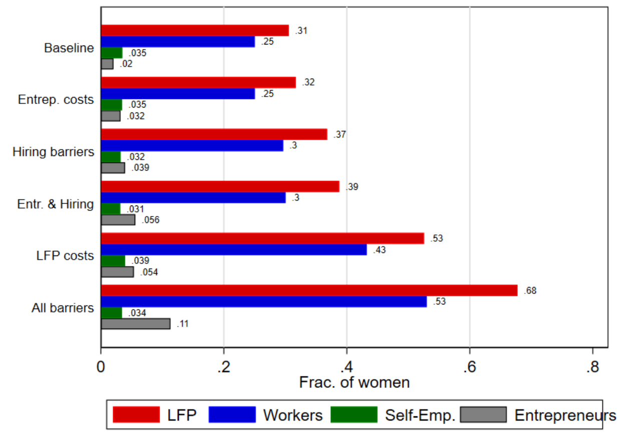
Notes: See Table F5 for a definition of all the outcome variables. Female takes the value 1 if the respondent is a female and 0 otherwise. Male mean is the average value of the outcome variable for male respondents. All regressions control for respondents' age, education, and income category along with year fixed effects. Robust standard errors are reported in parentheses. * is $p < 0.1$, ** is $p < 0.05$, and *** is $p < 0.001$.

impact our results in any meaningful way. In particular, the impact of removing gender barriers has a very similar impact on the allocation of women in the economy (Figures F1(a) and F1(b)), as well as on changes in real income (Figures F1(c) and F1(d)). If anything, the results are quantitatively larger in this case. This is because $\sigma_x^f > \sigma_{x,base}^f$. This in turn implies that when gender barriers are now removed, even more productive women become entrepreneurs, who hire other women, which increases FLFP and real income more than in the baseline.

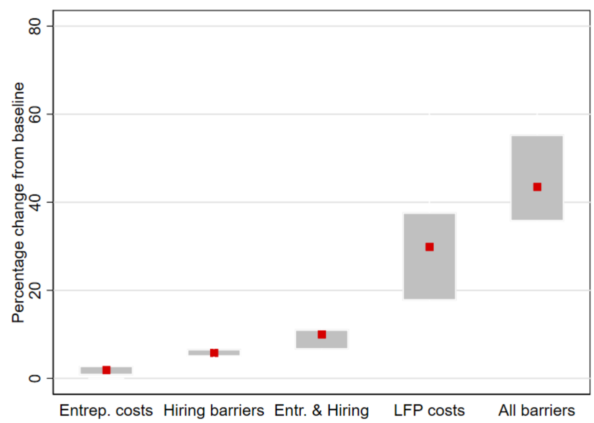
Figure F1: Gender-Specific Ability Distribution and Aggregate Impact



(a) Distribution of Women: Baseline Scenario



(b) Distribution of Women: Gender-Specific σ_x



(c) $\triangle$ Real Income: Baseline Scenario

(d) $\triangle$ Real Income: Gender-Specific σ_x

Notes: Figures (a)-(b) compare the distribution of women in the economy when the same σ_x is imposed for men and women (a), and when σ_x is allowed to vary by gender (b). Figures (c)-(d) report the corresponding changes in real income.

Table F5: Questions and Variables

Variable	Definition
Own	You are, alone or with others, currently the owner of a company you help manage, self-employed, or selling any goods or services to others.
Current L	Current firm size
Expected L	Expected firm size in the next 5 years
Risk	Fear of failure would prevent you from starting a business.
Competition	Right now, are there many, some, or no other businesses offering the same products or services to your potential customers? The variable takes the value 1 if there are some/many competitors.
New Product	Will all, some, or none of your potential customers consider this product or service new and unfamiliar? New Product takes the value 1 if "all" or "some" customers consider this product/service new.
New Technology	Have the technologies or procedures required for this product or service been available? The variable takes the value 1 if the technology has been around for less than 5 years.
Desirable	In your country, most people consider starting a new business a desirable career choice.
Prestige	In your country, those successful at starting a new business have a high level of status and respect.
Media	In your country, you will often see stories in the public media about successful new businesses.
Business Opp.	Are you involved in this start-up to take advantage of a business opportunity or because you have no better choices for work?
Independence	Which one of the following, is the most important motive for pursuing this opportunity: to have greater independence
Higher Income	Which one of the following, is the most important motive for pursuing this opportunity: higher income
Maintain Income	Which one of the following, is the most important motive for pursuing this opportunity: maintain income