Gender Gaps across the Spectrum of Development: Local Talent and Firm Productivity*

Nava Ashraf† Oriana Bandiera‡ Virginia Minni§
Víctor Quintas-Martínez¶

September 2026

Abstract

We ask whether the gendered division of work affects firm productivity across the spectrum of economic development. Personnel records of over 100,000 individuals hired by a global firm that operates in 100 countries reveal that the performance of female employees is higher where women are underrepresented in the candidate pool. This implies productivity gains from hiring more women, but realizing them would require increasing women's pay relative to men. The findings highlight how unequal gender norms in local labor markets create an equity-efficiency trade-off inside the firm, particularly in low-income countries with conservative gender norms.

JEL Codes: O12, O15, M5

*We thank our partners at the MNE for sharing the data, Sam Bowles, Stefano Caria, Anton Heil, Jonas Hjort, Erik Hurst, Rocco Macchiavello, Andrea Prat, Imran Rasul, and seminar participants at CIVOE, Harvard, HBS, IIES, IZA, LSE, MIT Sloan, NBER Summer Institute, Northwestern, Princeton, Toulouse, UCL, UMichigan for useful feedback. We thank Rishabh Malhotra, Donata Schilling, Misha Seong, Ayse Tumturk for excellent research assistance. Data access is provided through a collaboration agreement with the LSE Altruistic Capital Lab.

†STICERD and Department of Economics, LSE. n.ashraf1@lse.ac.uk

‡STICERD and Department of Economics, LSE. o.bandiera@lse.ac.uk

§University of Chicago, Booth School of Business, CEPR, IZA, and J-PAL and NBER. virginia.minni@chicagobooth.edu

¶Amazon. vquintas@amazon.com

1 Introduction

A firm seeking to hire employees in India or Turkey will choose from a pool of applicants where men outnumber women by a factor of four. A firm seeking to hire for the same job in Canada or Ethiopia will choose from a pool of applicants with a similar number of men and women. How do restrictions in the pool of talent by gender affect the productivity of firms?

The answer relies on measuring how men and women select differently into local labor markets and how this difference affects the talent pool from which firms hire. To the extent that gender norms create greater barriers to entry into the labor force for women, only those motivated and able to overcome those barriers will select into the labor force (Olivetti and Petrongolo, 2016; Mulligan and Rubinstein, 2008).[1] Testing for potential positive selection of women is important because it would imply that firms could increase productivity by changing the composition of the workforce.

Measuring potential differential selection by gender is difficult because applicants may look the same on observables at the stage of hiring (Sarsons, 2017). Once inside the firm, the same gender norms that created barriers to labor force participation create barriers to pay and career progression.

In this paper, we propose a method to measure potential positive selection by studying productivity by gender in a multinational firm where women face the same barriers inside the firm, but differential barriers in their local labor market by country and cohort. We combine individual-level data from the personnel records of a multinational company with macro-level, local labor market data on the labor force participation across 4 cohorts and the 101 countries where the multinational operates. The personnel records contain the individual earnings and career paths of approximately 100,000 employees.

We match the firm's administrative data with aggregate data at the gender-age cohort-country level from the World Bank on the ratio of women to men in the labor force, the labor force participation ratio (henceforth LFPR). We use the labor participation data to measure selection into the labor force across different countries, age cohorts, and genders. Importantly, the LFPR varies among countries with similar levels of income. Thus, differential labor force participation across genders does not solely capture the level of economic development.[2]

[1] This has a parallel in the ordeals literature on targeting, where by design ordeals (barriers) change the composition of applicants through self-targeting (Alatas, Purnamasari, Wai-Poi, Banerjee, Olken and Hanna, 2016). While ordeals can improve efficiency, in this case, barriers can distort the allocation of talent as they act differentially on a dimension - gender - that is orthogonal to productivity.

[2] For instance, Algeria and Ecuador had a similar level of GDP per capita in the 1980s, but the former's LFPR was less than 1/4 of the latter's LFPR.

Having data from the same firm operating in many countries allows us to disentangle firm policy from local conditions since the multinational adheres to global policies determined at its headquarters. On the other hand, local firms are embedded in local contexts, which can confound demand and supply factors. An additional advantage of our data is that we observe women and men with the same education, same tenure, and working in the same function in the same firm. Hence, the gender earnings gap is not influenced by differences in occupational choices that make comparisons between genders difficult.

We start by documenting stylized facts in the firm workforce. First, we show that the gender ratio in the firm mirrors the gender ratio in labor force participation across labor markets. In other words, the firm hires fewer women in countries and cohorts with low LFPR. This is in line with the firm using the same selection process in all countries; that is, they do not employ more pro-women policies in countries where norms keep women inside the home (or vice versa). The evidence is supported by the accounts of the firm's HR managers and is consistent with similarly centralized policies in the sample of 1,213 multinationals analyzed by Hjort, Li and Sarsons (2020). It also indicates that local economic conditions shape the firm workforce through the variation in labor supply by gender in each country.

Second, we analyze how differences in LFPR correlate with gender differences in pay and performance within the firm. Like Olivetti and Petrongolo (2008), we find that the gap between female and male earnings (henceforth, the gender pay gap) monotonically decreases as the LFPR increases. That is, when the LFPR is at its lowest, the gender pay gap is "inverted", and women are paid more than men with the same experience, same tenure, and working in the same function. The gap shrinks as the LFPR increases, and it converges to standard levels for the industry (negative 10%) when and where the LFPR is highest. Since we observe employees of different ages in the firm, we can exploit both cross-country and cross-cohort variation in the LFPR. The sign and size of the effect are similar whether we use within-country or within-cohort variation for identification, ruling out country-level confounders. This implies that the marginal, and hence the average, productivity of female employees is decreasing in the share of women working in the firm, as is standard in models of selection on gains (Lazear, 2021). Under the logic of selection, when a group faces a high cost to join the workforce, only the individuals whose returns are high enough will do so. As a result, the pool of workers available to firms becomes more positively selected.

Our third fact corroborates the intuition of differential selection by gender using other performance metrics. Under the rationale of positive selection, as the LFPR increases, the ability of the marginal female labor-force participant falls, in line with the existence of an ability threshold above which women choose to participate in the labor force. We find that, in low LFPR countries and cohorts, women are over-represented in the highest rungs of the hierarchy and are more likely to be promoted relative to their counterparts in high LFPR countries and cohorts. Relatedly, women are over-represented in the top decile of the wage distribution and under-represented in the bottom decile when LFPR is low. Moreover, there is a negative correlation between women's average performance and the LFPR, and it originates from the bottom percentiles of the wage distribution. The wages of the women at the bottom decile of the wage distribution decrease as the LFPR increases, while the wages of the women at the top decile remain constant.

Overall, the reduced form evidence is consistent with the interpretation that the women whom we observe in the firm in low LFPR countries had the ability to overcome higher barriers and are thus positively selected. This raises the question of why the firm does not use the differences in labor force participation rates by gender to its advantage. In particular, the firm could increase average productivity by "undoing" gender differences in local contexts and hiring more women. In the second part of the paper, we structurally estimate the parameters of the firm's compensation policies using a simple two-stage model of labor-force participation and firm matching. The model allows us to understand to what extent the firm's existing wage policy is optimal in terms of maximizing the firm's productivity and to simulate the effects of alternative firm wage policies and local economic conditions on firm productivity and pay inequality.

We first estimate the individual-level ability and the parameters of the firm's pay policy. We assume that log-pay consists of a component that is common to all people of the same gender in the same country-cohort group (for instance, country-specific regulations) and a component that is specific to the individual and proportional to his or her productivity. The structural estimates leverage the significant advantage of our data: we observe several employees in each country, gender, and cohort, allowing us to estimate both a fixed parameter common to all employees in the same cell (e.g., discrimination based on gender) and, using the variation in wages within cell, differences in individual productivity. Moreover, as the pay data comes from the firm's administrative records, we do not rely on noisily reported wages. In our case, there is no measurement error in the salary information.

Although we assume that the underlying distribution of productivity is the same for every individual, the productivity of the average worker may still differ across genders, countries, and age cohorts because of differential selection into the labor force. We use the variation in labor force participation (LFP) at the gender-cohort-country level to measure gender differences in the selection into the labor force. We find that, on average, across countries and cohorts, female employees have 0.10 higher ability (in units of standard deviations of the underlying productivity variable) than their male counterparts. The difference is larger in countries with weaker gender equity labor laws and with more conservative gender norms; for instance, the ability of women in the firm is 0.18 higher than that of men in the firm in countries with below-median LFPR, whereas this gap decreases to 0.07 in countries with above-median LFPR. The estimated calibrated ability correlates with other measures of worker performance that are not used in the estimation. With these estimates in hand, we analyze the effect of a number of counterfactual scenarios on firm productivity and inequality, which isolate the implications of changes in wage policy and labor-force participation while abstracting from other adjustment margins, such as recruiting, training, or hiring workers from other firms.

The first counterfactual compares the observed wage policy to the optimal contract that maximizes average productivity while keeping employment and the wage bill as binding constraints. We show that, given the productivity differences between men and women, the firm could increase productivity if they were to change the terms of the wage contract to attract more women. We find that the optimal contract has a lower base pay and a steeper performance gradient than the observed contract. This brings the firm's gender ratio close to one and increases productivity by 50% on average. However, we note that such a contract would significantly increase inequality within and between genders; most notably, the difference in pay between women and men would go up by 73 log points. This captures both differences in performance for the same job and differences in jobs as more able women climb the corporate ladder faster. Thus, whilst it is theoretically possible for the firm to benefit from policies that compensate for unequal labor market conditions, these policies would create a high level of inequality among employees of different genders.

In order to hire more women without excessively increasing inequality, the firm could increase women's pay without decreasing men's pay or have the same wage policy across genders. We estimate the effects of this in a second counterfactual. We let the firm optimize the wage policy with an additional constraint that imposes the wage policy parameters to be the same across genders. The increase in productivity is about 60% as large as under the unconstrained wage policy. Moreover, a gender pay gap of 14 log points remains because women's ability is higher on average.

These counterfactuals underscore the significance of a trade-off between efficiency and equality among workers *within the same firm*, which is inverted when compared to the efficiency-equality trade-off observed outside the firm. In the labor market, efficiency can be achieved by restoring equity and closing the gap in labor force participation between genders (Hsieh, Hurst, Jones and Klenow, 2019).[3] Instead, within a single organization that faces unequal participation in the labor market by gender, efficiency may require tolerating high within-firm inequality across genders. When we mute differences in gender LFP by equating the outside option of both genders, the trade-off between productivity and inequality disappears, and the firm can increase productivity by 11% and eliminate pay inequality by increasing the female share among its employees.

The final counterfactual simulates the effect of more stringent labor regulations that make it harder to link pay to performance. We find that firm productivity decreases by 17%, and the difference in pay between women and men falls by 6 log points. The results cast new light on pro-worker labor policies and the gender earnings gap. Intuitively, most pro-worker measures, such as restrictions on hiring and firing, make it harder to link pay to performance (Propper and van Reenen, 2010). Capping performance rewards makes it more difficult for the firm to attract the most positively selected female workers. The gender pay gap consequently narrows from the top rather than the bottom: pro-worker regulation compresses the gap while leaving the minority group worse off.

The main contribution of our paper is to show that gender disparities in local labor markets affect firm productivity by creating an equity-efficiency tradeoff within the firm. This phenomenon is likely to continue to grow in importance as firms continue to increase in size and expand into new regions (Hsieh and Rossi-Hansberg, 2023; Hazell, Patterson, Sarsons and Taska, 2022).

First, our analysis contributes to the literature on how firms set wages and organize their economic activity across space and how this is shaped by local economic conditions (Grossman and Helpman, 2008; Blinder and Krueger, 2013; Aghion, Bloom, Lucking, Sadun and Van Reenen, 2021). We show the consequences on firm productivity and inequality of across-country differences in the worker gender composition of local labor markets. In doing so, we elucidate how the tradeoff between equity and efficiency inside the same organization can be opposite to the usual economy-wide equity-efficiency tradeoff across organizations. In the macroeconomy, bringing convergence in labor force participation by gender or race contributes to higher growth via an improved allocation of talent (Hsieh et al., 2019). Conversely, a single firm facing unequal labor pools by demographic groups can potentially benefit from the resulting misallocation of talent by leveraging the positive selection of minority groups

[3] Hsieh et al. (2019) finds that reducing misallocation by lowering barriers across gender and race groups accounts for 41.5% of the increase in GDP per capita in the United States between 1960 and 2010.

to its advantage. And yet, fully capitalizing on this would require tolerating significant within-firm wage disparities among different groups, which may prove unfeasible. A growing strand of research in economics - as well as in psychology, sociology, and organizational behavior - has documented that individuals care about their pay relative to that of their co-workers (Charness and Kuhn, 2011; Fehr, Goette and Zehnder, 2009; Lemieux, MacLeod and Parent, 2012; Breza, Kaur and Shamdasani, 2018).

Second, our paper connects the literature on the barriers to female labor force participation (Goldin, 1995; Jayachandran, 2015; Olivetti and Petrongolo, 2016; Bursztyn, Cappelen, Tungodden, Voena and Yanagizawa-Drott, 2023) to the literature on the impact of diversity for firm productivity (Alesina and La Ferrara, 2005; Hamilton, Nickerson and Owan, 2012; Hjort, 2014; Bertrand and Duflo, 2017; Marx, Pons and Suri, 2021). Seen through the lens of selection, the link between diversity and productivity is underpinned by the traits of the minority due to the barriers they had to overcome rather than a direct "treatment" effect of diversity on productivity through, for instance, role model effects or changes in culture. In our data, we observe women and men with the same education, same tenure, and working in the same function in the same firm. This means that the gender gaps we observe are not affected by different occupational choices, which often complicate gender comparisons (e.g., Blau, 1977; Goldin, 2014; Card, Cardoso and Kline, 2016; Wiswall and Zafar, 2018; Andrew, Bandiera, Costa-Dias and Landais, 2021).

Across countries and time, the division of labor inside and outside the home has remained within the confines of norms that assign the largest share of housework to women. Because of this, the under-representation of women in spheres of influence and employment has led to significant efforts in both the private and public sectors to address the gap through extensive diversity initiatives (Bertrand, 2011; Olivetti and Petrongolo, 2016; Bertrand, 2020). Supporters of these initiatives argue that diversity per se could be beneficial for productivity and profits due, for example, to the nature of the production function or role model effects (Lazear, 1999; Athey, Avery and Zemsky, 2000; Hong and Page, 2001). We contribute to the debate by showing that to understand the consequences of gender parity on firms, it requires to first understand the selection into the workforce of women and men.

The paper is organized as follows. Section 2 presents the institutional context of the MNE and describes the data sources. Section 3 provides reduced form evidence on the link between the LFPR and differential proportion, pay, and performance in the firm between genders. Section 4 introduces the model that we then estimate using the firm's personnel data and country-cohort level LFPR, and Section 5 uses our estimates to evaluate the effects of different counterfactuals. Section 6 concludes by discussing policy implications and other issues for further research.

2 Context and data

2.1 Context

We collaborate with an MNE with headquarters in Europe and offices in more than 100 countries worldwide, as illustrated in Online Appendix Figure A.1. The MNE produces consumer goods; in 2019, it had a turnover of €20+ billion and employed over 120,000 workers, of which approximately 55% were white-collar workers. We focus on white-collar workers because blue-collar workers are only observed in two-thirds of countries where the MNE has production activities. Typical white-collar jobs in the MNE involve sales, engineering, marketing, HR, R&D for product development, and general managerial activities. The workers have homogeneous levels of human capital as applications require a college degree, and most employees have degrees in either business administration (50%) or engineering (20%).

To preserve the confidentiality of the firm, we generally refrain from reporting country labels for the entire sample of countries in the figures.

2.2 Data

Personnel records: Our sample covers the universe of employees between 2015 and 2019. We focus our analysis on local employees (non-expats), resulting in 100,819 distinct regular full-time workers over 2015-2019 in 101 countries (303,759 employee-year observations). The company is organized into a hierarchy of work levels that goes from work level 1 to 6 (C-Suite). For each employee, we observe: (1) work level and job title; (2) tenure in the firm and job; (3) annual performance score decided by the manager, and (4) total compensation (fixed plus variable pay in euros). The information on pay comes directly from the firm payroll records and hence does not have measurement error, unlike self-reported wages from labor surveys. Pay differences capture differences in performance between employees, encompassing off-peak salary increases as well as promotions. We look at four 10-year age cohorts within the company: 18-29, 30-39, 40-49, and 50-59[4].

A recent strand of evidence shows that multinationals pay higher wages in the countries where they operate (Verhoogen, 2008; Javorcik, 2015; Hale and Xu, 2016;

[4]The age band classification is decided by the firm. We do not have more granular data on age cohorts because of data privacy clauses. Due to a limited sample size of workers in age cohorts above 50-59, we only consider workers up to the 50-59 cohort.

Alfaro-Urena, Manelici and Vasquez, 2022; Hjort et al., 2020). We confirm that this is the case for our firm in Online Appendix Figure A.2 which shows that the firm's average wages are usually well above the countries' average wages, using both the average wages in the manufacturing sector from the ORBIS database and the ILO estimates for white-collar employees. Table 1 presents summary statistics separately by gender at the gender-cohort-country level, which is the relevant unit of analysis used in the structural estimation.[5] The first two columns are for the full sample of workers, and columns 3 and 4 are for the sample used in the structural estimation, where we only consider the cells with at least 30 workers of each gender. The female cells show lower average pay, age, tenure, probability of being in managerial positions, and probability of experiencing fast promotions. Overall in the company, 40% of workers are aged between 30-39 and the majority of workers are in WL1 (>70%).

Country-cohort level data: We combine the firm's administrative records with country-cohort data on labor force participation rates (LFP) of men and women from the World Bank. The World Bank labor force participation data are from the ILO modeled estimates database where the latest edition covers the years from 1990 onwards.[6] We match the worker-level data from the MNE with the country-level LFP data by pairing the workers' country and age cohort. Hence, the variation in LFP rates is at the gender-country-age cohort level. In particular, we match the age-country cohorts in the firm with the average LFP rate in the country in the decade of labor market entry, separately by gender (i.e. upon finishing their college education when they are in their 20s years of age).[7]. In this way, we can leverage variation in the LFPR both at the country level and at the cohort level, as LFPR varies in both dimensions. For example, employees aged 18-29 are associated with the LFP rates of the 2010-2020 decade, while employees of ages 30-39 are associated with the LFP rates of the 2000-2010 decade, and so on.[8]

In our main analysis, we use the LFP data for all individuals because the LFP data for individuals with tertiary education[9] are missing for almost 20% of the sample spanning 25 countries, particularly those with low female labor force participation (FLFP). However, we conduct a robustness exercise for our model calibration in Online Ap-

[5]As we explain in section 4, we let the structural parameters vary by gender, cohort, and country in the firm to allow for the wage policy to differ across countries.

[6]Accessed here on March 22, 2020.

[7]Since our sample consists of college-educated workers, the decade of labor market entry is assumed to be their 20s.

[8]Since data on the LFP for the 1980-1989 decade is not available, we use a linear interpolation of the LFP for that decade.

[9]Following the definition from the World Bank, tertiary education includes all formal post-secondary education, including public and private universities, colleges, technical training institutes, and vocational schools. See: `https://www.worldbank.org/ext/en/topic/education/tertiary-education`.

pendix D using LFP data only for individuals with tertiary education.

We also use other aggregate data for some additional analysis: GDP per capita at constant 2015 USD from the World Bank; the World Value Survey (waves 1-6); the Women, Business and the Law Index from the World Bank; the Restrictive Labor Regulations Index from the World Economic Forum; the gender development index (GDI), which is the ratio of female/male Human Development Index from the United Nations Development Programme; and the female to male ratio in years of schooling and the percentage in post-secondary education from the World Bank.

Figure 1 plots moments of the distribution of the labor force participation ratio of women's LFP against men's LFP (the LFPR)[10] at different deciles of GDP per capita. It shows that while the mean exhibits the well-known U-shape pattern (Goldin, 1995), the distributions largely overlap: there is variation in LFPR at every level of GDP per capita. For instance, the interquartile range of LFPR is broadly similar across the deciles of GDP per capita. Moreover, the 90th percentile of LFPR only ranges between 0.85 and 1. This indicates that there are country-cohort cells with high LFPR at every level of economic development, instead of being solely concentrated in high-income countries. This is essential for the analysis that follows because it allows us to partial out the differences in national income among the countries where the MNE operates.

3 Facts

In this section, we document stylized facts of the employment and wages of women and men at the firm and how these vary across different labor markets. In particular, we look at how the gender ratio of the firm workforce and its gender pay and promotion gaps correlate with the LFPR in the country. The main findings are that the firm gender ratio closely follows the LFPR in the country, particularly so in low LFPR countries, and that, when the LFPR is at its lowest, the gender pay gap is "inverted", with women being paid more than men with the same experience, tenure and working in the same function.

We interpret these findings through the lens of a simple selection model where we assume that men and women share the same underlying productivity distribution, while women face higher barriers to labor-force participation.[11] The model predicts

[10] We rescale women's LFP by men's LFP in order to account for country-cohort level differences in the probability of individuals to work. However, we note that the variation in men's LFP is tiny compared to the variation in women's LFP. In other words, the variation in LFPR is mostly driven by the variation in the female labor force participation, the numerator in the ratio.

[11] This assumption serves as a normalization for interpreting the reduced-form evidence. In the structural analysis (Section 4), we examine the robustness of the results to alternative normalizations, including a calibration based on labor-force participation among individuals with tertiary education.

when barriers are high, only the women who are most productive in the workplace will be observed there. As barriers fall, so does the average productivity of women who enter the labor force and are observed in the workplace. In line with the prediction of the Roy model, women are overrepresented in the highest rungs of the hierarchy and more likely to be promoted relative to their counterparts in high LFPR countries. Relatedly, the difference between men and women is driven by individuals in the lowest deciles of pay or promotion: in low LFPR countries, the firm hires many men from the left tail of the ability distribution.

3.1 Hiring gap

First, we investigate whether the proportion of men and women in the firm correlates with the level of LFPR in the different labor markets. Figure 2 panel (a) documents a large variation in female shares of employment within the MNE across countries, which matches the variation in LFPR across countries: the female to male ratio in the firm closely follows the same ratio in the labor force, particularly at lower levels of LFPR (<0.6). When the LFPR is above 0.6, the firm hires more women relative to men compared to the LFPR, although a formal test of differential slopes by above/below median LFPR does not indicate that the relationship between the firm and country ratios significantly varies with the level of LFPR. We also note that wages at hiring are the same between men and women (the gender pay gap in starting salary is equal to -0.013 with a p-value of 0.209), as also confirmed by HR managers at the firm. Overall, this evidence is in line with the firm using the same selection process in all countries, rather than, for example, employing more pro-women policies in countries with low female labor force participation (or vice versa). In other words, the firm adopts the same headquarters personnel policies worldwide rather than adapting to local economic conditions - as broadly found in the literature on multinationals' wage and price setting procedures (DellaVigna and Gentzkow, 2019; Hjort et al., 2020). As a result, countries' LFPRs "bind"; that is, barriers at the country level constrain the firm's talent pool.

3.2 Pay gap

We turn to look at how the gender pay gap changes with the LFPR. We define the gender pay gap as the difference between women's wages and men's wages (in logs). Pay in the firm captures differences in performance between employees, encompassing off-peak salary increases as well as promotions. Figure 2 panel (b) estimates a kernel-weighted local polynomial of the pay gap on LFPR. The correlation is negative:

at low levels of LFPR, we observe an inverted pay gap, that is, women earn between 22% and 45% more than men; the gap falls as the LFPR increases, and it plateaus at around -10% when the LFPR reaches 0.8. This is close to the average gender pay gap of -16% in Europe (Commission, 2019). Overall, Figure 2 panel (b) documents a negative correlation between the gender pay gap and the LFPR.

Table 2 looks at the gender pay gap using different sources of variation, thus also allowing us to check whether the sign of the correlation is consistent regardless of the source of variation that we use to identify it. We estimate the following model:

$$w_{iact} = \alpha LFPR_{ac} + \beta Female_i + \gamma LFPR_{ac} * Female_i + \mathbf{X}'_{iact}\boldsymbol{\Lambda} + \boldsymbol{\psi}_t + \epsilon_{iact} \tag{1}$$

where w is log wage of employee i in country c and year t for age group a. $\boldsymbol{\psi}_t$ represents year fixed effects to take out year-level macro shocks and $\mathbf{X}_{iact}$ is a vector of controls. We cluster standard errors at the same level as the RHS variable, that is country-cohort. The coefficient of interest is γ which measures the change in the pay gap as LFPR increases. We include different controls in $\mathbf{X}_{iact}$: column 2 controls for a quadratic function of tenure and function fixed effects, column 3 adds GDP per capita in logs so to show that the variation in the LFPR is not only a function of country income, column 4 adds cohort fixed effects to exploit only the variation across countries; column 5 replaces the cohort fixed effects with country fixed effects hence only exploiting the variation within countries. The comparison between columns 4 and 5 is particularly informative as it uses one source of variation at a time.

The estimates of γ are negative and precisely estimated in all specifications. In Online Appendix Table A.1, we use the LFP data for individuals with advanced education only.[12] The results are nearly unchanged when we adopt this measure. However, we lose almost 20% of the sample, particularly from countries with low FLFP. Hence, we employ the overall LFP as our default measure.[13]

These gender differences are not present at the hiring stage - the gender pay gap in starting salary is equal to -0.013 (p-value=0.209) - backing up the interpretation of the adoption of a global personnel policy set at headquarters as seen in column 6 of Table 2. However, we next look at the pay progression for new hires. We do this in columns 7 and 8 by estimating the same specification as in column 1 for new hires only. Women in low LFPR country-cohorts display faster pay growth and a higher probability of promotion, and this positive gender gap in realized performance decreases as the LFPR increases. This evidence on the sample of new hires supports the inter-

[12] As defined by the World Bank, an individual with advanced education has completed a short-cycle tertiary education or a college degree and/or above.

[13] The correlation between overall LFPR and LFPR for individuals with advanced education is 68%.

pretation that the firm is not taking into account the implications on productivity that different LFP rates by gender might entail when making decisions on hiring and wage offers. It is only after some time at the firm that women and men *ex-post* pay gaps start to diverge the lower the LFPR is, indicating positive selection of women into the firm which is not accounted for at the hiring stage.

This perhaps counterintuitive inverted pay gap can be understood through the logic of positive selection. In countries with low LFPRs, there are higher barriers to entry into the labor force for women, and only the most talented women overcome them. As a result, the lower the LFPR, the more positively selected is the pool of women hired by the MNE relative to men. Thus, the average productivity of women in the workplace will be negatively correlated with their share in the labor force, as is standard in models of selection on gains (Roy, 1951). In other words, high barriers mean that only the most productive women are likely to be found in the workplace. As these barriers diminish, the average productivity of women entering the workforce also decreases. Building on this reasoning, we would expect women to be disproportionately represented at the upper echelons of the organizational hierarchy and to receive promotions more frequently than their male counterparts in countries with a low LFPR. We examine this hypothesis further in the following subsection.

Robustness We conduct additional robustness checks in the Online Appendix and the γ estimate is stable throughout. Table A.2 in the Online Appendix shows that the patterns in Table 2 hold if we add fixed effects for the geographical region and if we split the sample by lower income and higher income countries (as defined by the World Bank). In Online Appendix Table A.3, we report the results when converting wages from euros into PPP 2017 $ using the PPP conversion rates of the ICP at the World Bank. The gender gap is unaffected, and the only change is the magnitude of the coefficient on LFPR, which shrinks to the level found when controlling for country fixed effects (column 5 in Table 2). This is what we would expect as differences in PPP exchange rates would not affect cross-country comparisons of the gender gap.

Separately, a potential concern is that wages depend on factors other than individual productivity. In particular, the value of women to the firm may be driven by the value of diversity—bringing in a "woman's perspective"—which would lead us to overestimate women's productivity and, if there are spillovers, also men's. To assess this, in Online Appendix Table A.4 we show that the results are also robust to controlling for team gender composition.[14] The estimate of γ remains negative and similar

[14] We construct leave-one-out measures of the share of women in each worker's sub-function, country, and month, and classify teams as high female share using either a 30 percent threshold or the sample median.

in magnitude in both high- and low-female-share teams, while the triple interaction between Female, LFPR, and high female share is small and statistically insignificant across outcomes. These findings suggest that team gender composition is unlikely to explain the relationship between female labor force participation and gender differences in pay and career outcomes.

Finally, we note that the results are driven by differences in fixed pay rather than in variable bonus, which constitutes a much lower proportion of overall salary (the median ratio of bonus to fixed pay is 13%) — see Online Appendix Table A.5. In the company, pay summarizes altogether most differences in performance between employees, encompassing off-peak salary increases as well as promotions.

3.3 Promotion gap

Figure 3 displays plots of women's performance measures against the LFPR. Panels (a) and (b) of Figure 3 show that women are over-represented at the top decile of the wage distribution and under-represented at the bottom decile when overall LFPR is low and converge when LFPR is close to 1. Moreover, Panels (c) and (d) show that when LFPR is low, women are over-represented in managerial positions and among those who get promoted quickly, but the two converge as participation rates get more equal. In Online Appendix Figure A.3, we show the plots with the men's performance measures, which display the opposite patterns: men's performance is positively correlated with the LFPR.

A Roy selection framework predicts that as the LFPR increases, women with lower ability enter the labor force while high-ability women already in the labor force are unaffected. We test this result empirically in Figure 4 by looking at how women's wages at different points of the distribution change as the LFPR increases. Since overall wage levels change across countries, we control for the respective wage measure for men. The first panel from the left shows that women's average wages decrease as the LFPR increases. The remaining two panels indicate whether this decrease is driven by the bottom or the top of women's wage distribution. It comes from the bottom of the wage distribution: the 10th percentile of women's wages decreases with LFPR (second panel), while there is no impact on the 90th percentile (third panel).[15] These differential patterns at different levels of the wage distribution are consistent with women facing a higher bar for entering the firm in countries with lower LFPR. As the LFPR increases, lower-ability women start to enter, hence decreasing the wages at the bottom of the distribution, while high-ability women are not affected, hence

[15] Results are robust to using other percentiles, for example, the 25th percentile.

leaving the wages at the top of the distribution unchanged.

The fact that the negative correlation is driven by the left tail of the wage distribution helps distinguish the selection mechanism from alternative explanations that would also generate a negative correlation between average female performance and LFPR. For instance, this pattern rules out that the high productivity of women in low LFPR countries, and hence their high wages, is because women bring different valuable inputs to the firm, and therefore, the marginal value of these inputs is high when the share of women is low. If this were the case, we would find that the productivity of the top percentiles of women would decrease as LFPR increased, which is the opposite of what we find in Figure 4. It also rules out the argument that women's performance increases relative to men's because men in low LFPR countries have better outside options.[16]

Taken together, the evidence is consistent with positive selection into the labor force. That is, the average productivity of women who work outside the home is higher the higher the share of women who work inside the home. This implies that the firm could increase productivity by increasing their employees' gender ratio above that observed in the labor force. Moreover, the potential productivity gains would be higher where the female share in the labor force is lower. Seen through the lens of selection, the benefits of gender equality inside the firm are higher when gender equality outside the firm is lower. Yet, panel (a) of Figure 2 shows that the firm does not hire more women relative to their population share. This suggests that the cost of hiring more women is also higher when the gender ratio in the labor force is lower and hence that the net benefit is negative. These costs cannot be identified in the data because we only observe the state of the world in which they are not incurred. In what follows we combine theory and individual-level data to estimate the parameters of the firm's pay policy. We will then use these to create the counterfactuals we need to evaluate the effect of alternative gender policies.

4 Model and estimation

The correlations in section 3 are consistent with women employed by the firm being positively selected on unobserved productivity where female labor force participation is low. This suggests a trade-off between efficiency and equality: the firm could raise the productivity of its workforce by employing more women, but attracting more able

[16]Besides being inconsistent with the distributional changes, the assumption that men's outside option is decreasing in LFPR is inconsistent with the fact that multinationals pay higher wages and offer more amenities across the board (e.g., Verhoogen, 2008; Javorcik, 2015; Hale and Xu, 2016; Alfaro-Urena et al., 2022).

women requires paying them more, which widens pay inequality between genders.

Quantifying that trade-off requires separating the firm's wage schedule from the distribution of productivity in the pool from which it hires; observed pay depends on both, and the reduced-form correlations do not identify them separately. To that end, this section specifies a simple model of participation and pay, shows which parameters are identified by the participation rate and the first three moments of the within-cell wage distribution, and reports the resulting calibration.

4.1 Framework

We model employment at the firm as a two-stage process. In the first stage, individuals decide whether to participate in the labor force, comparing the expected return to market work with the value of staying at home. In the second stage, employers hire from the pool of participants in the labor force, and we observe wages for those employed by the MNE.

Stage 1: Labor force participation. Consider a two-sector (Roy, 1951) model as formalized by (Borjas, 1987). The value of market work to individual i is the expected log-wage they anticipate across potential employers, which we take to be linear in productivity A_i:

$$\bar{y}_i^1 = \bar{\alpha}^1 + \bar{\beta}^1 A_i. \tag{2}$$

Here $\bar{\alpha}^1$ is the expected return to participating in the labor market for the average worker, and $\bar{\beta}^1$ the expected return to productivity. We do not observe $\bar{y}_i^1$ directly; instead, we observe realized wages only for workers employed at the MNE.

The value of staying out of the labor force is, symmetrically,

$$y_i^0 = \alpha^0 + \nu^0 N_i, \tag{3}$$

where α^0 is the average value of home production (capturing, for instance, norms that apply to all women in a given society) and N_i is individual heterogeneity around it.

We assume joint normality,

$$\begin{bmatrix} A_i \\ N_i \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right),$$

so that ρ is the correlation between an individual's productivity at work and their value of staying at home. Individual i participates if and only if $\bar{y}_i^1 \geq y_i^0$. Let

$$\sigma \equiv \left[(\bar{\beta}^1)^2 + (\nu^0)^2 - 2\rho\,\bar{\beta}^1\nu^0\right]^{1/2}$$

denote the standard deviation of the gain from participating. The participation condition is then a threshold crossing:

$$\eta_i \;\equiv\; \frac{\bar{\beta}^1 A_i - \nu^0 N_i}{\sigma} \;\geq\; \frac{\alpha^0 - \bar{\alpha}^1}{\sigma} \;\equiv\; \xi. \tag{4}$$

Since $\eta_i \sim \mathcal{N}(0,1)$, participation occurs with probability $1 - \Phi(\xi)$, and a higher α^0 (due to, for instance, stronger norms about gender roles) raises the participation threshold ξ and lowers participation.

The same threshold determines the composition of the participating pool:

$$\mathbb{E}[A_i \mid \eta_i \geq \xi] = \frac{\bar{\beta}^1 - \rho\,\nu^0}{\sigma}\,\lambda(\xi), \tag{5}$$

where $\lambda(\cdot) \equiv \phi(\cdot)/[1 - \Phi(\cdot)]$ is the inverse Mills ratio. As long as $\bar{\beta}^1 - \rho\,\nu^0 > 0$ (that is, as long as productivity is rewarded more in the market than at home) average productivity among labor force participants is increasing in the barrier ξ. This is the mechanism behind the reduced-form patterns: the higher the bar, the more able the marginal participant, and the more positively selected the pool from which employers hire.

Stage 2: Employment at the MNE. We assume that, within a gender-country-cohort cell, the MNE draws a representative sample of candidates from the pool of participants in the labor force and makes hiring decisions before observing productivity A_i.[17] Entry into the labor force is therefore the selection margin that generates differences in the productivity of the firm's workforce across countries and cohorts.

Three features of the setting support this assumption. The firm's female-to-male employment ratio tracks the country participation ratio closely (Figure 2, panel a); the gender gap in starting salaries is economically and statistically zero (Table 2, column 6); and the firm applies centrally determined hiring and pay policies across all countries (DellaVigna and Gentzkow, 2019; Hjort et al., 2020).

Productivity is revealed once workers are employed, and for them we observe real-

[17]The condition we actually require is weaker: that any selection into the firm on productivity does not differ by gender within a country-cohort cell, since what we estimate is the gender difference in selection.

ized log-pay, which we model as linear in productivity with firm-specific parameters:

$$y_i^1 = \alpha^1 + \beta^1 A_i. \tag{6}$$

In terms of notation, we use bars for the Stage 1 objects that govern participation ($\bar{\alpha}^1, \bar{\beta}^1$) and no bars for the MNE's wage policy (α^1, β^1), which is what the data identify.

We view (6) as a reduced-form relationship between pay and productivity, and take no stance on where its parameters come from. Outside options, bargaining, differences in labor-supply elasticities across groups, and the wage markdowns implied by monopsony power all act on pay through these two parameters: whatever is common to a gender-country-cohort cell is absorbed into α^1, and whatever changes how steeply pay rises with productivity is absorbed into β^1. What the model requires is only that, within a cell, higher-productivity workers are paid more.

4.2 Calibration

We let all parameters vary by gender (g), cohort (a), and country (c). This flexibility is what allows the estimates to absorb confounders: we impose no assumption that the firm pays men and women alike, or that it applies the same schedule across countries. The one object we hold fixed across cells is the distribution of (A_i, N_i) assumed in 4.1.

Imposing the same distribution of (A_i, N_i) on every cell is what allows us to make comparisons across them. A difference in the average productivity of the employed can be read as a difference in selection only if the pool being selected from is the same; otherwise it could equally reflect a difference in the underlying distribution. This matters most for the comparison the paper is about.

Even granting that innate talent is distributed identically at birth, men and women may face different opportunities to acquire human capital, so that realized productivity distributions differ. Two features of the setting limit the scope for this. Gender differences in educational attainment have narrowed sharply and are small relative to the gap in participation (Online Appendix Figure A.4), and the firm hires only college graduates, roughly 70% of them in business administration or engineering, so the comparison is within a narrow educational band.

Identification. Theorem 6 in Heckman and Honoré (1990) characterizes what is identified in a two-sector log-normal Roy-Borjas model when only wages in one sector and the share of workers in that sector are observed. Under our assumptions, participation and the observed wage distribution satisfy

$$LFP_{gac} = 1 - \Phi(\xi_{gac}) \tag{7}$$

$$\mathbb{E}[y^1_{igac} \mid \text{empl'd}] = \alpha^1_{gac} + \theta_{gac}\lambda(\xi_{gac}) \tag{8}$$

$$\text{Var}(y^1_{igac} \mid \text{empl'd}) = (\beta^1_{gac})^2 + \theta^2_{gac}\left[\xi_{gac}\lambda(\xi_{gac}) - \lambda(\xi_{gac})^2\right] \tag{9}$$

$$\begin{aligned}&\mathbb{E}[(y^1_{igac} - \mathbb{E}[y^1_{igac} \mid \text{empl'd}])^3 \mid \text{empl'd}] \\ &\quad = \theta^3_{gac}\left[2\lambda(\xi_{gac})^3 - 3\xi_{gac}\lambda(\xi_{gac})^2 + \xi^2_{gac}\lambda(\xi_{gac}) - \lambda(\xi_{gac})\right]\end{aligned} \tag{10}$$

where ξ_{gac} is the participation threshold and θ_{gac} captures the sign and strength of selection on productivity. A positive θ_{gac} means labor force participants are positively selected; a value near zero means little selection; and a larger θ_{gac} for women than for men means women are more strongly positively selected into the labor force. The composite expressions defining both parameters are in Online Appendix section B. In this model, selection is identified from the spread and the shape of the within-cell wage distribution rather than from its level: this is what individual personnel records make possible over aggregate wage data by group.

Notice that the parameters governing the value of staying at home, α^0_{gac}, ν^0_{gac}, and ρ_{gac}, are not separately identified, since we do not observe that value. The estimates of calibrated productivity below do not require them. The counterfactuals in the next section do, and there we need one further normalization, since two moments must pin down three parameters. We detail two alternative strategies in Online Appendix section B and show that the estimates are almost unchanged across them.

Implementation. Online Appendix Table 3 summarizes each parameter and the empirical moment it is matched to. For LFP_{gac} we use World Bank participation rates in the matching gender-cohort-country cell. For the mean, variance and skewness of y^1_{igac} in the firm we use the corresponding sample moments of residualized log-pay, where the residual comes from a regression of log(base pay + bonus) on year and function dummies, tenure and tenure squared. Residualizing on function means the estimates are identified within job families rather than across them, so the occupational composition of the firm's workforce is not driving the results. We exclude cells with fewer than 30 male or 30 female employees, leaving 260 cells across 84 countries. Online Appendix Figure A.5 shows that normality fits the observed wage distributions reasonably well, except where the data are markedly non-normal (for instance bimodal); this is a check on a distributional assumption that is not itself targeted in the calibration.

As a robustness check, in Online Appendix section D we recalibrate the model using participation among the tertiary educated, which restricts the comparison to a more homogenous pool of workers in terms of education. The estimates and counterfactual conclusions are unchanged. While this proxy is closer to the firm's relevant hiring pool given the educational requirements, we keep overall participation as the baseline because the tertiary-educated series is missing for 56 country-cohort cells disproportionately from low female LFP countries.

4.3 Structural estimates

In this subsection, we report the structural estimates and correlate them with other data not used in the estimation to facilitate their economic interpretation.

Summary statistics of the calibrated parameters are presented in Panel A of Table 4. On average across countries and age cohorts, we estimate fixed pay α^1_{gac} to be higher for men and returns to productivity β^1_{gac} to be higher for women, albeit with a lot of variation across cells. The labor force participation thresholds, ξ_{gac}, are higher for females, as a direct consequence of having lower LFP than their male counterparts. Finally, we estimate the selection with respect to the productivity parameter, θ_{gac}, to be positive on average, although it is also very heterogeneous across cells, including cells where it is negative.

Panel B of the same table summarizes the parameters of the value of staying out of the labor force, which are identified under an additional normalization $\nu^0_{gac} \equiv 1$. We exclude country-cohort cells for which no solutions or multiple solutions to the calibration moment equations exist, and cells for which we cannot compute the main set of counterfactuals discussed in section 5. We find that women have stronger average values of staying out of the labor force, α^0_{gac}, again consistent with having a lower LFP than their male counterparts. The correlation between the wage in the firm and the value of staying out of the labor force is similar on average for males and females, although with a lot of variation across cells.

4.3.1 Calibrated ability

Having obtained the parameters in the firm's wage policy, we can recover model-implied productivity as:

$$A_i = \frac{y^1_{igac} - \alpha^1_{gac}}{\beta^1_{gac}},$$

where y^1_{igac} is the residualized log-wage described before. Because we observe wages in the firm, we can impute productivity even though α^0_{gac}, ν^0_{gac} and ρ_{gac} are not separately identified, because inverting the wage equation uses only firm-side parameters. Under the normalization of $A_i \sim N(0,1)$, differences in productivity will be measured in units of standard deviations of the underlying productivity variable, which enters the log-wage equation.

Figure 5 plots the average calibrated productivity by the LFPR. Average productivity is approximately constant for male workers in most of the sample, whereas for female workers, average productivity is very high when FLFP is much smaller than MLFP (about 0.4 standard deviations higher than for men) and decreases as the LFPR approaches 1. This result is consistent with selection with respect to productivity, so the lower a group's LFP, the more positively selected they are. Figure 6 further shows that this is due to a shift in the entire productivity distribution of women. As the LFPR increases, there is a downward shift of the entire productivity distribution, except the right tail. The observation that changes occur in the left tail of the distribution aligns with the findings presented in section 3, which show that the ability of the marginal female worker decreases as the LFPR increases.

4.3.2 Validation

To validate our estimates we show that they correlate with other data not used to calibrate the model. We use two separate external variables to corroborate our calibrated productivity estimate using individual performance data from the firm's records. The results for this exercise are in Figure 7. Panel (a) correlates calibrated productivity against the firm's performance score that a manager gives every year, and panel (b) against pay growth in the first year (for new hires). The correlation of both indicators with our calibrated productivity is positive and strong.

In Online Appendix Figure A.6, we further assess whether the relationship between calibrated productivity and performance evaluations is similar across gender, parental leave status, and cohorts. Across all groups, higher calibrated productivity predicts higher performance evaluations, suggesting that the recovered measure captures meaningful variation in worker productivity rather than systematic differences in compensation or labor supply incentives. For example, if parental leave primarily affected wages rather than productivity, we would not expect the calibrated measure to predict performance evaluations similarly for workers with and without parental leave. However, the relationship is nearly identical across groups.

4.3.3 Interpretation

To aid the interpretation of the structural estimates, we show that the parameters of the home payoffs and the firm wage policy are correlated with country-level social norms and country-level labor laws, respectively.

Figure 8 panel (a) plots the gender gap in calibrated home value, $\alpha^0_{Fac} - \alpha^0_{Mac}$, against four World Values Survey items: "Men make better business executives than women do," "Pre-school child suffers with working mother," "Being a housewife is just as fulfilling as working," and "When jobs are scarce, men should have more right to a job than women." For all four, disagreement with the statement is strongly negatively related to the gap, with slopes of at least -0.4 and p-values below 0.1. Panel (b) plots the same gap against the World Bank's Women, Business and the Law Index[18] and finds a strong negative relationship, indicating that part of what restricts female participation is codified in law rather than only in norms. Together these support reading $\alpha^0_{Fac} - \alpha^0_{Mac}$ as the cost that prevailing norms impose on the average woman who works. The figure shows that the gap $\alpha^0_{Ftac} - \alpha^0_{Mtac}$ is strongly negatively correlated with laws allowing or facilitating women's labor. Therefore, part of the restrictions to FLFP due to gender norms may be embedded in the laws of certain countries.

Figure 8 panel (c) plots calibrated β^1_{gac}, which generates within-cell pay dispersion, against the World Economic Forum's Restrictive Labor Regulations Index.[19] Returns to productivity are lower where regulation is more restrictive, consistent with such regulation limiting the use of performance pay. This correlation is what motivates the labor-regulation counterfactual in subsection 5.4.

5 Counterfactuals

We use the estimates to ask four questions. Does the firm's observed pay policy maximize the productivity of its workforce, and if not, what would achieving it require? How much of that is attainable while equalizing the wage schedule between men and

[18]The WB Women, Business and the Law Index covers 190 countries through the period 1971–2020 and is structured around the life cycle of a working woman. It consists of eight indicators constructed around women's interactions with the law — mobility, pay, workplace, marriage, parenthood, entrepreneurship, assets, and pensions — for current laws and regulations (i.e. religious and customary laws are not considered unless they are coded). Hyland, Djankov and Goldberg (2021) provides an overview of the data documenting how gender discrimination by law affects women's economic opportunities.

[19]The WEF Restrictive Labor Regulations Index is available for the period 2008–2020, and it is based on an annual survey of the most problematic factors for doing business (e.g. corruption, taxes, inflation, etc.). The survey is administered to a representative sample of around 15,000 business executives in 150 countries. The Restrictive Labor Regulations Index includes measures related to labor-employer relations, wage flexibility, hiring and firing practices, performance pay, labor taxes, attraction and retention of talent.

women? How much of the firm's productivity is attributable to the unequal barriers women face outside it? And what happens when regulation limits the firm's ability to link pay to productivity?

All four work through the two-stage structure of subsection 4.1: individuals first decide whether to participate in the labor force, and the firm then hires from the resulting pool. Where participation is lower, the pool is more strongly selected and the workers available to the firm are more able on average. Each exercise below changes something that moves participation—the firm's pay schedule, the barriers women face entering the labor force, or a regulatory cap on performance pay—and the productivity of the firm's workforce responds because it is drawing from a differently selected pool.

The policies we compute below are optimal only within the narrow problem we impose on the firm, and we read the exercises as constrained thought experiments rather than as a description of how it would behave. That problem is narrow in three respects. First, we do not observe the firm's production function or the elasticity of demand it faces, so we cannot ask what maximizes profit; we ask instead how productive a workforce the firm could assemble without spending more on wages or employing fewer people than it does. Second, pay is not the firm's only instrument in practice, and we hold the others fixed: recruiting intensity, screening, mentoring, training and workplace flexibility. Third, adjustment operates only through the labor force participation margin, so in the model the firm attracts women into the labor force rather than hiring women who already work elsewhere. We therefore view the results as informative about the direction and shape of the trade-off between productivity and inequality rather than as point predictions about its magnitude.

5.1 Productivity-maximizing wage policy

Our first counterfactual asks how productive the firm's workforce could be if the firm used its wage schedule to change the pool of workers in the labor force. We cast the problem as one of allocating given resources: holding fixed the wage bill it spends and the number of workers it employs in each country-cohort cell, the firm chooses the pay schedule that delivers the most productive mix of workers those resources support.

Formally, the firm chooses α^1_{gac} and β^1_{gac} to maximize $LFP \cdot \mathbb{E}[A_i \mid \text{empl'd}]$ subject to a wage bill no greater and employment no smaller than at baseline. This is equivalent to assuming that the firm first sets its scale of operation, by a process we do not model, and then allocates that budget across workers at that scale. Both constraints are what make this an allocation problem rather than an expansion one: without the wage bill fixed the firm raises the ability of its workforce by paying more, and without the employment floor it raises average ability simply by employing fewer people, since selection is positive. Full expressions are in Online Appendix section C.

Our counterfactuals require that the firm's wage policy be able to move participation, since otherwise a change in the schedule would alter pay without altering the pool the firm hires from. We therefore let it do so, setting $\bar{\alpha}^1_{gac} = \alpha^1_{gac}$ and $\bar{\beta}^1_{gac} = \beta^1_{gac}$, which may be understood as treating the MNE as representative of the labor market its workers face. Because the firm's schedule now moves participation, the program also requires the home-value parameters that subsection 4.2 left unidentified; throughout we use the $\nu^0_{gac} \equiv 1$ normalization, and Online Appendix section B shows that the alternative leaves the estimates almost unchanged. Under this normalization we can recover α^0_{gac} and ρ_{gac}, and hence solve the program, in 114 of the 260 cells in the structural sample, covering 67 countries.[20]

Online Appendix Figure A.7, panels (a) and (b), compare the observed wage policy parameters with the solution to this program. The two do not coincide, and in some countries they are far apart. The direction of the gap is the same nearly everywhere: to raise productivity the firm should attract more women, exploiting their stronger positive selection. It does so by lowering fixed pay for men and raising it for women, while generally steepening the return to productivity, except where the employment or wage-bill constraint binds. Comparing rows 1b and 2b of Table 5, the female-to-male employment ratio rises from 0.71 to 0.92 and average productivity from 0.31 to 0.47 $[0.37, 0.51]$ standard deviations, a gain of 0.16 standard deviations (95% bootstrap CI for the difference: $[0.153, 0.235]$) that is an increase of around 50% relative to baseline. Productivity rises in every cell (Figure A.7, panel e).

Why does the firm not adopt this policy? One possible answer is that the productivity-maximizing wage policy would generate stark inequality between genders. Since women in the labor force are more positively selected in most cells, productivity maximization calls for paying women more in order to attract more of them, so that the pay gap (female minus male) would be even larger than what we observe. Averaging across cells, it would move from -0.10 to 0.63, an increase of 73 log points (Figure A.7, panel d). Such pay gaps may not be socially acceptable, nor would they be evenly distributed: pay inequality would rise most in the countries where barriers against women working are highest, so the firm would depart most from the local norm exactly where that norm is most restrictive.

In this exercise the firm adjusts pay alone, while we hold fixed its other instru-

[20] In some cells the calibration admits no solution with $\rho_{gac} \in [-1, 1]$, and in others it admits multiple solutions; we drop both. Of the remainder, all the counterfactual wage policies reported in Table 5, Panel B, can be computed in 114 cells.

ments: recruiting intensity, screening, mentoring, training and workplace flexibility. Since these could substitute for pay in attracting women, the 73 log points overstate the increase in pay inequality needed to obtain a productivity gain of this size. The firm also adjusts only along the labor force participation margin, drawing women out of the home rather than away from other employers. Since high-ability women already working elsewhere would be cheaper to attract, the 0.16 standard deviations understate the productivity gain available at a given cost.

5.2 Equal pay for equal ability

We next ask what the firm can achieve while paying men and women on the same schedule, by re-solving the problem with the added restriction $\alpha^1_{Fac} = \alpha^1_{Mac}$ and $\beta^1_{Fac} = \beta^1_{Mac}$ for all cohort-country cells, so that two workers of equal ability are paid equally regardless of gender. This is computable in 88 cells covering 55 countries; the binding obstacle is that once the schedules are tied, some cells admit no policy satisfying the employment, wage-bill and equality constraints at once.[21]

The changes go in the same direction as before but are smaller. The firm still raises fixed pay for women and lowers it for men, and still steepens variable pay (Figure A.8, panels a and b). The employment ratio rises from 0.71 to 0.78 (rows 1c and 6c), against 0.93 when the schedules may differ by gender. Productivity rises from 0.39 to 0.46 $[0.32, 0.51]$ standard deviations, a gain of 0.07 standard deviations (95% bootstrap CI: $[0.040, 0.114]$) or 18% of baseline. This gain is about 60% as large as the 0.12-standard-deviation gain available when the firm can set separate wage schedules by gender (row 2c).

Equalizing the schedule does not, however, equalize pay. The gender pay gap still rises by 24 log points, from -0.10 to 0.14, because women in the firm are more productive on average and a common schedule rewards that difference rather than removing it. Equal pay for equal ability is therefore not the same thing as equal pay on average because of differential selection into the labor force.

The two exercises together delimit which notions of equality are compatible with efficiency in this setting. A single schedule applied to both genders is compatible with about 60% of the available productivity gain, but not with equal average pay. Equality of average pay is not attainable through wage policy at all while the underlying barriers differ, because the productivity distributions the firm draws from are not the same. The next subsection shows that it becomes attainable, at no cost in productivity, once the barriers themselves are equalized. Constraints on the firm's pay schedule trade off

[21] Because this restricts the sample, we compare row 6c against the baseline and the optimum of 5.1 recomputed on the same 88 cells, rows 1c and 2c, rather than against Panel B.

against efficiency; removing the constraints women face outside the firm does not.

5.3 Equal outside options

We now ask how much of the pattern is attributable to the barriers themselves, by replacing women's home-value parameters with men's within each country-cohort cell: $\alpha^0_{Fac} = \alpha^0_{Mac}$, $\nu^0_{Fac} = \nu^0_{Mac}$ and $\rho_{Fac} = \rho_{Mac}$. We consider the short run, holding the firm's pay policy at baseline, and the long run, when the firm can adjust its policy optimally to the new environment.

In the short run (Online Appendix Figure A.9, and rows 1b and 3b of Table 5), male participation is unchanged by construction while female participation rises in most cells, so the employment ratio rises from 0.71 to 0.92. The pay gap is close to flat, moving from -0.10 to -0.12, because the women drawn in are less able than those already working and women's average pay falls slightly. The effect on productivity is theoretically ambiguous: the firm averages over a larger pool, raising LFP, but conditional on participating women are less positively selected, lowering $\mathbb{E}[A_i \mid \text{empl'd}]$. Empirically the second effect dominates and productivity falls from 0.31 to 0.28 $[0.11, 0.34]$ standard deviations. Equalizing the barriers women face does not by itself raise the productivity of the firm's workforce, so long as the firm's pay policy stays where it is. That is what subsection 5.1 would lead us to expect. The baseline policy is not the one that maximizes productivity given the pool the firm already draws from, and enlarging that pool does not help unless the firm also changes how it pays.

Allowing the firm to re-optimize confirms this and removes the trade-off. Once women's outside options match men's, the equal-pay restriction stops binding, rows 4c and 7c of Table 5 being identical, so the firm attains gender parity in employment and a zero pay gap at the same time (panels c and d of Figure A.10 and Figure A.11). Equalizing the outside option makes women cheaper to attract, which lets the firm lower fixed pay and steepen variable pay (panels a and b), drawing in higher-ability workers without breaching the employment or wage-bill constraints. Productivity reaches 0.55 $[0.45, 0.61]$ standard deviations in both cases. Relative to the best the firm could do with barriers as they are, the gain is 0.09 standard deviations (95% bootstrap CI: $[0.086, 0.175]$) when it is constrained to a common schedule (row 7c against row 6c), and 0.05 standard deviations (95% bootstrap CI: $[0.024, 0.071]$) when it is not (row 4c against row 2c). The trade-off between productivity and pay equality inside the firm is a consequence of unequal participation in the labor force: once the barriers to entry are equalized, the firm can pay men and women on the same schedule, employ them in equal numbers, and still be more productive than under any policy available to it today.

5.4 Regulatory limits on performance pay

A second reason the firm might not set the productivity-maximizing policy is regulation. Many pro-worker provisions limit how steeply pay can track productivity, effectively capping β^1_{gac}, and Figure 8, panel (c), shows that estimated returns to productivity are indeed lower where regulation is more restrictive. We ask what stricter regulation would cost by re-solving for the optimal policy with the added cap $\beta^1_{gac} \leq \max\{\beta^1_{Fa,\mathrm{FRA}}, \beta^1_{Ma,\mathrm{FRA}}\}$, using France because it sits at the 95th percentile of the WEF Restrictive Labor Regulations Index. Home-value parameters stay at baseline, but participation is free to respond to the new incentives.

Every effect goes in the expected direction. Comparing rows 2b and 5b of Table 5, productivity falls from 0.47 to 0.39 standard deviations, a decline of 0.08 standard deviations (95% bootstrap CI for the difference: $[-0.125, -0.075]$). The pay gap narrows from 0.63 to 0.57, by 6 log points. To keep satisfying the employment constraint with a capped gradient, the firm raises fixed pay in most cells, and more for women than for men, so the employment ratio rises from 0.92 to 0.99. Productivity falls in every cell (Online Appendix Figure A.12), so the sign is consistent even though the average magnitude is modest. The same comparison in the smaller Panel C sample, rows 2c and 5c, yields similar results: productivity falls from 0.51 to 0.46 standard deviations, a decline of 0.05 standard deviations (95% bootstrap CI for the difference: $[-0.104, -0.049]$); the employment ratio rises from 0.93 to 0.96; and the pay gap narrows from 0.61 to 0.54, by 7 log points.

The mechanism is that capping the gradient removes the instrument the firm was using to attract high-ability women, so the cost falls disproportionately on them and the compression of the pay gap comes from the top rather than the bottom. This illustrates a general drawback of pursuing gender equality by mandating identical treatment: when two groups are differently selected, treating them identically is not neutral between them.

6 Conclusion

This paper shows that the traditional division of labor, where men work in the market and women at home, affects firm productivity through the differential participation of women and men in the workforce. That is, when female labor force participation is low compared to men's, the few women who work outside the home are those with the highest productivity. We find that this inequity in the labor market could potentially positively impact firm productivity by bringing in highly talented women. However, the firm does not use it to its full advantage, as doing so would imply a large increase in within-firm inequality between genders; this is particularly so where gender disparities in society are the largest. Precisely where the productivity gains are highest are where the costs in terms of intra-organization inequality are largest. This tension between efficiency and equality limits the extent to which a single firm can gain by going against discriminatory norms (Becker, 1971).

Understanding differential selection by gender, or indeed by any under-represented group, is key to informing personnel policy as well as broader labor market policies aimed at addressing the gender pay gap. Our counterfactual estimates highlight the complementarity between public policies that attenuate gender barriers and firm policies to maximize productivity.

The inability of firms to compensate for labor market gender disparities influences how they design their policies at the hiring, pay, and promotion stages. Aiming for gender equity — in pay, promotions, and dismissals — can turn out to be inequitable because selection generates different distributions of productivity between genders. Perhaps counter-intuitively, gender equity policies within firms might end up hurting women as they limit the firm's ability to reward performance.

This paper focuses on firm wage and promotion policies and does not consider firm hiring strategies. Nonetheless, applying the principle of self-selection could address a frequent issue faced by firms: that potential productivity is not fully revealed at the point of hiring. In particular, by positive selection, under-represented groups should have higher productivity, other things equal. This implies that between two potential hires with the exact same observable qualities, the minority candidate has, on average, better unobservables. Awareness of the evidence of positive selection of under-represented groups could change how quality can be inferred, particularly if quality is not perfectly observable or objectively measured. If made aware of the "distance traveled" — that the very presence of a member of an under-represented group has information on the talent of that member — managers might change hiring decisions and screening criteria. The implications of increased awareness of the positive selection effect of minority candidates by leaders for the recruitment, promotion, and productivity of under-represented groups are left to future research.

7 References

Aghion, Philippe, Nicholas Bloom, Brian Lucking, Raffaella Sadun, and John Van Reenen, "Turbulence, firm decentralization, and growth in bad times," *American Economic Journal: Applied Economics*, 2021, *13* (1), 133–169.

Alatas, Vivi, Ririn Purnamasari, Matthew Wai-Poi, Abhijit Banerjee, Benjamin A Olken, and Rema Hanna, "Self-targeting: Evidence from a field experiment in Indonesia," *Journal of Political Economy*, 2016, *124* (2), 371–427.

Alesina, Alberto and Eliana La Ferrara, "Ethnic diversity and economic performance," *Journal of economic literature*, 2005, *43* (3), 762–800.

Alfaro-Urena, Alonso, Isabela Manelici, and Jose P Vasquez, "The Effects of Joining Multinational Supply Chains: New Evidence from Firm-to-firm Linkages," *The Quarterly Journal of Economics*, 2022, *137* (3), 1495–1552.

Andrew, Alison, Oriana Bandiera, Monica Costa-Dias, and Camille Landais, "Women and men at work," *IFS Deaton Review of Inequalities*, 2021.

Athey, Susan, Christopher Avery, and Peter Zemsky, "Mentoring and diversity," *American Economic Review*, 2000, *90* (4), 765–786.

Becker, Gary S, *The economics of discrimination*, University of Chicago press, 1971.

Bertrand, Marianne, "New perspectives on gender," in "Handbook of labor economics," Vol. 4, Elsevier, 2011, pp. 1543–1590.

_ , "Gender in the twenty-first century," in "AEA Papers and proceedings," Vol. 110 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2020, pp. 1–24.

_ **and Esther Duflo**, "Field experiments on discrimination," *Handbook of economic field experiments*, 2017, *1*, 309–393.

Blau, Francine D, *Equal pay in the office*, Lexington Books, 1977.

Blinder, Alan S and Alan B Krueger, "Alternative measures of offshorability: a survey approach," *Journal of Labor Economics*, 2013, *31* (S1), S97–S128.

Borjas, George J., "Self-Selection and the Earnings of Immigrants," *The American Economic Review*, 1987, *77* (4), 531–553.

Breza, Emily, Supreet Kaur, and Yogita Shamdasani, "The morale effects of pay inequality," *The Quarterly Journal of Economics*, 2018, *133* (2), 611–663.

Bursztyn, Leonardo, Alexander W Cappelen, Bertil Tungodden, Alessandra Voena, and David H Yanagizawa-Drott, "How Are Gender Norms Perceived?," *National Bureau of Economic Research*, 2023.

Card, David, Ana Rute Cardoso, and Patrick Kline, "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women*," *The Quarterly Journal of Economics*, 10 2016, *131* (2), 633–686.

Charness, Gary and Peter Kuhn, "Lab labor: What can labor economists learn from the lab?," *Handbook of labor economics*, 2011, *4*, 229–330.

Commission, European, "Report on gender pay gap in EU countries," 2019. `https://ec.europa.eu/info/policies/justice-and-fundamental-rights/gender-equality/equal-pay/gender-pay-gap-situation-eu_en#relatedlinks`, (accessed October 10, 2020).

DellaVigna, Stefano and Matthew Gentzkow, "Uniform pricing in us retail chains," *The Quarterly Journal of Economics*, 2019, *134* (4), 2011–2084.

Fehr, Ernst, Lorenz Goette, and Christian Zehnder, "A behavioral account of the labor market: The role of fairness concerns," *Annu. Rev. Econ.*, 2009, *1* (1), 355–384.

Goldin, Claudia, "The U-shaped female labor force function in economic development and economic history," 1995.

_ , "A grand gender convergence: Its last chapter," *American Economic Review*, 2014, *104* (4), 1091–1119.

Grossman, Gene M and Elhanan Helpman, "Fair wages and foreign sourcing," *Harvard University Press*, 2008.

Hale, Galina and Mingzhi Xu, "FDI effects on the labor market of host countries," *2016-25*, 2016.

Hamilton, Barton H, Jack A Nickerson, and Hideo Owan, "Diversity and productivity in production teams," in "Advances in the Economic Analysis of participatory and Labor-managed Firms," Vol. 13, Emerald Group Publishing Limited, 2012, pp. 99–138.

Hazell, Jonathon, Christina Patterson, Heather Sarsons, and Bledi Taska, "National wage setting," *University of Chicago, Becker Friedman Institute for Economics Working Paper*, 2022, (2022-150).

Heckman, James and Bo Honoré, "The Empirical Content of the Roy Model," *Econometrica*, 1990, pp. 1121–1149.

Hjort, Jonas, "Ethnic divisions and production in firms," *The Quarterly Journal of Economics*, 2014, *129* (4), 1899–1946.

_ , Xuan Li, and Heather Sarsons, "Across-Country Wage Compression in Multinationals," Technical Report, National Bureau of Economic Research 2020.

Hong, Lu and Scott E Page, "Problem solving by heterogeneous agents," *Journal of economic theory*, 2001, *97* (1), 123–163.

Hsieh, Chang-Tai and Esteban Rossi-Hansberg, "The industrial revolution in services," *Journal of Political Economy Macroeconomics*, 2023, *1* (1), 3–42.

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

Hyland, Marie, Simeon Djankov, and Pinelopi Koujianou Goldberg, "Do Gendered Laws Matter?," 2021.

Javorcik, Beata S, "Does FDI bring good jobs to host countries?," *The World Bank Research Observer*, 2015, *30* (1), 74–94.

Jayachandran, Seema, "The roots of gender inequality in developing countries," *economics*, 2015, *7* (1), 63–88.

Lazear, Edward P, "Globalisation and the market for team-mates," *The Economic Journal*, 1999, *109* (454), 15–40.

—, "Why are some immigrant groups more successful than others?," *Journal of Labor Economics*, 2021, *39* (1), 115–133.

Lemieux, Thomas, W Bentley MacLeod, and Daniel Parent, "Contract form, wage flexibility, and employment," *American Economic Review*, 2012, *102* (3), 526–531.

Marx, Benjamin, Vincent Pons, and Tavneet Suri, "Diversity and team performance in a Kenyan organization," *Journal of Public Economics*, 2021, *197*, 104332.

Mulligan, Casey B. and Yona Rubinstein, "Selection, Investment, and Women's Relative Wages over Time," *The Quarterly Journal of Economics*, 2008, *123* (3), 1061–1110.

Olivetti, Claudia and Barbara Petrongolo, "Unequal Pay or Unequal Employment? A CrossCountry Analysis of Gender Gaps," *Journal of Labor Economics*, 2008, *26* (4), 621–654.

— **and** —, "The evolution of gender gaps in industrialized countries," *Annual review of Economics*, 2016, *8*, 405–434.

Propper, Carol and John van Reenen, "Can Pay Regulation Kill? Panel Data Evidence on the Effect of Labor Markets on Hospital Performance," *Journal of Political Economy*, 2010, *118* (2), 222–273.

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

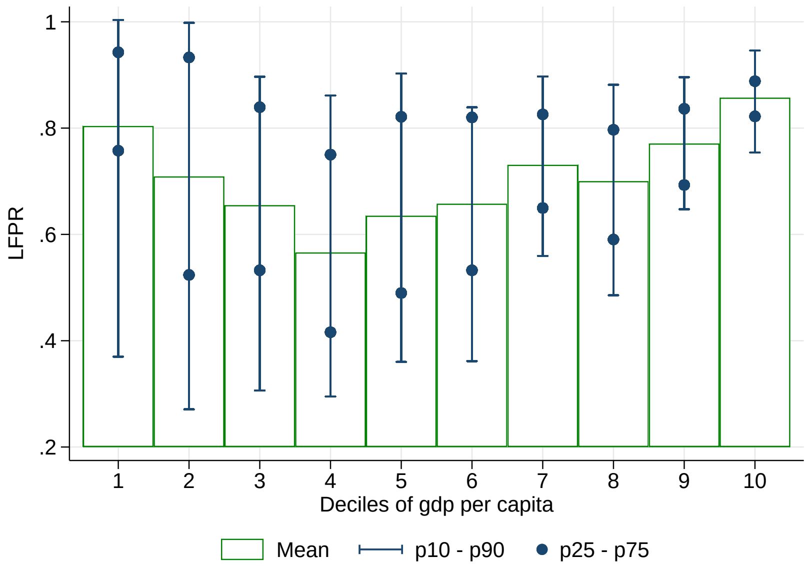
Sarsons, Heather, "Interpreting signals in the labor market: evidence from medical referrals," *Working paper*, 2017.

Verhoogen, Eric A, "Trade, quality upgrading, and wage inequality in the Mexican manufacturing sector," *The Quarterly Journal of Economics*, 2008, *123* (2), 489–530.

Wiswall, Matthew and Basit Zafar, "Preference for the Workplace, Investment in Human Capital, and Gender*," *The Quarterly Journal of Economics*, 08 2018, *133* (1), 457–507.

8 Figures

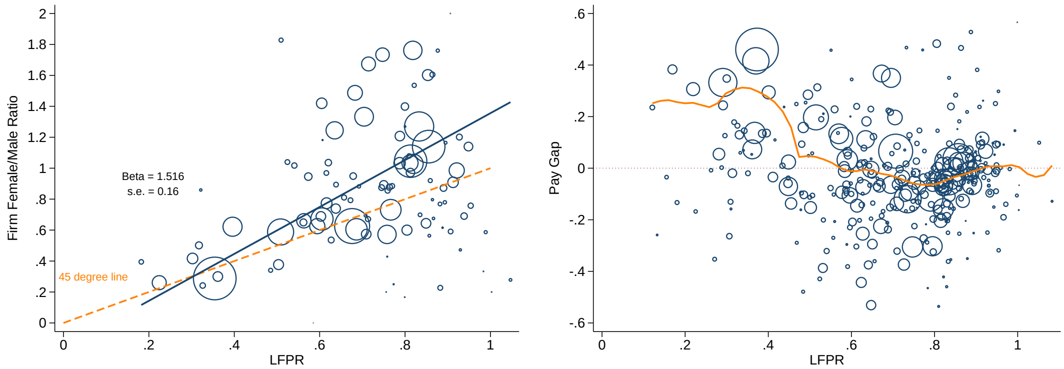
Figure 1: LFPR and GDP per capita



Notes. The figure plots the average, the interdecile range, and the interquartile range of the LFPR across deciles of GDP per capita across countries and cohorts. GDP per capita is at constant 2015 USD and is taken from the World Bank. The unit of observation is a country-cohort.

Figure 2: Gender ratio/pay gap and LFPR

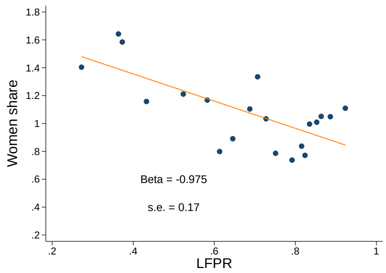
(a) Gender ratio gap **(b)** Gender pay gap



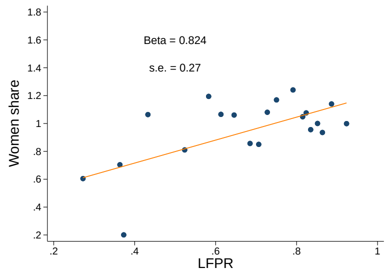
Notes. In panel (a), y-axis corresponds to the female/male employment ratio in the MNE while the x-axis corresponds to the LFPR in the countries. Each circle is a country and the size depends on the number of firm employees in the country. The orange line represents the 45-degree line. The unit of observation is a country. Panel (b) plots the gender pay gap (the difference between womens and mens wages in logs) against the LFPR. Each circle represents a country-cohort pair, whose size depends on the number of firm employees in the country-cohort cell. The orange line represents the smoothed values of a kernel-weighted local polynomial regression using analytical weights by employee size of each cohort-country cell.

Figure 3: Gender promotion gap and LFPR

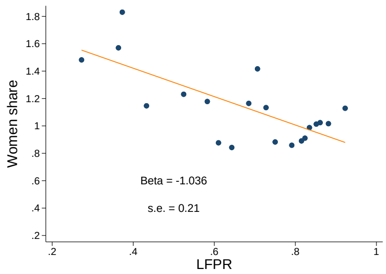
(a) Top decile of wages



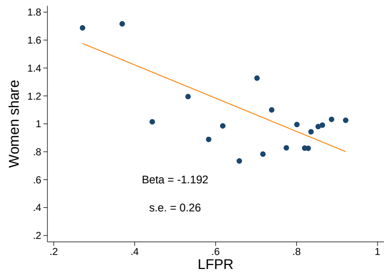
(b) Bottom decile of wages



(c) Managerial positions

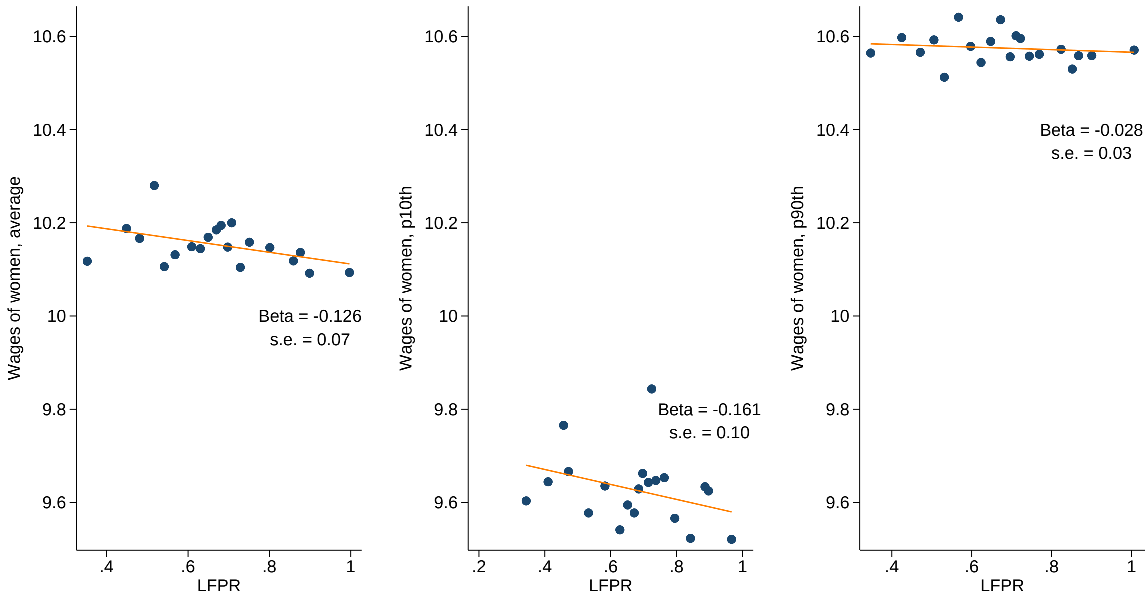


(d) High promotion speed



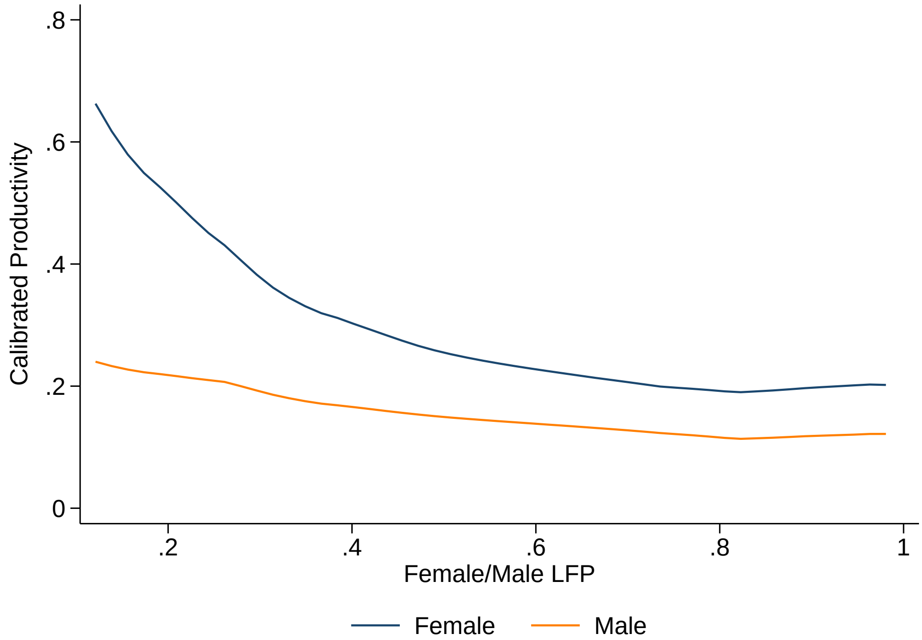
Notes. The figures are binned scatterplots and a linear fit of the share of women with different performance metrics (as a proportion of the female share in the firm) against the LFPR, in each country-cohort cell. Panel (a) looks at the top decile of wages; Panel (b) at the bottom decile; Panel (c) looks at the share of women in managerial positions; and Panel (d) at promotion speed to managerial positions, based on average promotion rates in the firm by labor market experience. In the regressions, we use analytical weights by employee size of each cohort-country cell. The unit of observation is a country-cohort.

Figure 4: Women's wages and LFPR



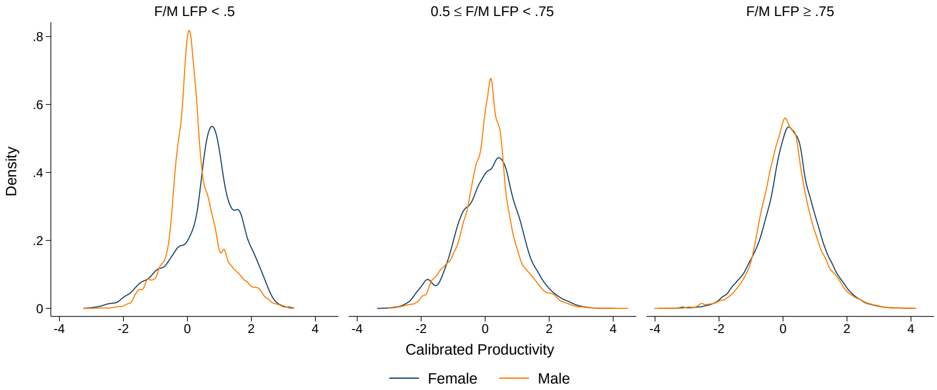
Notes. The figures are binned scatterplots and a linear fit of women's wages against the LFPR. The first figure from the left plots women's average wages, the middle one plots the 10th percentile, and the last one plots the 90th percentile. In the regressions, we control for the respective measure of men's wages, and we use analytical weights by employee size of each cohort-country cell and robust standard errors. The unit of observation is a country-cohort.

Figure 5: Calibrated productivity by gender and LFPR



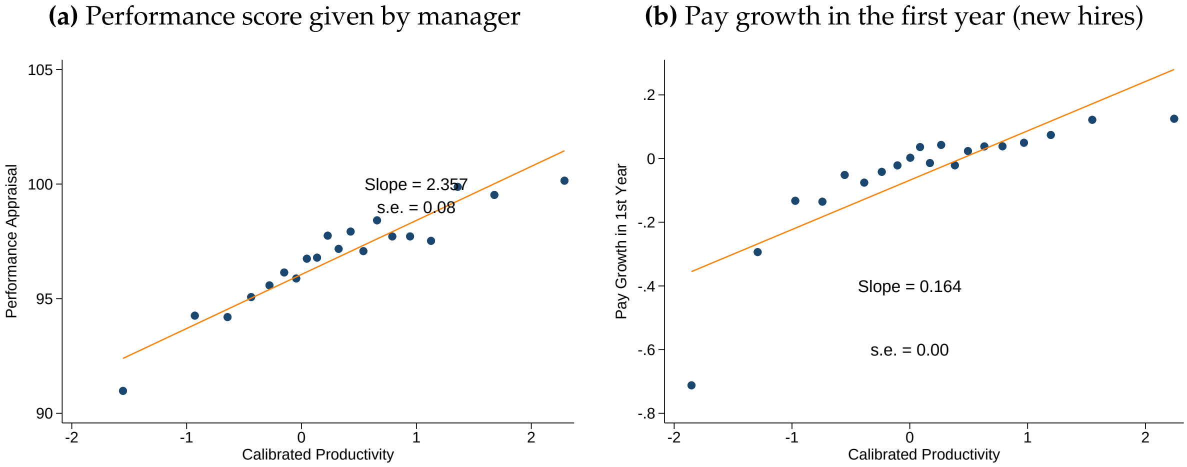
Notes. The figure plots the average calibrated productivity for our sample of firm workers by Female/Male LFP, smoothed through a local linear regression. The unit of observation is a worker.

Figure 6: Calibrated productivity distribution by gender and LFPR



Notes. The figure plots a kernel density estimate of calibrated productivity for our sample of firm workers by three LFPR groups: $[0, .5)$, $[.5, .75)$ and $[.75, 1]$. The unit of observation is a worker.

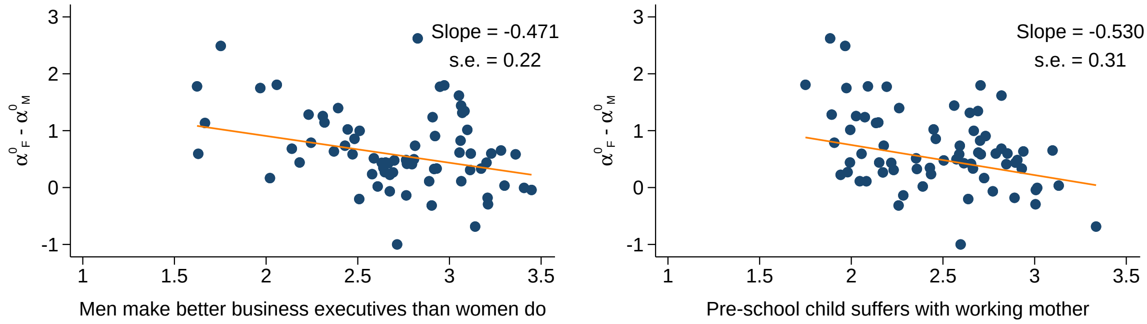
Figure 7: Calibrated productivity and non-targeted performance measures



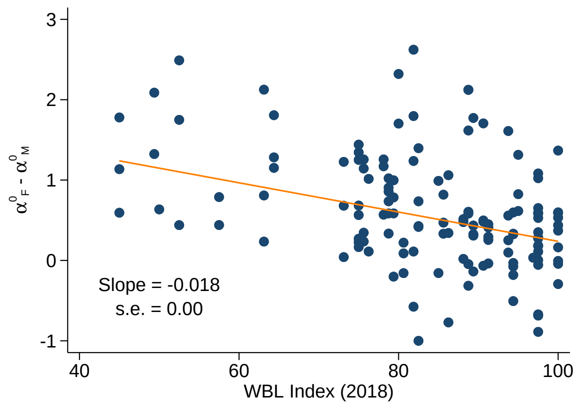
Notes. The figures are binned scatterplots and a linear fit of other performance measures (performance score, pay growth for new hires) against our calibrated productivity. The unit of observation is a worker-year. The performance score is an annual performance rating given by the line manager to each subordinate (continuous variable with values ranging between 0 and 150).

Figure 8: Calibrated parameters, norms and labor regulations

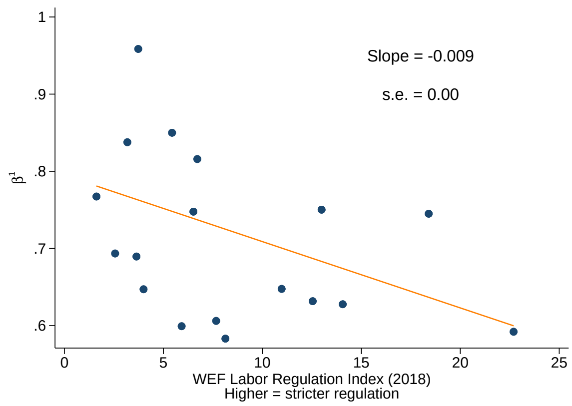
(a) World Value Survey



(b) Women, Business and the Law



(c) Restrictive Labor Regulations Index



Notes. Panel (a) shows scatterplots and fitted linear regressions of the gap in calibrated $\alpha^0_{Ftac} - \alpha^0_{Mtac}$ against four questions in the World Value Survey: (1) "Men make better business executives than women do," (2) "Pre-school child suffers with working mother," (3) "Being a housewife is just as fulfilling as working," (4) "When jobs are scarce, men should have more right to a job than women." For all questions, lower values of the index denote more agreement with the statement. Each dot is a country-cohort cell. Panel (b) shows the scatterplot and fitted linear regression of the gap in calibrated $\alpha^0_{Ftac} - \alpha^0_{Mtac}$ against the WB Women, Business, and the Law index in 2018. A higher value means fewer legal constraints on women's work. Each dot is a country cell. In all cases, the regression line uses cell size (at the firm) as analytic weights. Panel (c) shows the binned scatterplot and a linear fit of our calibrated parameters β^1_{gac} against the Restrictive Labor Regulations Index in 2018, using gender-country-cohort cell size (at the firm) as analytic weights.

9 Tables

Table 1: Summary statistics

	Full Sample		Structural Sample	
	Male (1)	Female (2)	Male (3)	Female (4)
Pay + Bonus (logs)	10.331	10.274	10.383	10.352
	(0.636)	(0.581)	(0.597)	(0.549)
Age	39.581	39.057	38.485	38.485
	(11.397)	(11.320)	(10.742)	(10.742)
Tenure	9.132	8.877	9.164	8.995
	(7.337)	(7.263)	(7.245)	(6.957)
Share in Work-level 2+	0.192	0.156	0.193	0.177
	(0.218)	(0.178)	(0.173)	(0.157)
Share with Fast Promotions	0.280	0.251	0.301	0.268
	(0.342)	(0.324)	(0.335)	(0.322)
Share Top Performers	0.175	0.144	0.173	0.151
	(0.123)	(0.096)	(0.072)	(0.063)
Econ, Business, and Admin	0.515	0.567	0.517	0.559
	(0.324)	(0.305)	(0.305)	(0.286)
Share in Sales Function	0.514	0.380	0.486	0.336
	(0.269)	(0.255)	(0.213)	(0.186)
Observations	377	366	260	260
Median no. workers in cell	115	95	234	191

Notes. This table reports summary statistics for the relevant sample of workers used in the analysis. An observation is a gender-cohort-country cell. This is the relevant unit in the structural estimation. Columns 1 and 2 present results for the full sample. In Columns 3 and 4, the sample is restricted to cells used in the structural estimation, where we exclude cells with fewer than 30 male or 30 female employees. Work level denotes the hierarchical tier (from level 1 at the bottom to level 6). The share of fast promotions only considers workers that achieve at least work-level 2 or higher. The sales function is the most common function (39%). Top performers are identified from the firm performance appraisal system based on the supervisor annual rating of worker performance.

Table 2: Gender pay gap and LFPR

	Full Sample					New Hires		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Pay + Bonus (logs)					Starting Salary	Pay Growth	Major promotion
Female	0.378	0.256	-0.007	0.253	0.197	-0.013	0.159	0.112
	(0.142)	(0.096)	(0.176)	(0.094)	(0.077)	(0.011)	(0.067)	(0.043)
LFPR	1.640	1.641	0.482	1.662	0.138		0.514	-0.025
	(0.282)	(0.212)	(0.160)	(0.204)	(0.235)		(0.151)	(0.073)
Female × LFPR	-0.564	-0.471	-0.312	-0.451	-0.376		-0.142	-0.118
	(0.194)	(0.135)	(0.113)	(0.132)	(0.103)		(0.087)	(0.058)
GDP per capita (logs)			0.291					
			(0.024)					
Female × GDP per capita (logs)			0.015					
			(0.017)					
Controls	No	Yes	Yes	Yes	Yes	Yes	No	No
Cohort FE	No	No	No	Yes	No	No	No	No
Country FE	No	No	No	No	Yes	No	No	No
N	303759	303759	302854	303759	303759	39241	8274	8274
R-squared	0.116	0.285	0.442	0.307	0.540	0.803	0.103	0.050

Notes. An observation is a worker-year. Controls include tenure, tenure squared, year FE and function FE. Column 6 restricts the sample to entry-level new hires and absorbs country-function-year fixed effects. Columns 7 and 8 restrict the sample to new hires at the entry level observed for at least four years. *Pay growth* is computed as the difference in log pay between the last year a worker is observed and the first year a worker is observed. *Probability of promotion* equals 1 if the worker was promoted to work-level 2 during the sample period. Columns 7 and 8 control for starting salary. Standard errors clustered at the country-cohort level.

Table 3: Summary of model parameters and empirical targets

Param.	Interpretation	Empirical Target
A. Structural Parameters:		
α^1_{gac}	Unconditional average log-wage in the firm	Average observed log-wage (controlling for selection)
β^1_{gac}	Returns to productivity in the firm	Variance of the observed log-wage (controlling for selection)
$\bar{\alpha}^1_{gac}$	Unconditional expected average log-wage in the labor market	Not separately identified. With an additional normalization, can be obtained from composite parameters ξ_{gac} and θ_{gac}.
$\bar{\beta}^1_{gac}$	Expected returns to productivity in the labor market	
α^0_{gac}	Unconditional average value of staying at home	
ν^0_{gac}	Dispersion of value of staying at home	
ρ_{gac}	Correlation between productivity and the value of staying at home	
B. Composite Parameters:		
ξ_{gac}	Participation threshold	LFP
θ_{gac}	Sign and strength of selection	Skewness of the observed log-wage (controlling for selection)

Notes. This table provides a summary of the parameters of the model and the empirical target that each parameter aims to match in the calibration strategy that we describe in subsection 4.2.

Table 4: Summary Statistics of Counterfactual Samples

	Male				Female			
	mean	sd	min	max	mean	sd	min	max
A. Full Structural Sample (260 country-cohort cells)								
α^1_{gac}	-0.04	0.58	-1.25	1.93	-0.31	0.79	-2.40	1.41
β^1_{gac}	0.70	0.18	0.21	1.57	0.78	0.25	0.25	1.90
ζ_{gac}	-0.91	0.17	-1.44	-0.27	-0.12	0.41	-1.14	1.25
θ_{gac}	0.29	0.65	-1.48	1.36	0.34	0.64	-1.81	1.41
B. Counterfactual Sample I (114 country-cohort cells)								
α^1_{gac}	-0.02	0.43	-1.25	1.15	-0.23	0.53	-2.40	0.89
β^1_{gac}	0.60	0.15	0.21	1.03	0.64	0.18	0.27	1.14
ζ_{gac}	-0.88	0.19	-1.30	-0.38	-0.20	0.40	-0.87	1.25
θ_{gac}	0.34	0.39	-0.72	1.00	0.33	0.43	-0.79	1.12
α^0_{gac}	-1.13	0.67	-2.51	0.62	-0.48	0.60	-1.90	1.08
ν^0_{gac}	1.00	0.00	1.00	1.00	1.00	0.00	1.00	1.00
ρ_{gac}	-0.26	0.68	-0.98	1.00	-0.20	0.62	-1.00	1.00
C. Counterfactual Sample II (88 country-cohort cells)								
α^1_{gac}	-0.05	0.38	-1.15	1.15	-0.34	0.47	-2.40	0.85
β^1_{gac}	0.61	0.14	0.21	1.03	0.65	0.17	0.27	1.14
ζ_{gac}	-0.89	0.20	-1.30	-0.38	-0.21	0.41	-0.87	1.25
θ_{gac}	0.40	0.35	-0.52	1.00	0.45	0.32	-0.47	1.12
α^0_{gac}	-1.23	0.58	-2.51	0.62	-0.61	0.56	-1.90	0.81
ν^0_{gac}	1.00	0.00	1.00	1.00	1.00	0.00	1.00	1.00
ρ_{gac}	-0.38	0.61	-0.98	1.00	-0.36	0.53	-1.00	0.99

Notes. This table summarizes the parameters by gender-country-cohort cells, using cell sizes as analytic weights. Panel A reports the parameters that are point identified without additional normalizations, as discussed in subsection 4.2, in the full structural sample (260 cells). Panel B additionally reports the parameters relating to the value of staying out of the labor force, under the normalization $\nu^0_{gac} \equiv 1$, excluding cells in which (a) we cannot calibrate α^0, ν^0 and ρ, or (b) we cannot compute one of the optimal wage policies for the main set of counterfactuals. Panel C reports the same parameters under the same normalization as Panel B, restricting the sample to the country-cohort cells for which we can compute the "optimal with equality" wage policy.

Table 5: Summary of Counterfactual Results

	Parameters		Results		
	α^1, β^1	α^0, β^0, ρ	Productivity	F/M Ratio	Pay Gap (F-M)
	A. Full Structural Sample (260 country-cohort cells)				
(1a)	Baseline	Baseline	0.25	0.67	−0.08
	B. Counterfactual Sample I (114 country-cohort cells)				
(1b)	Baseline	Baseline	0.31	0.71	−0.10
(2b)	Optimal	Baseline	0.47	0.92	0.63
			[0.37, 0.51]		
(3b)	Baseline	Men's	0.28	0.92	−0.12
			[0.11, 0.34]		
(4b)	Optimal	Men's	0.52	1.00	0.00
			[0.43, 0.55]		
(5b)	Strict Labor Reg.	Baseline	0.39	0.99	0.57
			[0.26, 0.42]		
	C. Counterfactual Sample II (88 country-cohort cells)				
(1c)	Baseline	Baseline	0.39	0.71	−0.10
(2c)	Optimal	Baseline	0.51	0.93	0.61
			[0.40, 0.56]		
(3c)	Baseline	Men's	0.35	0.91	−0.23
			[0.18, 0.41]		
(4c)	Optimal	Men's	0.55	1.00	0.00
			[0.45, 0.61]		
(5c)	Strict Labor Reg.	Baseline	0.46	0.96	0.54
			[0.32, 0.51]		
(6c)	Opt. with Equality	Baseline	0.46	0.78	0.14
			[0.32, 0.51]		
(7c)	Opt. with Equality	Men's	0.55	1.00	0.00
			[0.45, 0.61]		

Notes. This table summarizes the counterfactual results (average productivity, female-to-male employee ratio, and pay gap) across country-cohort cells, weighted by cell size. Productivity ($LFP \cdot \mathbb{E}[A_i \mid \text{empl'd}]$) is measured in units of standard deviations of the underlying latent variable A_i. The pay gap is the difference in log-wage (female − male). In the "optimal" wage policy, the firm maximizes ability subject to an employment and a wage bill constraint, see section C. In the "strict labor regulations" counterfactual, the firm solves the same problem as in the "optimal" wage policy, with an additional cap on returns to ability. In the "optimal with equality" wage policy, the firm maximizes ability subject to an employment, a wage bill constraint, and an equality between men and women restriction. Panel A reports the baseline results in the full structural sample (260 cells). Panel B reports a set of counterfactual results for 114 cells, which exclude those in which (a) we cannot calibrate α^0, ν^0, and ρ, or (b) we cannot compute one of the optimal wage policies for the counterfactuals reported. Panel C reports the same results for 88 cells, in which we can additionally compute the "optimal with equality" wage policy. The remaining cells are excluded primarily because the feasible region is empty (i.e., the firm cannot choose a wage policy that satisfies the LFP, the wage bill and the equality constraint at the same time). Square brackets report 95% bootstrap confidence intervals based on 300 bootstrap replications.