Cultural Constraints on Women's Productivity
Evidence from World Bank Lending

Yves Froude †

Aaditya Mattoo†

Bob Rijkers§ †

Michele Valsecchi †

August 2025

Abstract: Do unequal gender norms limit women's productivity? We study this question using three decades of World Bank project and HR data. Women are less likely than men to lead aid projects in countries with high gender inequality, despite no evidence of lower ability. Using a difference-in-differences design, we show that female-led projects suffer a performance penalty relative to men's in such settings, consistent with cultural constraints. These effects remain after controlling for project leader and country fixed effects and instrumenting female leadership with the gender composition of the pool of potential leaders.

Key words: gender, aid effectiveness, development

JEL codes: O10, O16, O22, J16, D73

§ Utrecht University, † Development Research Group, World Bank. We thank Alina Timoshkina, Cung Truong Huoang and Natalia Pecorari for excellent research assistance, Daria Tagloni, Jed Friedman, and Clement Joubert for excellent comments and help accessing the data. We also benefited from discussions with the leadership team from the People and Culture Vice Presidency of the World Bank, and from comments from and discussions with Radi Anguelova, S. Anukriti, Reena Badiani, Kathleen Beegle, Emmet Cunningham, Louise-Paul Delvaux, John Giles, Eeshani Kandpal, Tatjana Kleineberg, and Joanne Marie Mc Donough. This paper has been partly supported by the Umbrella Facility for Trade trust fund (financed by the governments of the Netherlands, Norway, Sweden, Switzerland and the United Kingdom) and the Strategic Research Partnership on Economic Development. We also acknowledge the generous financial support from the World Bank research support budget. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank of Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the countries they represent. All errors are our responsibility. The authors can be contacted at yfroude@worldbank.org, mvalsecchi@Worldbank.org, amattoo@worldbank.org and brijkers@worldbank.org.

1 Introduction

Across the world, women often achieve lower measured outcomes than men, even when they have similar skills and qualifications. Evidence from agriculture, entrepreneurship, and the labor market documents persistent productivity gaps, linked not only to unequal access to inputs, networks, and finance, but also to bias and cultural norms that constrain women's effectiveness (Udry (1996); Goldstein and Udry (2008); De Mel et al. (2009); Jayachandran (2015); Bertrand (2020)).

Identifying how gender norms shape such gaps is challenging: It is only rarely possible to observe the same individual performing the same task(s) across multiple cultural contexts. Men and women are often (self-)selected into different roles, and may differ systematically in observable and unobservable characteristics, complicating causal inference. Performance can also be difficult to measure, and observed performance differences may reflect variation in incentives or objectives rather than productivity.

We leverage a unique setting to examine how cultural constraints shape women's productivity, focusing on the performance of World Bank Task Team Leaders (TTLs) responsible for delivering aid projects. TTLs are directly accountable for project outcomes and interact intensively with recipient-country counterparts. They operate under a standardized organizational framework, face common incentives, and perform comparable tasks, yet are deployed across countries that vary widely in gender inequality. Using three decades of project and HR data, we track the same TTLs across countries to assess whether local gender norms systematically affect their effectiveness. This context provides an unusually clean test of cultural constraints: differences in outcomes can be attributed to the interaction between gender and local norms rather than variation in individual ability, tasks, roles, or incentives.

Using this novel dataset we examine whether gender inequality constrains the effectiveness of foreign aid. Specifically, we ask whether projects led by women perform systematically worse in countries with higher gender inequality, consistent with cultural constraints imposed by gender norms. To estimate causal effects, we exploit the rotation of TTLs across recipient countries in a difference-in-differences framework, controlling for individual and country fixed effects. To further address potential selection into being a project leader, we construct an instrumental variable based on the gender composition of the pool of potential TTLs, notably the share of women at the relevant professional grade within the unit hosting the project. This design isolates the effect of local gender norms from assignment patterns or ability differences, providing one of the first systematic, cross-country, micro-level tests of how cultural constraints shape women's productivity in high-skilled professional roles and the effectiveness of the projects they lead.

Turning to our main findings, we begin by documenting strong gender sorting. Women are significantly less likely to lead projects in gender-unequal countries, but we find no evidence of sorting by ability: women (and men) working in such countries are not, on average, less experienced, educated, or capable than other Task Team Leaders (TTLs). Women are over-represented in lower grade levels and more likely to lead projects in education, social protection, and health, and less likely in agriculture, energy, or extractives.

Second, and most important, using a difference-in-difference design, we show that development projects led by female TTLs suffer a performance penalty in gender-unequal

countries relative to those led by men. This result is robust to controlling for TTL and country fixed effects, further assuaging concerns that the result reflects sorting based on ability of the TTL and/or unobserved country characteristics. Finally, we show that these results are robust to a variety of tests, including addressing potential residual selection bias with an instrumental variable strategy, using the gender composition of staff at the same grade level in the hosting unit as an instrument for female project leadership. IV estimates are, if anything, more negative than OLS estimates, suggesting that women selected to lead projects in gender-unequal countries are more adept than they would be under a gender-blind assignment.

This paper builds on and contributes to several strands of literature. First, it complements prior work on gender and productivity by providing rare cross-country micro-level evidence on the role of cultural norms. While prior research shows that gender norms are associated with barriers to human capital accumulation, (Klasen (2018)), female entrepreneurship (Chiplunkar and Goldberg (2024), and talent allocation Hsieh et al. (2019)), evidence on how social norms constrain women's effectiveness in professional roles remains limited. By exploiting TTL rotations across countries with varying gender inequality, we isolate the effect of local norms on women's performance, complementing the work of Ashraf et al. (2022), which shows that women are positively selected in contexts where gender norms strongly favor men.

Prior research shows that gender inequality hampers growth, inter alia by erecting barriers to human capital accumulation (Klasen, 2018) and entrepreneurship (Chiplunkar and Goldberg, 2024), and inducing misallocation of talent (Hsieh et al., 2019). Yet, direct cross-country evidence on the role of norms remains scarce.[1]A central question in this literature is whether observed gender gaps in pay and employment reflect differences in women's characteristics or constraints imposed by prevailing gender norms. While the evidence increasingly points to the latter (Olivetti et al., 2024), direct cross-country evidence remains scarce. By leveraging the rotation of Task Team Leaders across countries with different gender norms, we isolate the impact of social norms on women's relative performance from individual characteristics. Alongside Ashraf et al. (2022) we provide some of the first systematic cross-country evidence on how cultural context shapes women's labor market participation and productivity.

Second and related, we contribute to the literature on aid effectiveness in general (for reviews see Qian (2015); Arndt et al. (2016)), and the success of individual projects in particular. Previous studies have shown that project manager characteristics, project preparation, and recipient country factors partly predict project outcomes (Kilby (2015), Denizer et al. (2013), Ashton et al. (2023), Geli et al. (2014)), with project manager characteristics being especially important. Building on this, we provide the first systematic evidence quantifying how project leader gender interacts with local gender norms to influence project outcomes, showing that local culture shapes project performance. Our findings indicate that, alongside institutional quality (Burnside and Dollar (2000, 2004)), gender-based cultural fit is a measurable and economically meaningful determinant of project success.

The remainder of this paper is organized as follows. The next section desribes the

[1]A notable exception is Ashraf et al. (2022), who document positive selection of women in countries with low female labor force participation

institutional context and the data. Section 3 describes our identification strategy. Results are presented in section 4. A final section concludes.

2 Institutional Context

The World Bank is the world's largest multilateral development lender, financing operations in almost all developing countries. Each project is managed by one or more Task Team Leaders (TTLs), who are responsible for project design, supervision, and outcomes, and who serve as the primary interlocutors with recipient governments. To isolate individual contributions, our main analysis excludes projects jointly managed by multiple supervisor TTLs.

All TTLs operate within a common institutional framework: projects must follow standardized preparation, supervision, and evaluation procedures, and outcomes are assessed using harmonized metrics, including ratings from the Bank's Independent Evaluation Group. This consistency allows project performance to be compared across countries while holding institutional rules constant.

TTL assignments are made by managers based on operational needs, technical expertise, and staff availability. Junior staff (grades GE–GF) typically manage smaller or less risky projects, while senior specialists and leads (GG–GH) take on larger, technically demanding, or politically sensitive operations. Individual staff preferences and career progression are also taken into consideration. Serving as TTL is widely regarded as a desirable role, providing visibility and the opportunity to shape development outcomes.

Women have historically been underrepresented in higher grades and certain technical sectors (e.g., energy, infrastructure). This creates systematic variation in the pool of potential female TTLs across units and grades, which we exploit in our identification strategy.

The World Bank offers a suitable setting to study the impact of gender inequality on development effectiveness. Because all projects follow common rules, procedures, and evaluation standards, project performance can be compared consistently across countries. Since many TTLs work in multiple countries, we can observe the same individual operating under different gender norms while holding institutional processes constant. Moreover, it is committed to gender equity, reducing concerns that the institution itself may be reinforcing gender unequal outcomes. Last but not least, it has excellent administrative data which it has been willing to share for research purposes.

3 Data

The project combines various datasets:

World Bank Project Data We construct a dataset of World Bank projects implemented between 1987 and 2023, combining multiple sources of information in three steps. First, we collect standard project-level data from the World Bank project portal, including approval date, expected and actual duration, sector, recipient country, and total amount committed (in USD).

Second, and most novel, we manually compile information on Task Team Leaders (TTLs). For each project, we identify the TTL at approval ("Designer TTL") and at completion ("Supervisor TTL"), drawing primarily from Implementation Completion Reports (ICRs), the World Bank's mandatory end-of-project self-evaluation documents. When ICRs are unavailable or incomplete, we supplement the data with intermediate project documents and Project Appraisal Documents (PADs), which are prepared by Bank staff just before consideration by the World Bank's Executive Board.

Third, we merge these data with outcome ratings from the World Bank's Independent Evaluation Group (IEG), which evaluates completed projects on a standardized six-point scale ranging from Highly Unsatisfactory (1) to Highly Satisfactory (6).These ratings offer a comparable measure of project performance across countries and over time. They have been shown to correlate with objective outcomes. For example, Andersen et al. (2022) show that World Bank aid to highly aid-dependent countries triggers outflows to tax havens, but only for projects that fail, i.e., those with an unsatisfactory outcome rating.

In our analysis, we restrict the sample to IDA- and IBRD-financed investment projects with published IEG outcome ratings, known TTLs, and executed in a single country.[2] Our main sample therefore consists of 4,367 projects with a single supervisor TTL. In robustness checks, we also report results for projects with multiple TTLs.

Human Resources Data We also access administrative human resources (HR) data of World Bank staff records for the period 1987-2024. These data include information on staff gender, nationality, age, education, family composition, date of entry, position, grade, salary, contract type, duty station, and organizational unit. In addition, we observe annual performance ratings for most staff, scored on a scale from 1 (totally unsatisfactory) to 5 (superb), with the vast majority receiving a 3. Higher ratings are associated with faster salary progression.

Gender equity indicators We use four commonly used cross-country indicators of gender inequality and uneven gender norms. First, the Global Gender Gap Index (GGGI), published by the World Economic Forum (WEF), measures gender gaps across economic participation, educational attainment, health, and political empowerment (Forum, 2021). Second, the Social Institutions and Gender Index (SIGI), compiled by the OECD Development Centre, measures discrimination embedded in social institutions across four domains: discrimination in the family, restricted physical integrity, restricted access to productive and financial resources, and restricted civil liberties (Centre, 2023). Third, the Gender Inequality Index (GII), published by the United Nations Development Programme (UNDP), reflects gender-based disadvantages in reproductive health, empowerment, and labor market participation (Programme, 2020). Finally the Women, Business and the Law (WBL) Index, produced by the World Bank, captures legal barriers to women's economic opportunities across ten domains covering the lifecycle of an economically active woman: safety, mobility, workplace,

[2]We exclude development policy loans, projects financed by trust funds, multisectoral projects, and projects spanning multiple countries. Our starting dataset contains 7,955 projects with known IEG outcome ratings, of which 220 cannot be assigned to a unique country and 15 lack information on the main sector. We further exclude 757 development policy loans, 102 projects for which we cannot identify the relevant organizational unit of the TTL in the HR data, and 1,129 projects with multiple supervisor TTLs.

pay, marriage, parenthood, childcare, entrepreneurship, assets, and pensions (Bank, 2023). This is a measure of *de jure* discrimination that may not capture reality on the ground. All four indices are standardized to have a mean of zero and a standard deviation of one, and rescaled such that higher values indicate greater gender inequality.

Other data We supplement these data with information on Gross Domestic Product (GDP) per capita from the World Development Indicators (WDI).

4 Empirical Strategy

Our study tests whether gender inequality in recipient countries impacts the effectiveness of World Bank projects differently depending on whether they are led by a female or male Task Team Leader (TTL). Specifically, we examine the hypothesis that projects led by female task team leaders suffer a performance penalty relative to those led by males in environments with unequal gender norms favoring men.

To test this hypothesis, we use a difference-in-difference design comparing how differences in project performance between female- and male-led projects vary with the level of gender inequality in the recipient country. Specifically, we regress indicators of project performance, notably IEG outcome scores and project delays, on a dummy for whether the project manager is female, a measure of gender inequality in the recipient country, and, crucially, their interaction. The coefficient on this interaction β is the parameter of focal interest, because it measures the female performance premium (or penalty) associated with operating in a more gender-inequitable environment. Because gender norms are highly persistent and many gender inequality measures are only recently available, we use time-averaged values of these proxies, treating gender inequality as time-invariant in our regressions.

The primary empirical concern is that any observed performance penalty (or premium) reflects confounders rather than gender inequality itself. Differences in TTL ability, project type, or country characteristics could bias estimates. Of particular concern is differential selection by gender across countries: if the characteristics of male and female TTLs vary systematically with gender inequality, observed performance gaps may reflect selection rather than genuine differences. To explore this, we will examine how proxies for TTL ability correlate with gender, gender inequality, and their interaction, controlling for TTL, project, and country characteristics.

In our main analysis, we address potential selection bias by including extensive controls and fixed effects. Our most stringent specifications include both country and TTL fixed effects. Country fixed effects absorb differences in time-invariant country characteristics, such as geography or culture (to the extent it is time-invariant). TTL fixed effects remove bias from time-invariant characteristics of TTLs, such as innate ability, and ensure identification comes from comparing how the same TTL performs across countries with different levels of gender inequality.

Our preferred specification is:

$$Y_{ict} = \beta Female_i \times Gender\ Inequality_c + \mu_i + c_c + \phi\, TTL_{it} + \theta\, P_{pt} + \psi\, C_{ct} + \varepsilon_{ict}$$

Where Y_{ict} denotes the outcome for project p in country c initiated in year t supervised by TTL i; $Female_i$ is an indicator for a female TTL; $Gender\ Inequality_c$ is a time-averaged measure of gender inequality in the recipient country; μ_i are TTL fixed effects and C_c are country fixed effects[3] TTL_{it} is a vector of time-varying TTL-level characteristics notably grade salary (which serves as a proxy of ability), age, age squared, dummies capturing prior project supervision experience (1, 2, or more than 2 projects); P_{pt} is a vector of project-level characteristics notably log commitments, log expected project duration, sector dummies, and financing source (IDA, IBRD or both); and C_{ct} is a vector of time-varying country characteristics, notably whether a country is at war, experiencing a financial crisis and the risk of natural hazafds. In our most stringent specifications, these country characteristics are interacted with the female TTL dummy to allow for the possibility that the effect of country characteristics other than gender inequality varies by the gender of the project leader.

While this is a highly stringent specification, potential residual selection concerns remain. Female TTLs operating in gender-unequal countries may be positively selected (e.g., only those particularly apt at operating in gender unequal contexts take such assignments) or negatively selected (e.g., those with fewer outside options operate in gender unequal countries). TTL fixed effects remove bias from cross-TTL selection, but may not fully account for TTL–recipient country specific complementarities.

To address these concerns, we implement an instrumental variables (IV) approach. Specifically, we instrument whether a project is assigned a female TTL using the gender composition of eligible TTLs by grade and by the unit hosting the project at the time of project approval. Intuitively, the instrument captures the likelihood that a project is led by a female TTL based on the pool of eligible female staff in that unit and grade, rather than project-specific characteristics. Because women have historically been less represented at senior grades (e.g., H-level lead specialists) than at junior professional grades (e.g., F-level professionals), and because certain sectors, such as human development or education, employ more women than others, such as transport or infrastructure, the probability that a project is led by a female TTL varies both by grade and unit hosting the project.

5 Who leads which projects?

To set the scene for the analysis that follows, we first examine the characteristics of female project managers and how they differ from men, as well as those of the project they manage. We subsequently assess whether women are less likely to lead projects in countries with unequal gender norms, and whether the ones that do have special characteristics that could impact project outcomes. This not only helps contextualize the results, but also helps

[3]Their inclusion precludes separate identification of TTL the direct effect of gender inequality on project performance.

us assess potential selection bias, which is one of the primary obstacles to identifying the impact of gender norms on gender differences in project performance.

5.1 Gender Differences in Task Team Leaders and Their Projects

Over the full sample period, only about one-third of World Bank projects have been led by women, despite women comprising almost half of the institution's professional staff. Women are thus less likely to be project managers.

Table 1 compares the characteristics of male and female task team leaders (TTLs) and the projects they manage. Panel A begins with demographic characteristics of the TTLs. On average, female TTLs are younger and have less experience, whether measured by years at the World Bank, professional grade, or number of projects managed. Figure 2 shows that women are more prevalent among lower-grade TTLs: 69% of entry-level analyst who are TTLs (GE level) are female, compared with 14% of lead specialists who serve as TTLs (GH level). This partly reflects the historic underrepresentation of women at higher grades, although gender representation has dramatically improved in more recent cohorts. Female TTLs also have somewhat lower educational attainment (despite most holding graduate degrees), earn lower salaries on average, and are less likely to be married.

Panel B documents differences in project characteristics. Importantly, we find no evidence of unconditional performance differences between male and female TTLs. Average project outcome ratings do not differ significantl across genders, and neither the expected nor the actual project duration varies by TTL gender. Projects led by female TTLs are somewhat smaller on average in terms of total amount of funding committed.

Gender sorting by sector is notable Women are more likely to lead in projects in social protection (48% of all TTL's are female) education (47%), and health (46%), and less likely to lead projects in agriculture (19%), transportation (16%) and energy and extractives (14%) (Appendix Figure 1). This sectoral sorting partly reflects the gender composition of staff working in those areas.

There is also strong sorting across countries, as is shown in panel C. Female TTLs tend to manage projects in countries with significantly lower gender inequality. These countries also have higher GDP per capita, stronger democratic institutions, lower incidence of conflict, and fewer financial crises. Across all four measures of recipient-country gender inequality, projects led by women are consistently in countries with lower inequality, though the magnitudes vary: ranging from 0.157 standard deviations lower according to the World Economic Forum (WEF) measure, 0.106 standard deviations lower according to the OECD's Social Institutions and Gender Index (SIGI), 0.08 lower according to the Women, Business and the Law (WBL) index, to 0.027 standard deviations lower according to the Gender Inequality Index (GII) published by the United Nations Development Programme (UNDP).

The likelihood that a project is led by a female TTL is strongly associated with the share of women among potential TTLs, proxied by the share of staff at the same grade level as the TTL in the unit hosting the project, as is shown in Panel D and Appendix Figure 3 which presents a binscatter of project leadership against the share of women at the relevant grade. We leverage this variation in the supply of potential TTLs in our instrumental variables strategy.

Table 1: Descriptive Statistics by Gender of the TTL

	Men	Women	Difference
A. TTL Characteristics			
Age	42.582	39.148	-3.435***
Ph.D	0.323	0.277	-0.046*
Is married	0.830	0.538	-0.292***
Salary	46412.567	39464.378	-6948.189***
1 prior project	0.145	0.146	0.001
2 prior projects	0.067	0.055	-0.013
GE - analyst	0.018	0.098	0.080***
GF - specialist	0.240	0.392	0.152***
GG - senior specialist	0.614	0.398	-0.217***
GH - expert	0.071	0.028	-0.043***
B. Project Characteristics			
IEG Outcome Rating	4.057	4.013	-0.044
Project Delay (days)	593.701	594.210	0.509
Project Delay >0	0.769	0.762	-0.007
Log Commitments (USD)	17.538	17.323	-0.215***
IDA	0.433	0.427	-0.006
IBRD	0.533	0.540	0.007
IDA & IBRD	0.034	0.033	-0.001
Agriculture	0.141	0.079	-0.061***
Education	0.077	0.169	0.092***
Energy & Extractives	0.125	0.048	-0.077***
Financial Sector	0.037	0.059	0.022**
Health	0.064	0.132	0.068***
Industry Trade Services	0.053	0.067	0.014
Info & Communication	0.016	0.018	0.002
Public Administration	0.163	0.164	0.001
Social Protection	0.041	0.094	0.052***
Transportation	0.162	0.078	-0.084***
Water Sanitation Waste	0.121	0.092	-0.029**
C. Recipient Country Characteristics			
Gender Inequality (WEF)	0.240	0.083	-0.157***
Gender Inequality (SIGI)	0.249	0.144	-0.106**
Gender Inequality (UNDP)	0.419	0.392	-0.027***
Gender Inequality (WBL)	0.099	0.019	-0.080**
GDP per capita (2015 US$)	2804.499	3215.942	411.443***
Polity	1.772	2.566	0.794***
War	0.016	0.008	-0.008**
Crisis	0.499	0.485	-0.014
Natural Hazards	5.044	4.912	-0.132
D. Unit Characteristics			
Female share unit-grade	0.237	0.550	0.314***

Notes: This table presents descriptive statistics on the characteristics of female and male Task Team Leaders. Column 1 presents averages for male TTLs, column 2 averages for female TTLs, and column 3 the difference. *, **, *** indicate significance at the 10%, 5%, and 1% significance level, respectively.

5.2 Which projects are led by women?

To assess which projects are led by women, Table 2 presents probit regressions in which the dependent variable equals one if the project manager is female. We add progressively more elaborate sets of explanatory variables. We are especially interested in examining whether women are less likely to lead projects in countries with unequal gender norms.

Column 1 only includes gender inequality using the gender gap index from the WEF as an explanatory variable. We focus primarily on this measure in the main text but present robustness to using alternative measures in the appendix. Gender inequality is negatively correlated with the probability of having a female TTL. A one standard deviation reduces the likelihood of having a female by 4.5% (the mean is 28.8%), but the predictive power of gender inequality appears limited given that the pseudo R2 is 0.00582.

The association between recipient country gender inequality and having a female TTL is robust to controlling for characteristics of the Task Team Leader, as is done in column 2, which furthermore shows that female TTLs are more prevalent at more junior professional grades. The association attenuates but remains statistically significant when controlling for project characteristics, as is done in column 3. The likelihood of having a female TTL is lower when projects are larger and varies considerably across sectors. It is also robust to controlling for country characteristics as is done in column 4.

Finally, in column 5 we add controls for the share of women in the pool of potential TTLs by controlling for the share of women at the grade level of the TTL in the unit hosting the project, both itself and interacted with gender inequality. These variables, which will serve as our instrument, have very strong predictive power.

In summary, the negative association between gender inequality and the probability of a female TTL is robust across specifications. Appendix Table 6 replicates these regressions using alternative gender inequality proxies. The association between gender inequality and having a female TTL is typically negative but not always statistically significant.

5.3 (How) Are Female Task Team Leaders in Gender-Unequal Contexts Different?

The previous subsection showed evidence of sorting by gender. Such sorting may create selection bias: female Task Team Leaders (TTLs) who take on projects in gender-unequal countries may systematically differ from those working in more gender-equal settings in ways that also affect project performance. For our purposes, the key concern is *differential* selection bias by gender: if gender differences in TTL characteristics vary with gender inequality, then observed gender gaps in project performance could reflect selection rather than genuine performance differences.

To assess whether such selection bias might be at play, and whether it differs by gender, Table 3 examines how various proxies for TTL ability correlate with gender, gender inequality, and their interaction. Panel A reports unconditional correlations; Panel B adds controls for TTL and project characteristics; and Panel C further includes country fixed effects and time-varying country covariates. Evidence of systematic correlations would indicate that our main estimates may be biased by differential sorting, whereas the absence of such patterns would mitigate concerns about confounding due to selection.

Table 2: Which projects are led by women? Probit models

	(1)	(2)	(3)	(4)	(5)
Gender Inequality (WEF)	-0.045***	-0.041***	-0.031***	-0.026**	-0.042**
	(0.010)	(0.012)	(0.010)	(0.012)	(0.017)
Age		-0.003**	-0.004***	-0.004***	-0.001
		(0.001)	(0.001)	(0.001)	(0.001)
Ph.D		-0.011	-0.044**	-0.042*	-0.046**
		(0.021)	(0.022)	(0.022)	(0.020)
GE - analyst		0.338***	0.309***	0.323***	0.004
		(0.036)	(0.035)	(0.036)	(0.032)
GF - specialist		0.143***	0.124***	0.124***	-0.030*
		(0.027)	(0.024)	(0.025)	(0.017)
GH - expert		-0.050	-0.055	-0.047	0.012
		(0.050)	(0.044)	(0.045)	(0.035)
Log Commitments			-0.017**	-0.015*	-0.012
			(0.008)	(0.008)	(0.008)
Education			0.236***	0.227***	0.056**
			(0.032)	(0.032)	(0.028)
Energy and Extractives			-0.057	-0.057	-0.054
			(0.036)	(0.037)	(0.035)
Financial Sector			0.142***	0.137***	0.075*
			(0.047)	(0.048)	(0.040)
Health			0.231***	0.217***	0.050
			(0.037)	(0.039)	(0.034)
Industry/Trade			0.162***	0.158***	0.090**
			(0.043)	(0.045)	(0.037)
Info/Comm			0.092	0.087	0.076
			(0.062)	(0.063)	(0.056)
Public Admin			0.064**	0.072**	0.038
			(0.031)	(0.032)	(0.027)
Social Protection			0.237***	0.240***	0.100***
			(0.044)	(0.047)	(0.037)
Transportation			-0.040	-0.042	-0.025
			(0.037)	(0.038)	(0.031)
Water/Sanitation			0.000	-0.009	-0.023
			(0.035)	(0.035)	(0.031)
GDP per capita (log)				0.017	0.004
				(0.014)	(0.013)
Polity				-0.001	-0.000
				(0.002)	(0.001)
War				-0.037	0.026
				(0.037)	(0.033)
Crisis				0.004	0.004
				(0.019)	(0.016)
Natural Hazards				0.000	0.002
				(0.006)	(0.005)
Female share unit-grade					0.889***
					(0.025)
Female share*G. Inequality (WEF)					0.074**
					(0.036)
Approval year dummies	No	No	Yes	Yes	Yes
Financing source dummies	No	No	Yes	Yes	Yes
N	3988	3409	3406	3219	3219
Observed p	0.288	0.305	0.305	0.307	0.307
Pseudo R2	0.00582	0.0591	0.132	0.128	0.292

Notes: This table presents estimates of the likelihood that the Task Team Leader is a woman. Marginal effects are presented with standard errors clustered by recipient country in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% significance level, respectively.

Table 3: Testing for Potential Ability Bias

	(1) Age	(2) PhD.	(3) # projects	(4) salary	(5) SRI (prior)	(6) IEG rating (prior)
Panel A: No Controls						
TTL female	-3.077***	-0.0394	-0.0262	-6310.8***	-0.0338	-0.159
	(0.389)	(0.0305)	(0.0519)	(1050.6)	(0.0361)	(0.126)
Gender Inequality (WEF)	0.219	0.0220	0.00910	-946.2	-0.0149	-0.108
	(0.193)	(0.0170)	(0.0191)	(768.4)	(0.0190)	(0.0868)
TTL female*GI (WEF)	-0.597	0.0126	-0.0636	798.4	0.0292	-0.150
	(0.437)	(0.0221)	(0.0474)	(939.3)	(0.0323)	(0.144)
Mean Dep Var	41.58	0.309	0.321	44368.9	3.380	4.107
N	3987	3408	3987	3930	976	661
R2	0.0795	0.0209	0.0609	0.115	0.0136	0.0534
Panel B: With TTL and Project Controls						
TTL female	-0.930**	-0.0550	0.0869	-414.7	-0.0256	-0.0884
	(0.348)	(0.0308)	(0.0477)	(804.3)	(0.0358)	(0.140)
Gender Inequality (WEF)	0.167	0.0233	0.00977	-174.8	-0.0183	-0.0283
	(0.166)	(0.0160)	(0.0178)	(714.4)	(0.0202)	(0.0790)
TTL female*GI (WEF)	-0.365	0.0112	-0.0509	1497.0*	0.0416	-0.177
	(0.297)	(0.0195)	(0.0401)	(598.4)	(0.0316)	(0.146)
TTL controls	No	No	No	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
Project Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean Dep. Var	41.58	0.309	0.321	44368.9	3.380	4.107
N	3983	3406	3983	3359	938	628
R2	0.338	0.0808	0.152	0.451	0.0920	0.103
Panel C: with TTL, Project, and Country Controls						
TTL female	-0.949**	-0.0501	0.0914	-67.74	-0.0364	-0.113
	(0.340)	(0.0304)	(0.0467)	(766.7)	(0.0430)	(0.130)
TTL female*GI (WEF)	-0.369	0.00952	-0.0552	1086.4	0.0190	-0.0513
	(0.332)	(0.0230)	(0.0384)	(628.8)	(0.0361)	(0.191)
TTL controls	No	No	No	Yes	Yes	Yes
Grade FE	Yes	Yes	Yes	Yes	Yes	Yes
Project Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Country Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean Dep Var.	41.58	0.309	0.321	44368.9	3.380	4.107
N	3764	3213	3764	3168	882	588
r2	0.382	0.122	0.193	0.516	0.214	0.298

Notes: Standard errors clustered are clustered two-ways, by recipient country and Task Team Leader, and presented in parentheses. TTL controls include age, education, and dummies for having been a TTL of one or 2 or more projects in the past, project controls include log commitments, financing type dummies and sector dummies. Country controls are log GDP per capita, the polity score, a dummy for being at war, a dummy for experiencing a financial crisis and a measure of the risk of natural hazards. *, **, *** indicate significance at the 10%, 5%, and 1% significance level, respectively.

Starting with unconditional correlations in panel A, female TTLs are on average younger than male ones (column 1) as already discussed above, and also earn significantly less (column 2), yet we do not find evidence that (male) TTLs who are assigned to gender unequal countries differ from those sent to more equal ones in terms of their age (column 1), education (column 2), project experience (column 3), salary (column 4), prior average performance (Salary Review Increase - SRI) ratings (column 5) or in terms of the outcome ratings of projects they already completed (column 6). More importantly perhaps there is no evidence of differential sorting by gender: none of the interactions between gender and gender inequality are statistically significant. Project leaders in more gender-unequal environments therefore seem neither positively nor negatively selected.

These results are robust to controlling for TTL and project characteristics as is done in Panel B. The female salary penalty is not longer statistically significant, if anything, it seems that women that are TTLs in gender unequal countries, have slightly higher salaries (though this effect is only significant at the 10% level).

Finally the results are also robust to including country fixed effects and time-varying country characteristics, as is done in panel C. Note that in this case the direct effect of gender inequality can no longer be separately identified (for it is collinear with the country fixed effects). Most importantly there is no evidence of differential sorting based on ability by gender, and, more generally, at best very limited evidence for potential ability bias. Overall, the results suggest there is no systematic sorting into gender-unequal countries on the basis of ability. These findings are robust to adding country fixed effects (Panel C), which compare male and female TTLs within the same country.

6 Does Gender Inequality Constrain Development Effectiveness?

Our main findings are presented in Table 4, which examines the determinants of IEG project outcome ratings using only projects for which we observe a single supervisor TTL. We first present OLS estimates and then discuss instrumenal variables estimates. The first specification includes only approval-year dummies and a dummy for whether the task team leader (TTL) is a woman. The estimated female performance premium is small (0.057) and statistically insignificant (see also Table 1), indicating that unconditionally, projects led by women do not perform significantly better or worse than those led by men.

Once we add TTL characteristics, project controls, and country fixed effects (column 2), the female performance premium becomes more negative and statistically significant. All else equal, projects led by women receive outcome ratings that are 0.091 points lower than those led by men. This difference is modest relative to the dependent variable's standard deviation of 1.114 but not negligible. The contrast between the unconditional and conditional estimates suggests that women are more often assigned to projects that, on average, are easier to implement successfully, which may partly reflect concentration in more favorable country contexts and can mask underlying differences when no controls are

included.[4]

Table 4: Impact of Gender and Gender Norms on Project Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Dependent variable: Project Outcome Rating								
	OLS							IV	
TTL female	-0.057	-0.091**	-0.074*	-0.068*	-0.084**				
	(0.041)	(0.040)	(0.040)	(0.040)	(0.040)				
Gender Inequality (Inv. GGI, STD)			-0.097**	-0.115***					
			(0.037)	(0.041)					
TTL female × Gender Inequality			-0.074*	-0.070	-0.076*	-0.254**	-0.251**	-0.318*	-0.333*
			(0.043)	(0.045)	(0.044)	(0.105)	(0.103)	(0.189)	(0.187)
Approval year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
TTL controls	-	Yes	-	Yes	Yes	Yes	Yes	Yes	Yes
Project controls	-	Yes	-	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	-	Yes	-	-	Yes	Yes	Yes	Yes	Yes
TTL FE	-	-	-	-	-	Yes	Yes	Yes	Yes
Country controls × female TTL	-	-	-	-	-	-	Yes	-	Yes
Observations	4,286	4,133	3,998	3,872	3,871	2,813	2,813	2,813	2,813
R-squared	0.020	0.157	0.028	0.068	0.146	0.532	0.535	0.012	0.017
Mean Dep. Var.	4.044	4.051	4.073	4.077	4.077	4.110	4.110	4.110	4.110
SD Dep. Var.	1.149	1.145	1.141	1.138	1.138	1.123	1.123	1.123	1.123
KP stat	–	–	–	–	–	–	–	19.15	19.00
N. TTL	2,159	2,093	2,073	2,016	2,015	947	947	947	947
N. rec. countries	147	141	117	116	115	108	108	108	108

Notes: Standard errors are two-way clustered by recipient country and Task Team Leader and reported in parentheses. TTL controls include salary, age, education, professional grade dummies, and prior project experience dummies. Project controls include log commitments, financing type, and sector dummies. Country FE includes GDP per capita, natural hazards risk, war, and financial crisis dummies. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

The third column investigates a potential source of gender differences in performance, namely local cultural norms. It replicates column 1 and adds a measure of gender inequality from the World Economic Forum, as well as its interaction with the female TTL dummy. Both gender inequality itself and its interaction with being a female TTL are strongly statistically significant. A one standard deviation increase in gender inequality is associated with a 0.097-point lower outcome rating for projects led by men. For female TTLs, the reduction is even larger, with an additional 0.074-point decline on average. Thus, female-led projects perform worse than male-led projects in countries where gender norms favor men.

To assess the robustness of this key finding, column 4 adds controls for TTL and project characteristics, which reduces the coefficient on gender inequality. At the same time, the coefficient on the interaction between female TTL and gender inequality changes only slightly (from -0.074 to -0.070) and loses some statistical significance. Column 5 introduces country fixed effects, which helps reduce concerns that other time-invariant country characteristics such as the quality of long-run institutions or geography could be driving the effect. In this specification, the coefficient on the interaction between gender inequality and female TTL becomes slightly more negative and is again statistically significant at the 10 percent level.

[4]Another possibility is that they work on projects that are easier to implement. In the appendix we show that the female performance penalty arises only once we control for country fixed effects.

In other words, the evidence that projects led by women perform worse than those led by men in contexts with norms favoring men appears robust.

The next OLS specification, presented in column 6, includes both TTL and recipient country fixed effects. This is a highly demanding specification, as identification is achieved by comparing performance differences within a given TTL across countries. Inclusion of TTL fixed effects reduces degrees of freedom considerably, but the coefficient on the interaction between gender inequality and female TTL becomes even more negative and is now significant at the 5 percent level. A one standard deviation increase in gender inequality is associated with a 0.254 lower project outcome rating for female TTLs relative to male TTLs, on average. For example, a female TTL leading a project in Togo (90th percentile, index 1.38) would likely achieve a project outcome rating 0.56 points lower relative to her male counterparts than if she led the same project in Estonia (10th percentile of the gender inequality index, –0.82). Thus, female-led projects underperform relative to male-led projects in countries where local gender norms favor men, highlighting how cultural context can shape project effectiveness.

Finally, in the last OLS specification, presented in column 7, we add controls for being a female TTL and other country characteristics to rule out the possibility that factors other than gender norms are driving the effect. The resulting coefficient on the interaction between being a female TTL and gender inequality is largely unchanged, suggesting that this is indeed not the case.

While our OLS specifications controlling for country and TTL fixed effects are already highly demanding, columns 8 and 9 present instrumental variables (IV) estimates of the same specifications as columns 6 and 7, respectively. Both IV estimates are somewhat more negative and statistically significant at the 10% level. The Kleibergen–Paap statistics indicate that the instruments are not weak. According to the IV specification presented in column 8 a one standard deviation increase in gender inequality is associated with a 0.318-point lower project outcome rating for female TTLs, compared to 0.254 in the corresponding OLS specification (column 6). In other words, the OLS estimates are likely biased downward. One possible explanation is that women who take on projects in gender-unequal countries are, on average, particularly adept at navigating the local context, which partially offsets the observed performance penalty.

These results also hold when using alternative proxies for gender inequality, as shown in Appendix Table 5, which replicates the specifications from columns 3–9. The results consistently show a negative and typically significant interaction between being a female TTL and gender inequality.

7 Conclusion

Using a novel dataset linking World Bank project outcomes with administrative human resources data, we show two main findings: women are less likely to lead projects in countries with high gender inequality, and projects led by women in these contexts face a performance penalty relative to those led by men, even though the women who do lead are not less capable. The results attest to cultural constraints on women's productivity.

Our results do not imply that donors should avoid appointing women to lead projects

Table 5: Determinants of Project Outcomes – Alternative Gender Inequality Measures

	Dependent variable: IEG Outcome						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Gender Inequality - UNDP - GII							
TTL female	-0.070*	-0.056	-0.071*				
	(0.039)	(0.038)	(0.038)				
Gen. Ineq. (GII, STD)	-0.236***	-0.294***					
	(0.045)	(0.034)					
TTL female × Avg GII (STD)	-0.056	-0.070*	-0.082*	-0.113	-0.279**	-0.395	-0.561*
	(0.044)	(0.042)	(0.042)	(0.123)	(0.138)	(0.260)	(0.299)
Observations	4,156	4,015	4,014	2,943	2,943	2,943	2,943
KP stat						22	17.53
Panel B: Gender Inequality - SIGI (2023)							
TTL female	-0.057	-0.047	-0.066				
	(0.043)	(0.041)	(0.041)				
Gen. Ineq. (SIGI, STD)	-0.131***	-0.176***					
	(0.040)	(0.042)					
TTL female × SIGI (STD)	-0.083*	-0.102**	-0.116**	-0.146	-0.225**	-0.503**	-0.578**
	(0.046)	(0.045)	(0.047)	(0.108)	(0.107)	(0.207)	(0.225)
Observations	3,853	3,673	3,672	2,686	2,686	2,686	2,686
KP stat						28.68	23.84
Panel C: Gender Inequality - WBL ("de jure" index)							
TTL female	-0.068*	-0.058	-0.084**				
	(0.039)	(0.038)	(0.038)				
Avg. Gen. Ineq. (Inv. WBL, STD)	-0.183***	-0.187***					
	(0.053)	(0.052)					
TTL female × Inv WBL (STD)	-0.074	-0.092*	-0.122**	0.047	-0.044	-0.340	-0.424
	(0.054)	(0.052)	(0.052)	(0.134)	(0.125)	(0.289)	(0.282)
Observations	4,283	4,149	4,146	3,057	3,057	3,057	3,057
KP stat						23.96	23.37
Approval year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
TTL controls	-	Yes	Yes	Yes	Yes	Yes	Yes
Project controls	-	Yes	Yes	Yes	Yes	Yes	Yes
Country controls	-	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	-	-	Yes	Yes	Yes	Yes	Yes
TTL FE	-	-	-	Yes	Yes	Yes	Yes
Country Controls × TTL female	-	-	-	-	Yes	-	Yes

Notes: Standard errors are in parentheses. TTL controls include salary, age, education, professional grade, and prior experience leading projects. Project controls include log commitments, financing type, and sector dummies. Country controls include log GDP per capita, polity score, and dummies for war or financial crisis. *** p<0.01, ** p<0.05, * p<0.1.

in countries where norms favor men, as such appointments may generate demonstration effects and other positive spillovers beyond the scope of this study. Rather, they show that entrenched gender norms can undermine both equity and efficiency. The success of development interventions depends, in part, on addressing these social barriers.

References

Andersen, Jørgen Juel, Niels Johannesen, and Bob Rijkers, "Elite capture of foreign aid: Evidence from offshore bank accounts," *Journal of Political Economy*, 2022, *130* (2), 388–425.

Arndt, Channing, Sam Jones, and Finn Tarp, "What is the aggregate economic rate of return to foreign aid?," *The World Bank Economic Review*, 2016, *30* (3), 446–474.

Ashraf, Nava, Oriana Bandiera, and Virginia Minni, "Gender roles and the misallocation of labour across countries," 2022.

Ashton, Louise, Jed Friedman, Diana Goldemberg, Mustafa Zakir Hussain, Thomas Kenyon, Akib Khan, and Mo Zhou, "A Puzzle with Missing Pieces: Explaining the Effectiveness of World Bank Development Projects," *The World Bank Research Observer*, 2023, *38* (1), 115–146.

Bank, World, "Women, Business and the Law 2023," 2023. Accessed: 2025-08-04.

Bertrand, Marianne, "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.

Burnside, Craig and David Dollar, "Aid, policies, and growth," *American economic review*, 2000, *90* (4), 847–868.

_ **and** _ , "Aid, policies, and growth: reply," *American economic review*, 2004, *94* (3), 781–784.

Centre, OECD Development, "Social Institutions and Gender Index 2023 Global Report," 2023. Accessed: 2025-08-04.

Chiplunkar, Gaurav and Pinelopi Koujianou Goldberg, "Aggregate implications of barriers to female entrepreneurship," *Econometrica*, 2024, *92* (6), 1801–1835.

Denizer, Cevdet, Daniel Kaufmann, and Aart Kraay, "Good countries or good projects? Macro and micro correlates of World Bank project performance," *Journal of Development Economics*, 2013, *105*, 288–302.

Forum, World Economic, "Global Gender Gap Report 2021," 2021. Accessed: 2025-08-04.

Geli, Patricia, Aart C. Kraay, and Hoveida Nobakht, "Predicting World Bank Project Outcome Ratings," Technical Report 7001, The World Bank 2014.

Goldstein, Markus and Christopher Udry, "The profits of power: Land rights and agricultural investment in Ghana," *Journal of political Economy*, 2008, *116* (6), 981–1022.

Hsieh, Chang-Tai, Erik Hurst, Charles I Jones, and Peter J Klenow, "The allocation of talent and us economic growth," *Econometrica*, 2019, *87* (5), 1439–1474.

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

Kilby, Christopher, "Assessing the impact of World Bank preparation on project outcomes," *Journal of Development Economics*, 2015, *115*, 111–123.

Klasen, Stephan, "The impact of gender inequality on economic performance in developing countries," *Annual Review of Resource Economics*, 2018, *10* (1), 279–298.

Mel, Suresh De, David McKenzie, and Christopher Woodruff, "Are women more credit constrained? Experimental evidence on gender and microenterprise returns," *American Economic Journal: Applied Economics*, 2009, *1* (3), 1–32.

Olivetti, Claudia, Jessica Pan, and Barbara Petrongolo, "The evolution of gender in the labor market," in "Handbook of labor economics," Vol. 5, Elsevier, 2024, pp. 619–677.

Programme, United Nations Development, "Human Development Report 2020: The Next Frontier—Human Development and the Anthropocene," 2020. Accessed: 2025-08-04.

Qian, Nancy, "Making progress on foreign aid," *Annu. Rev. Econ.*, 2015, *7* (1), 277–308.

Udry, Christopher, "Gender, agricultural production, and the theory of the household," *Journal of political Economy*, 1996, *104* (5), 1010–1046.

8 Appendix

8.1 Additional Tables and Figures

Figure 1: Share of Female TTLs by Sector

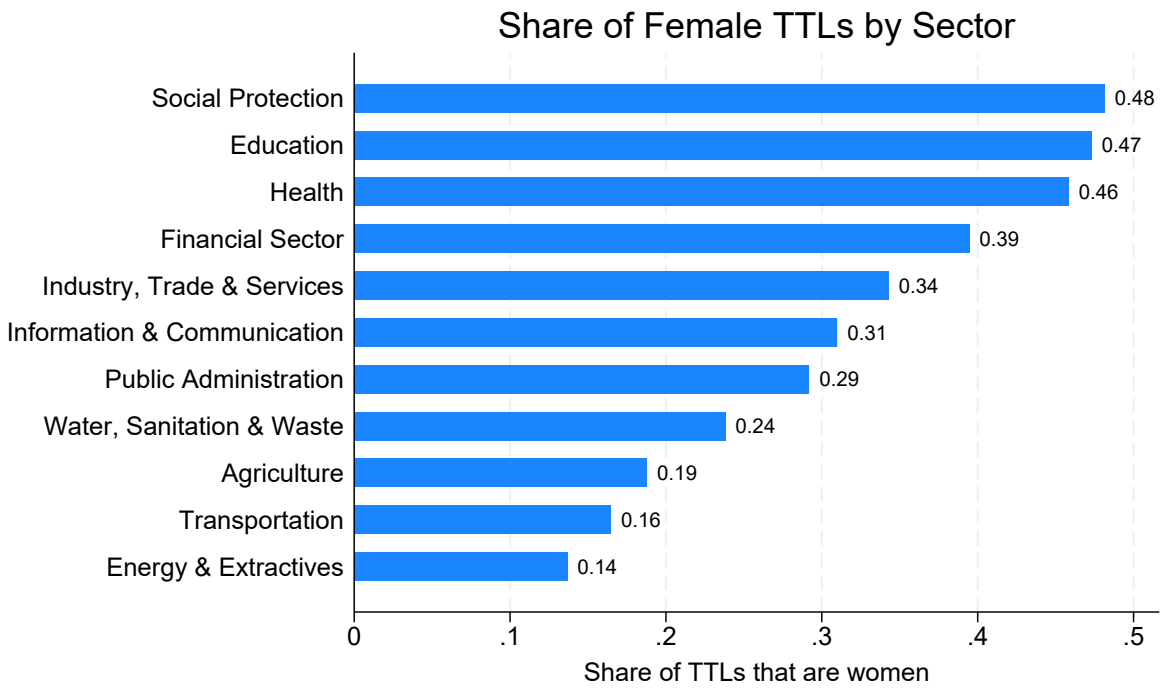


Figure 2: Share of Female TTLs by Grade

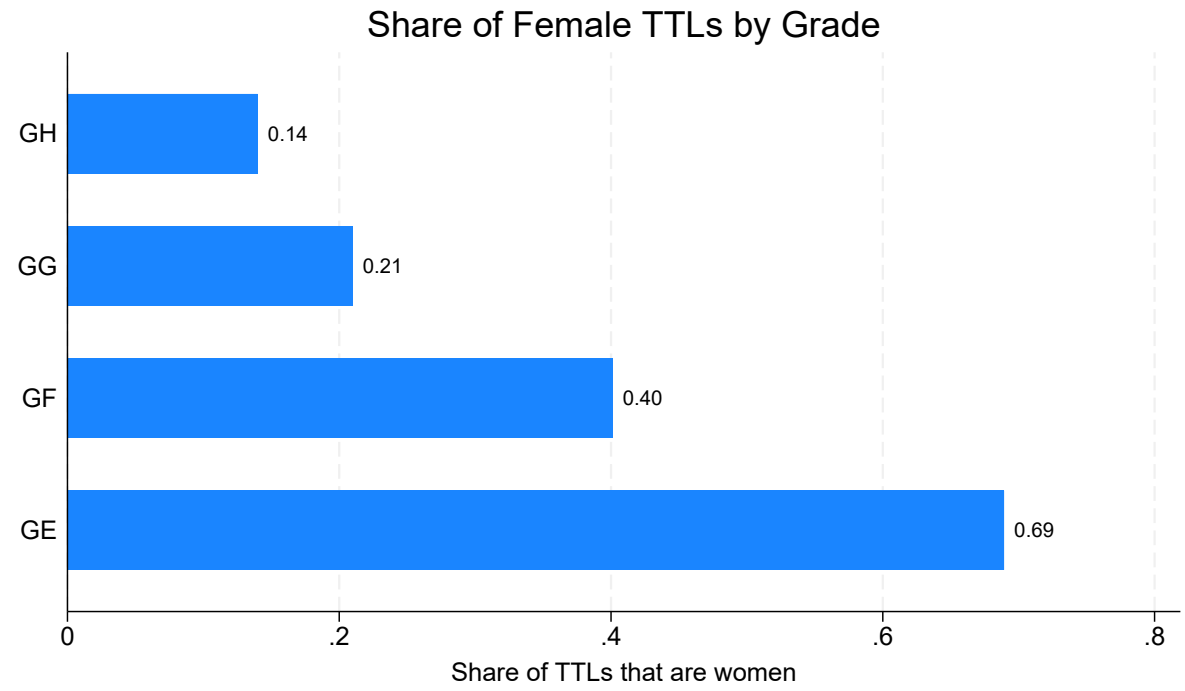


Figure 3: Prevalence of female TTLs vs Share of Staff that are Female

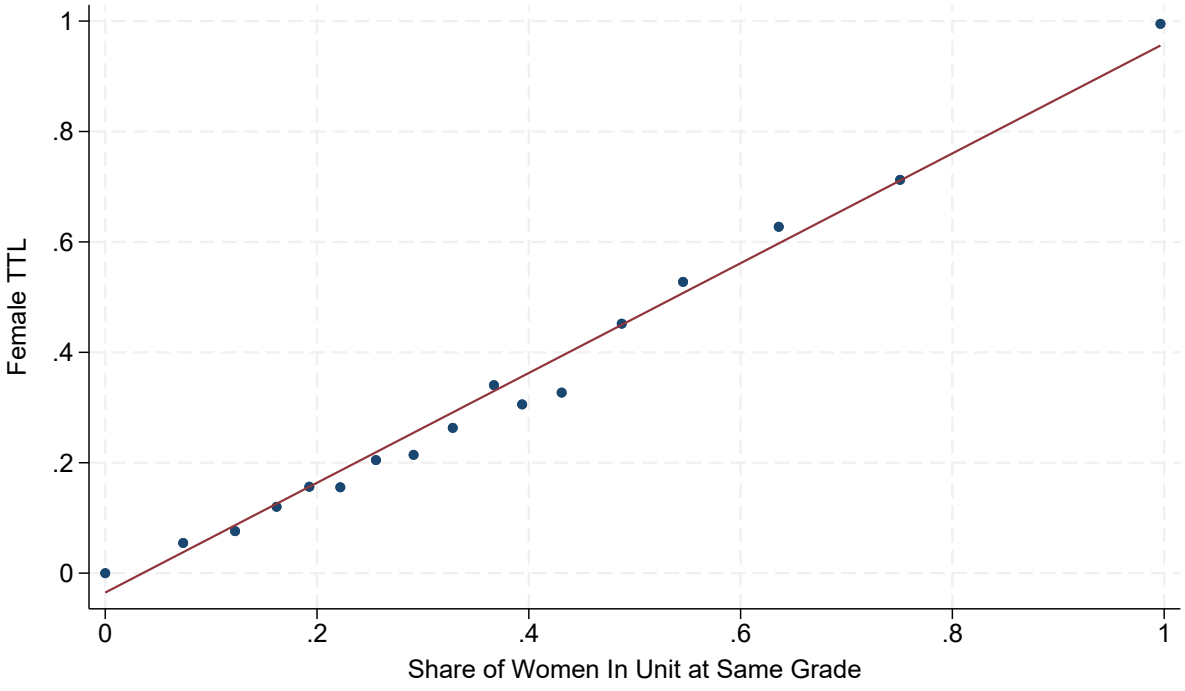


Table 6: TTL is a woman - Probit models with alternative gender inequality indicators

	(1)	(2)	(3)	(4)	(5)
Panel A: OECD					
Gender Inequality (SIGI))	-0.030**	-0.022*	-0.009	-0.003	-0.009
	(0.012)	(0.013)	(0.011)	(0.013)	(0.019)
Female share unit-grade					0.889***
					(0.026)
Female share unit-grade*GI (SIGI)					0.051
					(0.035)
N	4284	3674	3671	3362	3362
obs_p	0.291	0.307	0.307	0.306	0.306
pseudo_rs	0.00182	0.0543	0.127	0.127	0.291
Panel B: UNDP					
Gender Inequality (UNDP)	-0.210**	-0.221***	-0.209***	-0.210***	-0.069
	(0.085)	(0.084)	(0.062)	(0.081)	(0.093)
Female share unit-grade					0.936***
					(0.073)
Female share unit-grade*GI(UNDP)					-0.112
					(0.177)
N	4210	3611	3608	3336	3336
obs_p	0.288	0.304	0.304	0.304	0.304
pseudo_rs	0.00468	0.0588	0.128	0.128	0.291
Panel C: Gender Inequality WBL					
Gender Inequality (WBL)	-0.028**	-0.021*	-0.011	-0.008	-0.022
	(0.011)	(0.011)	(0.010)	(0.013)	(0.021)
Female share unit-grade					0.896***
					(0.024)
Female share unit-grade*GI (WBL)					0.071*
					(0.040)
N	4284	3674	367‘	3362	3362
TTL controls	No	Yes	Yes	Yes	Yes
Project controls	No	No	Yes	Yes	Yes
Country controls	No	No	No	Yes	Yes
N	4284	3674	3671	3362	3362
obs_p	0.291	0.307	0.307	0.306	0.306
pseudo_rs	0.00182	0.0543	0.127	0.127	0.291

Notes: This table presents estimates of the likelihood that the Task Team Leader is a woman. Marginal effects are presented with standard errors clustered by recipient country in parentheses. TTL characteristics include age, having a Ph.D, grade dummies (GE-analyst, GF-specialist, GH-expert). Project controls include log commitments, financing source dummies, and sector dummies. Country controls include log GDP per capita, polity score, war, crisis, and risk of natural hazards. *, **, *** indicate significance at the 10%, 5%, and 1% significance level, respectively.